FLUCTUATIONS AND ENTROPY IN THE ENERGETICS AND FUNCTION OF PROTEIN COMPLEXES

Vignesh Kasinath

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2014

Supervisor of Dissertation

_____

A. Joshua Wand, PhD., Benjamin Rush Professor of Biochemistry and Biophysics

Graduate Group Chairperson

_____

Kim A. Sharp, PhD., Associate Professor

Dissertation Committee

Kim A. Sharp, PhD., Associate Prof. of Biochemistry and Biophysics (Committee Chair)
Gregory Van Duyne, PhD., Jacob Gershon-Cohen Prof. of Medical Science
Heinrich Roder, PhD., Professor of Biochemistry and Biophysics (Fox Chase Cancer Center)
Ravinder Reddy, PhD., Professor of Radiology
Walter S. Englander, PhD., Jacob Gershon-Cohen Prof. of Medical Science
Charalampos Kalodimos, PhD., Professor of Chemistry and Chemical Biology (Rutgers)

To my parents whose support and encouragement were the only universal constants that mattered. . .

Acknowledgements

First and foremost I would like to thank Prof. Joshua Wand, who has been/will be an amazing mentor and whose insight and guidance were crucial for my thesis work. In addition to his support, I thank him for his patience, particularly in listening and providing critical comments on my many crazy and sometimes incorrect ideas. I would also like to thank Prof. Kim Sharp, who has also been very supportive and from whom I have learnt a lot from. I cannot thank him enough for his time to both teach me and brainstorm ideas with me. The knowledge and insight of both my mentors, Josh and Kim, has been an enriching experience and has served as a wonderful motivation through my thesis work.

I am also particularly thankful to Dr. Kathleen Valentine for teaching me the 'ABCD' of working with NMR and for her never ending enthusiasm for science. Our discussions were very interesting and something I looked forward to all the time.

I would also like to thank the past and present Wand lab members (John Gledhill, Nathaniel Nucci, Matt Stetz, Jackwee Lim, Yinan Fu, Igor Dodevski, Sabrina Bedard, Veronica Moorman, Kyle Harpole, Christine Jorge, Evan O'Brien, Brian Fuglestad, Bryan Marques). In particular, I would like to thank Yinan Fu, Matt Stetz, Jackwee Lim, John Gledhill and Nathaniel Nucci for helpful discussions and suggestions.

Special thanks to my friends Dr. Chandrasekhar Gopalakrishnan, Srivatsa Ray, Preeti Sareen, Dr. Purushottam Dixit, Manas Rachh, Dr. Karthik Shekhar, Kaushik Jayaram, Sadanand Kamath, Tannishtha Sanyal and Kartik Mohta who have kept me company and put up with me in the worst of times!

Finally, I would like to thank the people responsible for where I am today, my parents. Their support and encouragement was more important than they may realize. For my dad, who has likely spent more time in book stores than at home to get me that one extra book which might make a difference, I hope he thinks I have done well so far! For my mom, who has always reminded me to 'Keep Calm and Enjoy the Science'.

ABSTRACT

FLUCTUATIONS AND ENTROPY IN THE ENERGETICS AND FUNCTION OF PROTEIN COMPLEXES

Vignesh Kasinath

A. Joshua Wand, PhD.

Conformational entropy is a potentially important thermodynamic parameter contributing to protein function. Quantitative measures of conformational entropy are necessary for an understanding of its role but have been difficult to obtain experimentally. We have recently introduced an empirical calibration method that utilizes the changes in conformational dynamics as a proxy for changes in conformational entropy. This approach raises several questions with regards to the microscopic origins of the measured conformational entropy and also the validity of the empirical method to different protein-ligand complexes. One of the goals in this work was to probe the microscopic origins of the link between conformational dynamics and conformational entropy. Molecular dynamics simulations of seven proteins gave an excellent correlation of the calculated side chain motion with those of the corresponding measured side-chain motion derived from NMR relaxation. The simulations show that the motion of methyl-bearing side chain motion are sufficiently coupled to that of other side chains and serve as excellently reporters of the overall side chain conformational entropy. These results validate the use of experimentally accessible measures of methyl motion - the NMR-derived generalized order parameter - as a proxy from which to derive changes in protein conformational entropy. A slightly modified weighting scheme to project the change in dynamics of experimental methyl dynamics into conformational entropy results in validating the generality of the dynamical proxy approach to measure conformational entropy changes of any protein-ligand interaction. Furthermore, In this work this approach has been extended to 35 different protein-ligand complexes each with different ligand types (small molecules, DNA/RNA, peptides and proteins) and binding affinities (10^{-4} - 10^{-15} M). There is excellent agreement between the NMR-measured conformational dynamics derived measure of conformational entropy and the total binding entropy, essentially

postulating a 'Universal Entropy Meter'. This universal entropy meter can be utilized to measure the conformational entropy change for any protein-ligand interaction.

The second major goal of this work is to understand the role played by conformational entropy in very high affinity interaction. A dominant view of very high affinity interactions involving proteins is that they are largely driven by both large favourable interactions at the interaction interface and by an increase in solvent entropy due to the creation of dry or solvent depleted interface. The role of conformational entropy, though admitted as a potentially favourable contribution, remains largely obscured by this view. The vast majority of high affinity interactions contain an abundance of aromatic amino acids at the interaction interface. In order to characterize the contributions of aromatic amino acid side chains, a new ^{13}C isotope enrichment scheme is introduced which eliminates all the complexities associated with the NMR relaxation measurement of aromatic amino acid side chains. The application of this scheme to different protein systems reveal that aromatic amino acid side chains display a wide range of dynamics than previously understood. Utilizing both the NMR-measured conformational dynamics of methyl-bearing and aromatic amino acid side chains, I find that protein conformational entropy can indeed play a pivotal role in achieving very high affinity interactions. In the barnase-barstar protein-protein complex (K_d = 10^{-14} M) and the histamine-binding protein bound to small molecule histamine (K_d = $2.5x10^{-9}$ M), large favourable changes in conformational entropy offsets unfavourable entropic contributions from both solvent and rotational-translational contributions. Furthermore, the binding of barnase to the 4mer DNA (dCGAC) is found to be driven by favourable changes in conformational entropy in the absence of any enthalpic contribution to binding. Collectively, these results reveal that conformational entropy can provide large favourable contributions to the binding of both large and small ligands by offsetting other potentially unfavourable contributions to binding. This suggests that conformational entropy could represent the missing piece of the puzzle in the quest for the rational design of high-affinity and high-specificity pharmaceutical drugs.

Contents

Acknowledgements iii

Abstract iv

List of Tables ix

List of Figures xii

1 Introduction 1
- 1.1 Thermodynamics . . . 1
  - 1.1.1 Entropy change upon ligand binding . . . 5
- 1.2 Nuclear Magnetic Resonance (NMR) . . . 7
  - 1.2.1 Introduction to NMR [1] . . . 7
  - 1.2.2 Probing molecular motions using NMR . . . 10
  - 1.2.3 Backbone Relaxation and side chain Relaxation . . . 12
    - 1.2.3.1 Backbone Relaxation . . . 16
    - 1.2.3.2 Side chain Relaxation . . . 18
- 1.3 Conformational Entropy from Protein Dynamics: The History . . . 22
  - 1.3.1 Model Dependent Approach: Harmonic Oscillator . . . 24
  - 1.3.2 Model Independent Approach: Empirical Calibration . . . 25
  - 1.3.3 Questions that need to be addressed . . . 27

2 Insights into the microscopic origins of protein conformational entropy 30
- 2.1 Introduction . . . 30
- 2.2 Results . . . 35
  - 2.2.1 Simulation versus Experiment: O^2_{axis} . . . 39
  - 2.2.2 Rotamer Entropy from MD and $\mathbf{O^2_{axis}}$ . . . 43
  - 2.2.3 Total protein conformational entropy . . . 51

2.3 Reformulation of the "Entropy Meter" . . . 55
2.4 Summary and Outlook . . . 59

3 Aromatic amino acid side-chain relaxation 64
3.1 Introduction . . . 64
3.2 Establishing ^{13}C labelling in aromatic amino acids . . . 69
3.3 Fast dynamics of aromatic amino acid side-chain in calmodulin . . . 74
3.4 Summary and Discussion . . . 80

4 A sharp thermal transition of fast aromatic ring dynamics in ubiquitin 85
4.1 Introduction . . . 85
4.2 Effect of temperature on fast motions of aromatic rings . . . 88
4.3 Examining the cooperative nature of the aromatic ring motions . . . 98
4.4 Pressure induced fluctuations in aromatic dynamics . . . 102
4.5 Summary and Outlook . . . 106

5 Conformational entropy in very high-affinity interactions 110
5.1 Introduction . . . 110
5.2 Thermodynamics of protein-ligand systems: barnase-barstar; HBP(D24R)-histamine . . . 113
5.2.1 Very high affinity: barnase-barstar . . . 113
5.2.2 Very high affinity: histamine-binding protein:histamine . . . 115
5.2.3 Moderate affinity: barnase-dCGAC . . . 117
5.2.4 Weak binder: HBP(D24R)-serotonin . . . 119
5.2.5 Summary . . . 121
5.3 Results . . . 122
5.3.1 NMR chemical shift assignments . . . 122
5.3.2 Increase in conformational entropy upon barnase-barstar high affinity interaction . . . 123
5.3.3 Dynamics of the interface waters in barnase-barstar complex . . . 131
5.3.4 Conformational entropy can act as a driving force for binding interactions . . . 136
5.3.5 Conformational entropy offsets unfavourable entropic contributions to binding . . . 141
5.3.6 Conformational entropy modulates binding affinity . . . 143
5.4 Conclusions and Discussion . . . 144

6 A Universal Entropy Meter 150
6.1 Introduction . . . 150
6.2 Results . . . 154
6.2.1 Expanding the 'Entropy Meter' . . . 155

6.2.2 Recalibration of solvent entropy for a 'Universal Entropy Meter' . . 158
6.3 Conclusions 162

7 Conclusions **166**
7.1 Summary 166
7.2 Future Directions 170

A Appendix Chapter 2 **172**

B Appendix Chapter 4 **192**

C Appendix Chapter 5 **196**

D Appendix Chapter 6 **212**

Bibliography **216**

List of Tables

1.1 Amino acids and at least one type of interaction they participate in 5

2.1 Proteins used in MD simulations 38
2.2 Characteristics of the protein set used for MD simulations 38
2.3 Comparison of MD simulations and Experimental O^2_{axis} statistics 42
2.4 Correlation of normalized entropy (S_b) to MD O^2_{axis} 48

3.1 Lipari-Szabo Model-Free Parameters for Aromatic Ring Motion in Calcium-Saturated Calmodulin grown in $[4-^{13}C]$-erythrose 79
3.2 Lipari-Szabo Model-Free Parameters for Aromatic Ring Motion in Calcium-Saturated Calmodulin grown in $[1-^{13}C]$-glucose 81

4.1 Temperature dependence of the macromolecular tumbling time of ubiquitin 90

5.1 Thermodynamics of binding of the different protein-ligand complexes . . . 121

A.1 Methyl order parameters and Side chain entropies of ubiquitin from MD simulations 172
A.2 Methyl order parameters and Side chain entropies of alpha3D from MD simulations 174
A.3 Methyl order parameters and Side chain entropies of CaM-smMLCK from MD simulations 175
A.4 Methyl order parameters and Side chain entropies of CaM-nNOS from MD simulations 179
A.5 Methyl order parameters and Side chain entropies of Adipocytle lipid binding protein (ADBP) from MD simulations 183
A.6 Methyl order parameters and Side chain entropies of Hen egg white lysozyme (HEWL) from MD simulations 186
A.7 Methyl order parameters and Side chain entropies of Cytochrome c2 (Cyt c2) from MD simulations 189

B.1 Temperature dependence of ubiquitin aromatic order parameter and internal correlation times 192

B.2 Temperature dependence of ubiquitin aromatic order parameter and internal correlation times II . . . 192
B.3 Pressure dependence of ubiquitin F45 aromatic order parameter. 193
B.4 Pressure dependence of ubiquitin Y59 aromatic order parameter. 193
B.5 Pressure dependence of ubiquitin F4 aromatic order parameter. 193
B.6 Pressure dependence of ubiquitin F45 aromatic internal correlation times. 194
B.7 Pressure dependence of ubiquitin Y59 aromatic internal correlation times. 194
B.8 Pressure dependence of ubiquitin F4 aromatic internal correlation times. . 194
B.9 Temperature dependence of ubiquitin aromatic order parameter from 120ns MD simulations . . . 194
B.10 Pressure dependence of ubiquitin F45 aromatic order parameter from MD simulations. . . . 195
B.11 Pressure dependence of ubiquitin Y59 aromatic order parameter from MD simulations. . . . 195
B.12 Pressure dependence of ubiquitin F4 aromatic order parameter from MD simulations. . . . 195

C.1 Backbone order parameters of free barnase and barnase in complex with barstar . . . 196
C.2 Backbone order parameters of free barstar and barstar in complex with barnase . . . 198
C.3 Methyl side chain order parameters of free barnase and barnase in complex with barstar . . . 200
C.4 Methyl side chain order parameters of free barstar and barstar in complex with barnase . . . 202
C.5 Backbone order parameters of free barnase and barnase in complex with 4mer dCGAC . . . 203
C.6 Methyl side chain order parameters of free barnase and barnase in complex with 4mer dCGAC . . . 205
C.7 Aromatic side chain order parameters of free barnase and barnase in complex with barstar . . . 207
C.8 Aromatic side chain order parameters of free barstar and barstar in complex with barnase . . . 207
C.9 Aromatic side chain order parameters of free barnase and barnase in complex with dCGAC . . . 208
C.10 Barnase hydration measurement (NOE/ROE) in the free state and complex states . . . 208
C.11 Barstar hydration measurement (NOE/ROE) complex state bound to barnase . . . 210

D.1 List of all protein-ligand complexes used in Chapter 7 . . . 212

D.2 Thermodynamics and NMR relaxation parameters for the calibration of the 'Entropy Meter'. 213
D.3 Thermodynamics and NMR relaxation parameters for the calibration of the 'Entropy Meter'. 214

List of Figures

1.1 ITC of a protein binding to a ligand . . . 3
1.2 External magnetic field induced splitting of spin states . . . 8
1.3 Relaxation of nuclei in external magnetic field . . . 11
1.4 Time scales of motion . . . 13
1.5 Time Correlation of bond vector fluctuations . . . 14
1.6 Backbone order parameter distribution in ubiquitin . . . 17
1.7 side chain methyl group order parameter distribution in ubiquitin . . . 20
1.8 Order parameter distribution of aromatic ring side chain in calmodulin . . . 21
1.9 Model dependent correlation of $-T\Delta S_{conf}$ vs $-T\Delta S_{total}$ . . . 25
1.10 Model independent correlation of $-T\Delta S_{conf}^{NMR}$ vs $-T\Delta S_{conf}^{ITC}$ . . . 27

2.1 Convergence of angular time correlation functions . . . 41
2.2 Comparison of average O_{axis}^2: MD vs Experiment . . . 43
2.3 Comparison of site-to-site O_{axis}^2: MD vs Experiment . . . 44
2.4 Correlation of MD O_{axis}^2 and rotamer entropy . . . 46
2.5 Correlation of site-to-site MD O_{axis}^2 and rotamer entropy . . . 47
2.6 A representative rotamer landscape for methyl side-chains . . . 49
2.7 Rotamer entropy contributions from within well fluctuations . . . 50
2.8 Methyl groups as reporters of total conformational entropy . . . 54
2.9 Empirical calibration of CAP binding to DNA . . . 56
2.10 New universal "Entropy Meter" calibration . . . 58

3.1 Carbon labelling of aromatic amino acid . . . 69
3.2 Selective isotope enrichment of aromatic residues . . . 72
3.3 ^{13}C labelling efficiency of erythrose . . . 73
3.4 ^{13}C incorporation dictates labelling efficiency . . . 75
3.5 ^{13}C HSQC Flavodoxin C55A . . . 76
3.6 $(4\ -^{13}C)$-erythrose versus $(1\ -^{13}C)$-glucose: R_1, $R_{1\rho}$ relaxation . . . 77
3.7 Comparison of ^{13}C HSQC: Erythrose vs Glucose . . . 80

4.1 Relative localization of aromatic rings in ubiquitin . . . 89

4.2 Experimental temperature dependence of ps-ns aromatic ring motion in ubiquitin . . . 92
4.3 van't Hoff analysis of the fast motions of aromatic side chains in ubiquitin 93
4.4 Experimental temperature dependence of ps-ns aromatic ring correlation times in ubiquitin . . . 94
4.5 Temperature dependence of MD derived ps-ns aromatic ring motion in ubiquitin . . . 95
4.6 Aromatic ring flip motions above 300 K in ubiquitin . . . 97
4.7 Time correlation functions of aromatic rings from MD simulations . . . 99
4.8 Packing efficiency of ubiquitin as a function of temperature . . . 100
4.9 Contributions of the side chains to the pressure virial . . . 101
4.10 Pressure sensitivity of fast motions of aromatic rings in ubiquitin . . . 103
4.11 Pressure sensitivity of fast motions of aromatic rings in ubiquitin from MD simulations . . . 104
4.12 Non-linear pressure response of methyl motions in ubiquitin . . . 106
4.13 Experimental temperature dependence of ps-ns aromatic ring motion in $Ca^{2+} - Calmodulin - smMLCK$ . . . 108

5.1 Minimal structural perturbation upon barnase-barstar complex formation . 114
5.2 Interface of barnase-barstar consist of several water molecules . . . 115
5.3 Structural perturbations and thermodynamics of HBP(D24R) binding to histamine . . . 117
5.4 Water mediated interactions of histamine binding to HBP(D24R) . . . 118
5.5 Structural pertubations and thermodynamics of barnase-dCGAC complex 119
5.6 Dissociation constant of barnase-dCGAC . . . 120
5.7 Binding of serotonin to HBP(D24R) in the presence and absence of histamine . . . 121
5.8 ^{15}N amide chemical shift assignments of free barnase and free barstar . . 124
5.9 ^{13}C-methyl chemical shift assignments of free barnase and free barstar . 125
5.10 ^{15}N and ^{13}C-methyl chemical shift assignment of barnase-barstar complex 126
5.11 ^{13}C-aromatic chemical shift assignments of free barnase and free barstar 127
5.12 Changes in motion of the polypeptide backbone of barnase and barstar upon formation of the complex . . . 128
5.13 Correlation of ^{2}H vs ^{13}C O^2_{axis} of free barnase and complexed barnase . . 129
5.14 Distribution of the changes in O^2_{axis} in barnase-barstar complex . . . 130
5.15 Distribution of the changes in motion of aromatic side chains in barnase-barstar complex . . . 131
5.16 Water-protein interaction at the interface of barnase-barstar complex . . . 132
5.17 Reverse micelle encapsulation mediated hydration measurement of barnase-barstar complex . . . 134

5.18 Reverse micelle encapsulation mediated hydration measurement of free barnase . . . 135
5.19 Motional changes in the polypeptide backbone of barnase upon dCGAC binding . . . 137
5.20 Motional changes in the methyl-bearing amino acid side chains of barnase upon dCGAC binding . . . 138
5.21 Motional changes in the aromatic amino acid side chains of barnase upon dCGAC binding . . . 139
5.22 Reverse micelle encapsulation mediated hydration measurement of barnase-dCGAC complex . . . 140
5.23 Motional changes of the backbone amide and methyl side chains in HBP(D24R) upon histamine binding . . . 142
5.24 Motional changes of the methyl side chains in HBP(D24R)/HBP(D24R)-histamine upon serotonin binding . . . 144
5.25 Interplay between the different thermodynamic components of the binding interactions . . . 148

6.1 Binding energetics of the different protein-ligand complexes . . . 154
6.2 Calibration of the dynamical proxy for protein conformational entropy . . . 158
6.3 A Universal Entropy Meter for quantitative estimates of protein conformational entropy . . . 161
6.4 Dissection of the free energy of binding into the different energetic components . . . 165

Chapter 1

Introduction

1.1 Thermodynamics

Thermodynamics is a part of science which defines the rules that govern any process which involves energy and work. This seemingly simple science lays the foundation from which energetics of chemical reactions can be understood. Not surprisingly, all biological processes are subject to thermodynamics. Almost all biochemical processes (chemical processes associated with cellular function) are effected through proteins, three-dimensional molecules made up of amino acids. Proteins are involved in replication [2], signalling [2, 3], enzymatic reactions [2] and many many other vital processes. Understanding how proteins carry out this spectrum of functions is one of the central questions in biology.

All biochemical processes are accomplished by interaction among different elements (proteins, nucleic acids) participating in the process. Hence, understanding protein

function is to gain insight into the mechanism governing interaction of different elements in the biochemical process, i.e., protein-protein interactions, protein-DNA/RNA interaction, protein-small molecule interaction, together labelled as protein-ligand interactions. The factors influencing these interactions can be understood by studying the thermodynamics involved. The propensity for any reaction depicted as

$$[Protein] + [Ligand] \rightleftharpoons [Protein - Ligand] \tag{1.1}$$

to be thermodynamically favourable or unfavourable can be understood using:

$$\Delta G = \Delta H - T\Delta S \tag{1.2}$$

Equation 1.2 is one of the most important equations and will be referred to time and again throughout this thesis. What does Equation 1.2 represent? It defines the energy involved when a ligand binds a protein to form a protein-ligand complex as indicated in 1.1 [4]. In detail, ΔG, represents the change in *Gibbs-Helmholtz Free Energy* associated with the reaction 1.1 while ΔH represents net *Enthalpy change* which provides insight into the structural aspects involved in the interaction of protein and ligand and finally $-T\Delta S$ represents the energy associated with the net *Entropy change* for the reaction. Understanding the fundamental aspects of proteins, namely their three-dimensional structure and dynamics, which contribute to enthalpy and entropy and thereby its function is one of the important questions tackled in biology.

Calorimetric measurements have aided in the measurement of changes in enthalpy and free energy and indirectly changes in entropy for reactions such as those in 1.1 [5]. Figure 1.1 shows an example isothermal calorimetry (ITC) measurement for a reaction similar to 1.1.

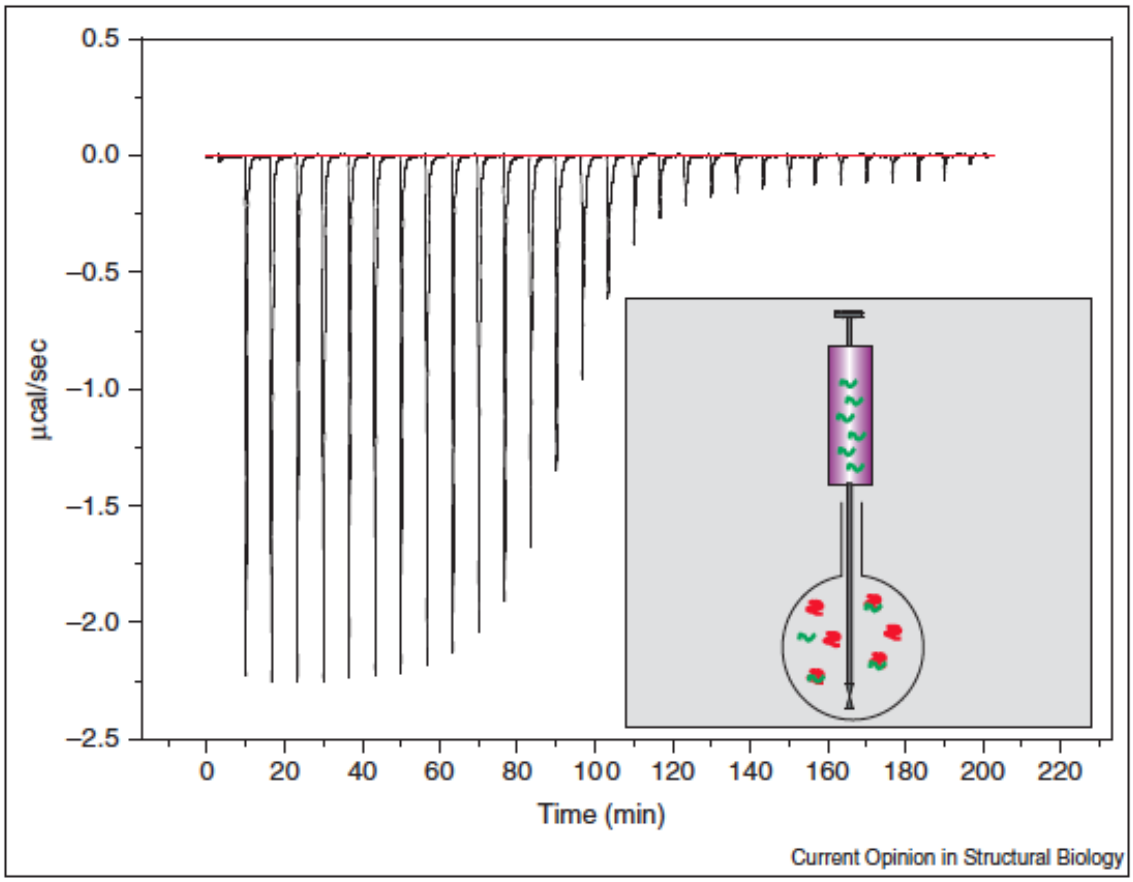


Figure 1.1: ITC Measurement of a protein binding to a ligand (S Leavitt, E Freire, *unpublished data*)

The basis of ITC measurements rely on the fact that titrating ligand to a protein or vice-versa, gives rise to the characteristic heat of dilution. The amount of power required (y-axis, Figure 1.1) to maintain a constant temperature difference between reaction and

reference cell is then calculated based on the amount of heat of dilution. The area under each peak is the heat associated with the binding process. The free energy change of the reaction 1.1 is then calculated using the affinity constant (K_a) associated with the binding of protein to ligand or vice-versa. This affinity constant is derived from a straightforward analysis of 1.1.

$$\Delta G = -RTln(K_a) \tag{1.3}$$

Over the last few decades, an intense focus on the 3-dimensional structure of proteins has resulted in a thorough understanding of the origins of enthalpy associated with reactions such as 1.1 [6, 7]. A survey of several protein-ligand interactions reveal that, *van der Waals*, *electrostatic*, and *hydrogen bond* interactions are the primary sources of the associated enthalpy for 1.1 [6, 8]. These vary between different proteins as they depend on the amino acid composition in the protein/ligand binding site. The non-uniform contribution of the different amino acids to each of these different types of interaction has however hindered the computation of the energetics governing protein-ligand interactions [9] .

Table 1.1 lists all the amino acids and at least one type of interaction they may participate in, respectively. What about *Entropy*?

Amino Acid	Type of Interaction
Alanine	van der Waals
Arginine	Electrostatic
Asparagine	Hydrogen bond
Aspartic acid	Electrostatic
Cysteine	Hydrogen bond
Glutamic acid	Electrostatic
Glutamine	Hydrogen bond
Glycine	van der Waals
Histidine	Hydrogen bond
Isoleucine	van der Waals
Leucine	van der Waals
Lysine	Electrostatic
Methionine	Hydrogen bond
Phenylalanine	van der Waals
Proline	van der Waals
Threonine	Hydrogen bond
Trptophan	Hydrogen bond
Tyrosine	Hydrogen bond
Serine	Hydrogen bond
Valine	van der Waals

Table 1.1: Amino acids and at least one type of interaction they participate in

1.1.1 Entropy change upon ligand binding

The total entropy change association with reaction 1.1 consists of several different contributions, namely:

- Solvent Entropy ($\Delta S_{solvent}$)
- Conformational Entropy (ΔS_{conf})
- Rotational/Translational Entropy (ΔS_{RT})

$$-T\Delta S = -T[\Delta S_{solvent} + \Delta S_{conf} + \Delta S_{RT}] \tag{1.4}$$

$$\Delta G = \Delta H - T[\Delta S_{solvent} + \Delta S_{conf} + \Delta S_{RT}] \tag{1.5}$$

Solvent entropy contributions have been estimated using the three-dimensional structures of protein in the free and ligand bound state [10]. These calculations have become more reliable wherein data from several proteins are used to parametrize the relationship between buried surface area and solvent entropy [10]. Very briefly, upon complex formation, buried waters are released giving rise to a gain in solvent entropy. This forms the basis for relating the change in buried surface area to the solvent entropy change upon complex formation.

Why is it important to understand the role played by conformational entropy (ΔS_{conf})? The typical change in conformational entropy associated with the unfolding of proteins from folded to unfolded state is of the same order of magnitude or even one order of magnitude less than that of the residual conformational entropy of the folded protein [11]. In other words, there is substantial residual conformational entropy in the folded state of proteins [11]. This suggests that conformational entropy of proteins could in theory contribute substantially to the free energy of reaction 1.1. Yet, protein conformational entropy contributions to 1.1 have been largely **ignored** or assumed to be **zero**. Protein conformational entropy has only recently been investigated in detail due to the difficulty associated with quantitative measurements of the same [12–16]. Previously,

estimates of conformational entropy exclusively came from computer based calculations aided by molecular dynamics (MD) simulations [11]. These however depend on the accuracy of MD simulations and were not sufficient to provide accurate quantitative measurement of conformational entropy changes. The experimental measurement of protein conformational entropy has proven challenging.

Nuclear Magnetic Resonance (NMR) has proved to be crucial in addressing this problem.

1.2 Nuclear Magnetic Resonance (NMR)

1.2.1 Introduction to NMR [1]

The nucleus of atoms, i.e. neutrons and protons, possess an intrinsic property, *spin*. Every nucleus is associated with a characteristic spin value, (eg) $\frac{1}{2}$ (^1H, ^{13}C, ^{15}N), 1 (^2H), ..., which in turn is related to the magnetic moment (μ) of that particular nucleus through its respective spin quantum number (S) and gyromagnetic ratio (γ) according to 1.6.

$$\mu = \gamma S \tag{1.6}$$

A spin $\frac{1}{2}$ nuclei, say ^1H or ^{13}C, has its spin in two possible states:

- spin-up state ($\frac{1}{2}$)

- spin-down state ($-\frac{1}{2}$)

These two states are degenerate when the nuclei is isolated, i.e. the two states are of equal energy. However, when this spin $\frac{1}{2}$ nuclei is placed in an external magnetic field, the interaction of these two spin states with the external magnetic field results in the two states being of different energy. This difference in energy (E) can be related to the

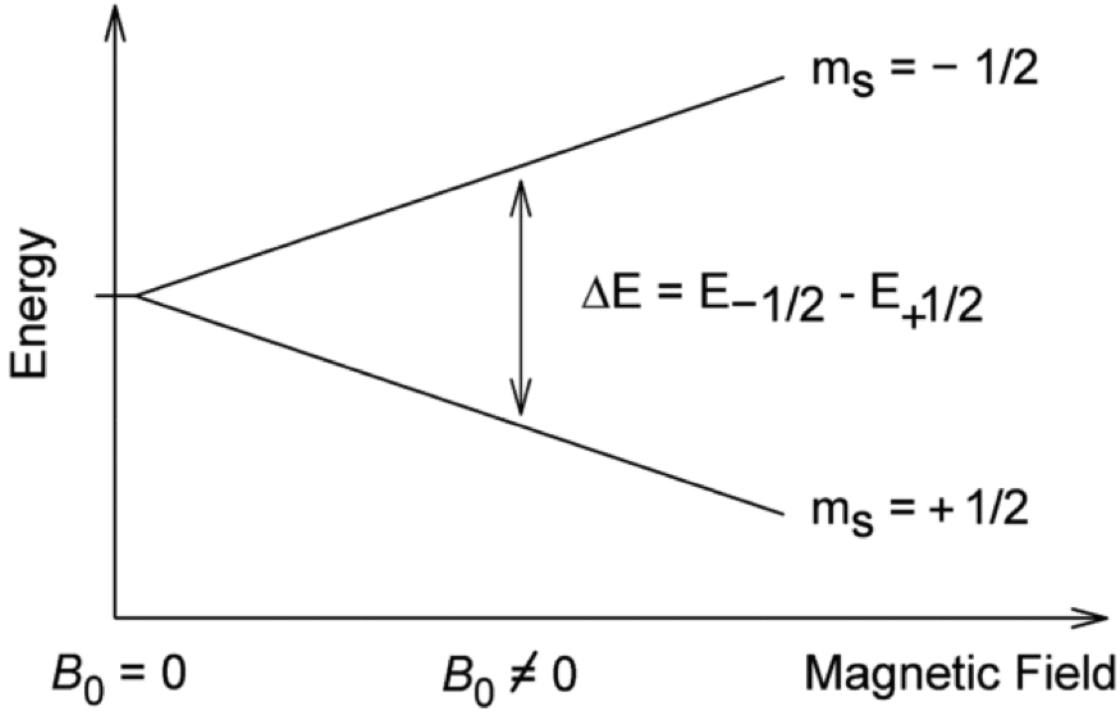


Figure 1.2: External magnetic field induced splitting of spin states

magnetic moment of the nuclei and the field strength of the external magnetic field (B_O) using 1.7.

$$E = \mu.B_O \tag{1.7}$$

Interestingly, due to this energy difference between the spin states, the precession of the nuclear magnetic moment relative to the external magnetic field will be different for the two states. One of the spin states will be aligned with the external magnetic field and the other aligned against. A precessing nuclear spin is then said to be in "resonance" when the energy difference between the spin states is the same as the energy applied through the external electromagnetic radiation. In other words, if the frequency of the applied electromagnetic radiation matches with the difference in energy of the two spin states then the nuclei absorbs energy and precesses in an external magnetic field and we observe resonance.

$$\begin{aligned} \Delta E = \hbar\omega_o &= \mu.B_O \\ \omega_o &= \frac{\gamma B_O}{2\pi} \end{aligned} \tag{1.8}$$

The precession frequency (ω_o) is called *Larmor* precession rate. Does the larmor frequency as defined above suggest that for all spin $\frac{1}{2}$ nuclei (with same γ) in an external magnetic field it takes the same value? **No**. In a protein, different nuclei are surrounded by various other nuclei and that each nuclei is surrounded by its own shell of electrons. This gives rise to the effect of *shielding*. This shielding of the nuclei from external magnetic field by the surrounding electrons reduces or increases the magnetic field experienced by the nucleus. Also, due to the complex nature of protein structure and that the chemical environment is so heterogeneous throughout the protein, the larmor frequency experienced by nuclei in different parts of the protein are different. This shift

in NMR larmor frequency due to the nature of chemical environment is termed *chemical shift*.

The ability of NMR to probe the chemical shift of all spin active probes in a macromolecule, (eg) protein, makes it a very powerful technique. In summary, every single NMR spin active nuclei in a protein experiences distinct chemical environments (baring some superimposed symmetry considerations) which in turn can be probed by NMR. Thus, NMR provides the tools for identifying individual atomic nuclei and the perturbations they undergo due to some external influence, (for example) binding of a ligand. Upon studying such perturbations, we gain valuable insight into the fundamentals of the binding process.

1.2.2 Probing molecular motions using NMR

It has been well established that proteins are not just static three-dimensional static structures. Proteins indeed undergo significant rotational, translational, and librational motions at the individual atomic level [11]. NMR provides a very concise way to probe some of these motions. At equilibrium, the spins in proteins are all aligned to the external magnetic field by some degree characteristic to the particular nuclei type. However, when these spins are perturbed to produce a non-equilibrium distribution of spin states, these spins *relax* back to their equilibrium distribution. This phenomenon of how fast spins tend to return to their equilibrium distribution is termed *relaxation*. NMR relaxation

is a powerful tool to study the molecular motions of atoms/bond vectors within proteins [17]. Each spin returns to its equilibrium distribution at a different rate which is influenced by the chemical environment surrounding it. There are two types of relaxation processes that can occur:

- Longitudinal/Spin-lattice relaxation (T_1)
- Transverse/Spin-spin relaxation (T_2)

The external magnetic field in NMR is typically applied along the z-axis. Longitudinal relaxation deals with the relaxation of the z component of the spins returning back to their equilibrium population while transverse relaxation deals with the transverse (xy) component of the spins. As mentioned earlier, the relaxation times for different nuclei

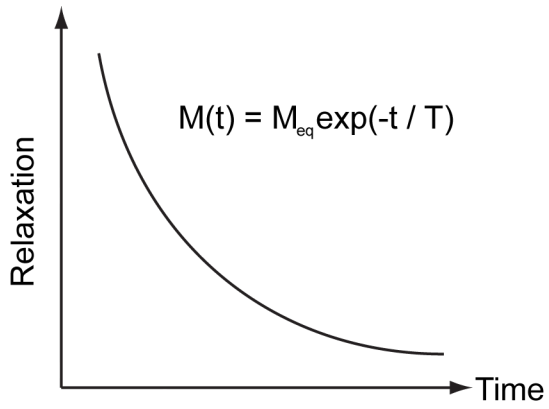


Figure 1.3: Relaxation of nuclei in external magnetic field

are strongly dependent on the surrounding environment, i.e., the nearby nuclei. The relaxation times contain vital information regarding the motion of the bond vector containing the nuclei as well as the time scale at which it fluctuates in the interior of the

protein. Extracting this information paves the way for probing the molecular motions of proteins. The theory behind extracting this information has been well documented and hence will not be discussed here [17]. However, a brief account of the two main applications of relaxation, *Backbone (main chain) relaxation* and *side chain relaxation*, and their corresponding theory will be presented. These two methods have been extensively used throughout this thesis. The following sections will outline in brief the idea and theory behind these methods as well as their connection to molecular motion of proteins.

1.2.3 Backbone Relaxation and side chain Relaxation

Proteins, which are polymers composed of many different amino acids, can be broadly classified to be made of two main elements:

- Backbone or Main chain
- Side chain (methyl, aromatic, and others)

Due to the complicated nature of these relaxation methods, probing the motions of all these elements simultaneously is a major challenge in NMR. Hence, motions pertaining to these elements are probed individually.

Probing the relaxation of protein backbone, i.e. the amide bond, provides information on the motions experienced by the main chain alone without regard for the side chains. The

spectrum of molecular motions in proteins span a variety of time scales (1.4) and NMR has proven useful to probe each of these time scales. The different type of motions probed by NMR include:

- Librational motions (very fast motions, < nanoseconds)
- Molecular tumbling (nanoseconds)
- Structural transitions, ligand binding (microseconds-milliseconds)
- Molecular diffusion (milliseconds-seconds)
- Hydrogen exchange (seconds-minutes)

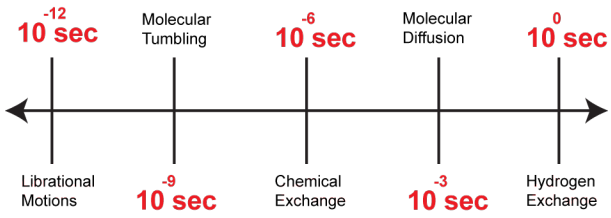


Figure 1.4: Time scales of motion

Backbone and side chain relaxation methods described here are primarily focussed on the fast time scale librational motions (less than nano seconds), these include amide bond fluctuations and side chain rotations, respectively. The measured backbone and side chain relaxation times are related to their respective degree of motions through the *Lipari-Szabo* model-free order parameter (O^2) [18]. What does order parameter represent? Order parameter represents the degree of flexibility of a bond vector and takes

on values in the range of $[0, 1]$ where 0 represents complete flexibility and 1 represents a rigid bond vector in the molecular frame. In other words, the order parameter is the limiting value of the autocorrelation function of a bond vector's position with time. The area under figure 1.5 (τ_e) represents the time scale of fluctuations for the corresponding bond vector (referenced as internal correlation time). The order parameter values

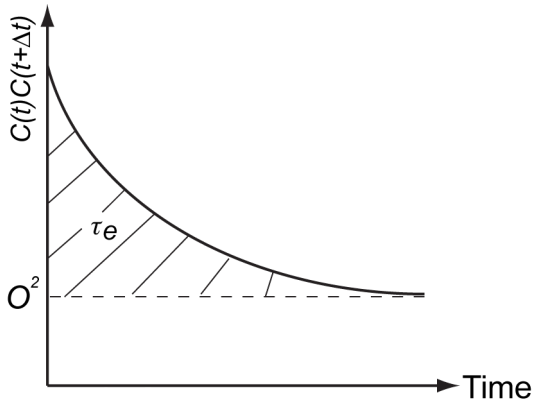


Figure 1.5: Time Correlation of bond vector fluctuations

extracted from backbone relaxation provide information on the molecular motions of the amide bond vector about its axis whereas the order parameter values extracted from other side chain relaxation experiments provide information on the molecular motions of the side chains about their respective axis. The order parameter value for any particular amide ^{15}N nuclei depends on its interaction with the surrounding nuclei. These interactions can be broadly classified into two categories:

- Dipolar Interactions (The interaction of magnetic dipole of different nuclei)

- Chemical Shift Anisotropy (The effect of the anisotropic electronic environment around nuclei)

For example, consider an amide ^{15}N nuclei, the dominant interactions affecting its relaxation are the dipolar interaction with the attached amide proton (hydrogen) and its (^{15}N) chemical shift anisotropy (CSA). The dipolar interaction from surrounding nuclei, (eg) ^{1}H, and others ... do indeed affect the relaxation properties. However, since dipolar interactions depend on the inverse 6^{th} power of the distance between the nuclei (r^{-6}), contributions from these surrounding nuclei are limited. Similar to dipolar interactions, chemical shift anisotropy is also dependent on the environment surrounding the nuclei of interest, in this case the amide ^{15}N. Depending on whether the nature of anisotropy is isotropic, axial or completely anisotropic, the effect on relaxation is different.
Hence, considering the appropriate effect of both dipolar interactions and chemical shift anisotropy influences the calculation of dynamic parameters, namely the order parameter (O^2), and internal correlation time (τ_e), from relaxation measurements [1, 18, 19].

I have developed a computer program (**pyRelxn2A**) which can incorporate all of these complex interactions and aid in extracting the dynamics information (i.e. O^2 and τ_e) from NMR-relaxation data. The analysis of all the relaxation data in this thesis were performed with **pyRelxn2A**.

1.2.3.1 Backbone Relaxation

Backbone relaxation measurements which comprise longitudinal relaxation (T_1) , transverse relaxation (T_2), and heteronuclear nuclear overhausser effect (NOE), together (Equation 1.9) provide a complete description of the different motions occurring in the main chain chain of proteins [20]. The relaxation of ^{15}N nuclei of the amide backbone is measured using ^{15}N relaxation of the ^{15}N amide nuclei in proteins.

$$\begin{aligned}
\frac{1}{T_1} &= \frac{1}{4}d^2\{J(\omega_H-\omega_N)+3J(\omega_N)+6J(\omega_H+\omega_N)\}+c^2J(\omega_N) \\
\frac{1}{T_2} &= \frac{1}{8}d^2\{J(\omega_H-\omega_N)+3J(\omega_N)+6J(\omega_H+\omega_N)+4J(0)+6J(\omega_H)\} \\
&+ \frac{c^2}{6}[4J(0)+3J(\omega_N)]+R_{ex} \\
NOE &= 1+\frac{T_1d^2}{4}.\frac{\gamma_H}{\gamma_N}\{6J(\omega_H+\omega_N)-J(\omega_H-\omega_N)\}
\end{aligned} \tag{1.9}$$

The spectral density function (J) in equation 1.9 corresponds to the probability of motions occurring at a given larmor precession frequency ω and is defined for isotropic motion as:

$$J(\omega)=\frac{2}{5}\left\{\frac{O^2\tau_m}{1+(\omega\tau_m)^2}+\frac{(1-O^2)\tau_f}{1+(\omega\tau_f)^2}\right\} \tag{1.10}$$

The relaxation times (T_1 and T_2) depend on a number of quantities, namely the larmor

frequencies of amide nitrogen (^{15}N, ω_N), amide proton (^1H_N, ω_H), spectral density function (J), molecular tumbling time (τ_m), τ_f (= $\frac{1}{\tau_m}$ + $\frac{1}{\tau_e}$) , internal correlation time or time scale of bond vector fluctuations (τ_e), and the order parameter (O^2).

The analysis of backbone relaxation prior to side chain relaxation is critical since an accurate estimate of molecular tumbling can be calculated through backbone relaxation. This is due to the fact that the main chain of most proteins, if not all, is predominantly rigid with flexible motions restricted to the unstructured regions of the proteins, namely loops, amino and carboxy termini. Depicted in figure 1.6 are the amide backbone order parameters (O^2_{NH}) for ubiquitin and it is evident that the structured regions, helix and beta-sheet, are predominantly rigid whereas the loops regions exhibit conformational flexibility. The calculation of molecular tumbling from the ^{15}N relaxation times

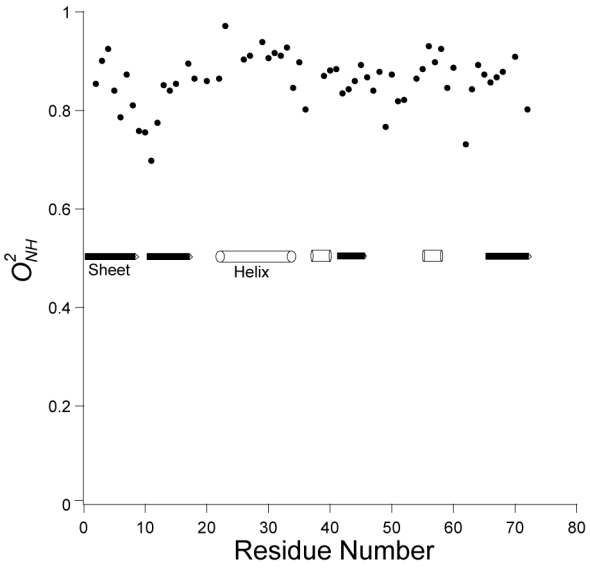


Figure 1.6: Backbone order parameter distribution in ubiquitin

is achieved using the in-house software (*pyRelxn 2A*) to fit for the functional forms of

Equation 1.9 and 1.10 and extract τ_m. The molecular tumbling can be isotropic, axially symmetric or completely anisotropic depending on the size and shape of the protein [21]. Therefore depending on the tumbling model used, the spectral density function changes and *pyRelxn 2A* can accordingly be used to appropriately accommodate for the modified functions to calculate the molecular tumbling time(s). The functional importance of understanding the molecular motions highlights changes in motions of backbone directly linked to the function of protein such as catalytic activity or structural transition resulting in motional changes [22, 23]. The calculated molecular tumbling model and time will need to be carried forward to analyse the relaxation of side chains in order to extract information pertaining to their molecular motions, i.e. side chain order parameter and their time scales of motion. In the next section, side chain relaxation of methyl-bearing amino acid side chain is described as these are the most well characterized and are also used extensively in this thesis. The theory described however, is applicable to other side chains namely aromatic rings with minor modifications. Backbone relaxation, methyl side chain relaxation, and aromatic side chain relaxation are the three relaxation methods used in this thesis.

1.2.3.2 Side chain Relaxation

Side chain relaxation methods probe the molecular motions of side chain nuclei of amino acids in proteins. The two main side chains that will be considered in this thesis are:

- Methyl side chain (Order Parameter: O^2_{axis})
- Aromatic side chain (Order Parameter: O^2_{aro})

Unlike the amide backbone of proteins, side chains have been observed to undergo a heterogeneous distribution of motion in the fast time scale ($\sim 10^{-12}$) by both molecular dynamics and NMR [17, 24]. How does NMR probe these motions? These fast time scale motions are probed similarly to that of the amide backbone, i.e., using relaxation methods (T_1 and $T_{1\rho}$). However, unlike the amide backbone where the relaxation of ^{15}N nuclei is measured, for side chains ^{13}C or ^{2}H relaxation is measured.

The relaxation of backbone amides involves an AX spin system, i.e., (A-X: $^{15}N -^{1} H_N$; $A{\equiv}^{15}N$ and $X{\equiv}^{1}H$) where as side chain relaxation can be AX (in the case of aromatic ring: $^{13}C -^{1} H$; $A{\equiv}^{13}C$ and $X{\equiv}^{1}H$) or AX_3 (in the case of methyl groups: $^{13}C^{1}H_3$). The relaxation properties of an AX_3 spin system is more complicated than an AX spin system hence the AX_3 spin system is usually reduced to an AXY_2 spin system where Y is deuterium ^{2}H nuclei which does not affect dipolar interactions of ^{13}C spins significantly. It is important to note here that methyl side chain relaxation can also be measured through deuterium relaxation, i.e., relaxation of ^{2}H nuclei in an AYX_2. However, deuterium relaxation is not detailed here since ^{13}C relaxation is used exclusively in this thesis. For a detailed review of these different methods refer to Igumenova et al [17].

The calculation of dynamics parameters (O^2 and τ_e) from the measured relaxation times (T_1 and $T_{1\rho}$) are similar to backbone amide relaxation wherein they involve Equation 1.9 with the larmor frequency of ^{15}N (ω_N) being replaced by the larmor frequency of ^{13}C

(ω_C). However, the main difference here being that the molecular tumbling time (τ_m) calculated from the amide backbone relaxation for the same protein at identical concentration is used. The side chain methyl order parameters calculated from Equation

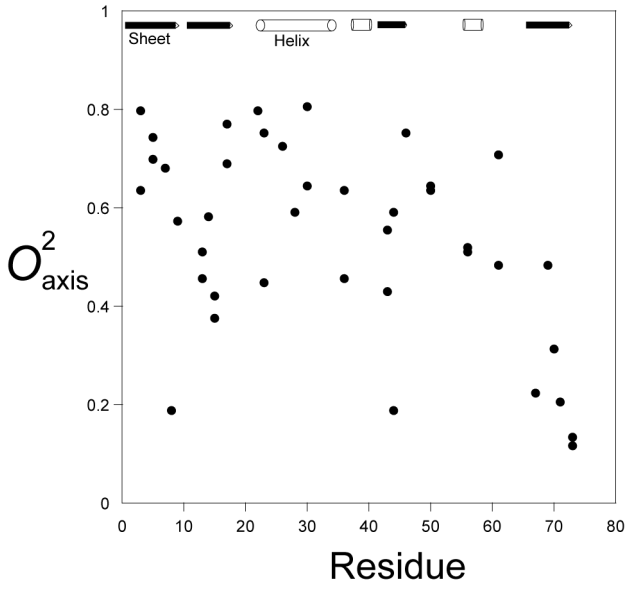


Figure 1.7: Methyl group order parameter distribution in ubiquitin

1.9 represent the motion of $^{13}C-^{1}H$ bond in the methyl group which can be related to the motions of the symmetry axis of the methyl group, i.e., $^{13}C_\beta-^{13}C_i$, where $^{13}C_i$ corresponds to the carbon in the AXY_2 spin system. This transformation is of practical significance as the order parameter (O^2_{axis}) corresponds to the motion of the entire methyl group rather than the motions of an individual $^{13}C-^{1}H$ bond. However, in the case of the aromatic side chain relaxation, the calculated order parameter represents the $^{13}C-^{1}H$ bond motion or in other words the motions of the ring about its symmetry axis. As pointed to earlier, the distribution of methyl order parameters such as those

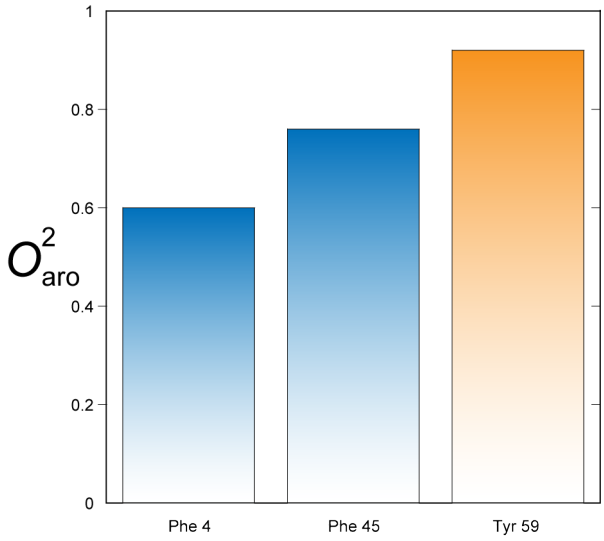


Figure 1.8: Order parameter distribution of aromatic ring side chain in calmodulin

seen in Figure 1.7 is heterogeneous [17, 25]. Aromatic side chains on the other hand have only recently been shown to possess such a heterogeneous range of dynamics and as such were previously thought to be quite rigid, similar to the backbone amides [26] (refer to Chapter 3).

The order parameters or dynamics of side chains combined with those of backbone aid in understanding how ligand binding affects protein function. The changes induced in the internal motions of proteins due to ligand binding contribute to the changes in conformational entropy. NMR-relaxation measurements of internal motions of protein side chains, namely methyl bearing amino acid side chains, have been used to provide quantitative estimates of changes in protein conformational entropy [13, 14]. Recent studies have shown that ligand binding modulates the conformational entropy of proteins thereby affecting its function/activity [27].

1.3 Conformational Entropy from Protein Dynamics: The History

One of the simplest definitions of entropy comes from the Boltzmann's principle which defines entropy as the measure of the number of possible states in a system at thermodynamic equilibrium. However, the Gibbs formulation (Equation 1.11) which is not too dissimilar to the Boltzmann formulation (Equation 1.12) defines the entropy in two levels, a macroscopic state and microscopic states. The macroscopic state of the system is defined by the number of different micro states visited by the system. In other words entropy is defined within the context of a distribution of micro states accessible for any particular macro state. The advantage of the Gibbs definition is that it is valid even for systems far away from equilibrium. Equation 1.11 and 1.12 define the entropy according to Boltzmann and Gibbs formulation, respectively.

$$S = -k_B ln\Omega \tag{1.11}$$

$$S = -k_B \sum_{i=1}^{n} p_i ln p_i \tag{1.12}$$

Ω represents the number of states accessible to the system whereas p_i represents the probability of the i^{th} state being visited when the system fluctuates.

Order parameters measured from NMR inherently contain information regarding the number of states visited, i.e., residues which are dynamic, visit more states and display a lower order parameter. The transformation of order parameter to entropy is essentially

reduced to extracting the number of states visited. The following set of equations relate the order parameter to entropy:

$$\begin{aligned} O^2 &= \iint d\Omega_1 d\Omega_2 p_{eq}(\Omega_1) p_{eq}(\Omega_2) P_2(\cos(\theta_{12})) \\ p_{eq} &\equiv \frac{\exp(\frac{-U(\Omega)}{kT})}{Z} \\ A &\equiv -kT\,lnZ \\ H &\equiv kT^2 \frac{\partial lnZ}{\partial T} \\ S &= klnZ + kT\frac{\partial lnZ}{\partial T} \end{aligned} \tag{1.13}$$

As evident from Equation 1.13, the calculation of the partition function (Z) from order parameter is required for the eventual transformation to entropy. The partition function (Z) is however dependent on the free energy (A) of the system and hence the potential $U(\Omega)$. To this end, much research effort has been directed at examining the effect of different energy potentials [15, 16]. One of the simplest potentials examined is that of the harmonic oscillator where the potential take the form:

$$U = \frac{1}{2}k\theta^2 \tag{1.14}$$

A detailed analysis of the the nature and characteristics of the different potential will not be presented here but it suffices to say that the different potentials reveal interesting phenomenon about the nature of side chain dynamics [15, 16]. However, they also

have shortcomings which are a result of the inherent assumptions associated with the form of the potential ($U(\Omega)$). For example, the harmonic oscillator model among many things does not account for correlated motions of side chains. In the densely packed environment of proteins, motions of side chains are highly correlated [17].

1.3.1 Model Dependent Approach: Harmonic Oscillator

With the above physical basis for the relationship between order parameter and entropy, Wand & co-workers have examined the effect of protein conformational entropy (if any) on the binding of ligands, small peptides, to the protein, Calmodulin [13]. Armed with measurements of methyl side chain dynamics of Calmodulin bound to several different peptide ligands, the protein conformational entropy was calculated (Equation 1.15).

$$S_{conf} = \sum p_i S_i^h - k_B \sum p_i ln p_i \tag{1.15}$$

The entropy here (Equation 1.15) was modelled by a multidimensional harmonic well. The first term (S_i^h) represents the entropy due to the fast intra-well motion or fluctuations about a particular conformation. The second term represents the classical Boltzmann entropy, corresponding to the number of different conformations. The potential used here corresponds to that of a simple harmonic oscillator (Equation 1.14). One acute shortcoming of this methodology is that the entropy calculated is dependent on

the potential energy function, but this shortcoming was addressed when evaluating the changes in conformational entropy (ΔS_{conf}) since the changes in order parameter (ΔO^2) are insensitive to the model used.

There was an excellent correlation between the changes in protein conformational entropy calculated using the above model and the changes in total binding entropy for the calmodulin:peptide complexes [13]. This was an unprecedented result as it stated that not only is conformational entropy important, its contribution could be of the same order of magnitude as that of the total binding entropy.

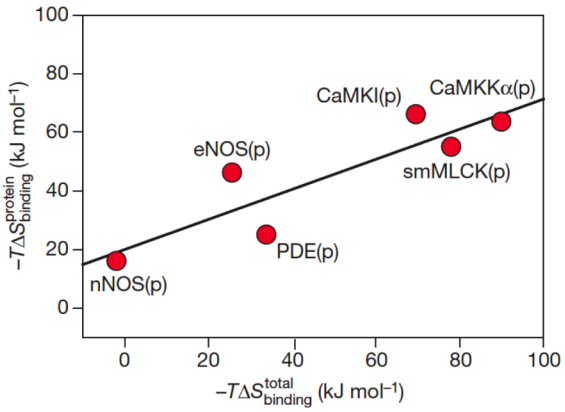


Figure 1.9: Correlation of change in protein conformational entropy with change in total binding entropy for Calmodulin

1.3.2 Model Independent Approach: Empirical Calibration

One of the main drawbacks of the model dependent approach as mentioned earlier is the inherent assumption of the form of potential used as well as the fact that the effect of correlation motion is neglected in simple potentials such as the harmonic oscillator. In

order to overcome this limitation, Wand & co-workers proposed an empirical calibration of the conformational entropy through the measured methyl side chain order parameters [14]. This is inherently a model dependent approach where in a linear relationship is assumed between the order parameter and the conformational entropy. The use of a linear relationship is the simplest possible and results in the least number of external variables needed to consider. The empirical relationship used could be written as follows:

$$\Delta S_{conf} = m[(n^{protein} \bullet < \Delta O^2_{axis} >^{protein} + n^{ligand} \bullet < \Delta O^2_{axis} >^{ligand})] + \Delta S_{other} \tag{1.16}$$

where 'm' corresponds to the empirical calibration slope and 'n' corresponds to the number of amino acids in the molecule. The conformational entropy of proteins calculated using Equation 1.16 correlated well to that calculated from the total binding entropy after accounting for the change in solvent entropy of the protein.

$$\Delta S_{total} = \Delta S_{solvent} + \Delta S^{protein}_{conf} + \Delta S^{ligand}_{conf} + \Delta S_{other} \tag{1.17}$$

This suggested that protein conformational entropy can indeed be calculated accurately from the NMR measured side chain dynamics albeit with some constraints. These constraints being that there are enough methyl probes to accurately represent the entire protein, i.e. good coverage, and that the assumption of methyl group dynamics are correlated to the motions of the rest of the protein. I have addressed several of these

issues in this thesis (refer to Chapter 2).

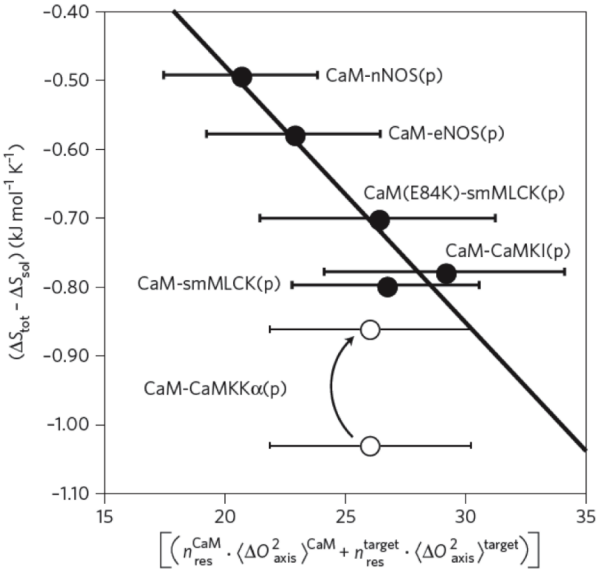


Figure 1.10: Correlation of change in protein conformational entropy measured by NMR versus that calculated from calorimetric measurements for Calmodulin

1.3.3 Questions that need to be addressed

The measurement of protein conformational entropy presents a significant step forward in understanding the subtleties of protein thermodynamics. However, the overall significance of conformational entropy with respect to the contribution and influence over protein:ligand interactions remains unclear. Is there a global role played by conformational entropy which is applicable to all protein:ligand interactions irrespective of the nature of the ligand? Furthermore, are methyl groups dynamics sufficient to measure protein conformational entropy? What about protein:protein interfaces which are largely devoid of methyl probes? How valid is the assumption that methyl group motions in side

chains are correlated to the rest of the protein? How does protein conformational entropy changes contribute to protein-ligand complexes spanning low to high affinities? These are just some of the many questions that still need to be answered. The most important question which has so far not been addressed is whether protein conformational entropy changes can be included in the pharmaceutical drug design process?

In this thesis, I attempt to address these questions with the aid of several techniques, including NMR-relaxation and MD simulations. First, the assumptions of the empirical calibration model (Equation 1.16) will be addressed through the use of MD simulations of several different proteins. Some interesting new results which have emerged out of this MD simulation study will also be outlined in detail (refer to Chapter 2). Secondly, a new approach to supplement the measure of protein conformational entropy is presented (Chapter 3). This involves aromatic amino acid side chains. Aromatic amino acids are ubiquitous in protein:protein, protein:RNA and protein:small molecule interactions. Moreover, they are highly populated in the interfaces of these interactions [28]. However, they remain largely unexplored due to the several complications associated with their NMR-relaxation properties. I have developed a new isotope labelling method which overcomes these issues. The application of this scheme to different systems and the wealth of interesting information we have learnt will be outlined (Chapter 3 and 4). In Chapter 5, I have combined both methyl side chain dynamics and aromatic side chain dynamics to study the interaction of a very high affinity protein-protein interaction. The role played by protein conformational entropy in such very high-affinity interactions

is examined in detail. One of the fundamental assumptions associated with very high affinity interactions is scrutinized here with the methods developed. Finally, in order to address the generality of the interpretation of the 'Entropy Meter' formulation, I have combined calorimetric and NMR dynamics information for 35 different protein-ligand complexes to demonstrate that the 'Entropy Meter' is indeed universal (Chapter 7). The generalized entropy meter presented here is a significant step forward in utilizing protein conformational entropy changes as an additional dimension for rational design of pharmaceutical drugs.

Chapter 2

Insights into the microscopic origins of protein conformational entropy

2.1 Introduction

The forces that govern the stability of the native state of proteins are complex and difficult to unravel and dissect through experiment, simulation or theory [29, 30]. This is particularly the case for the residual conformational entropy of proteins. Protein conformational entropy is a central component of the thermodynamics governing protein function [11, 31, 32]. Historically, it has been impossible to experimental measure the

contribution of residual protein conformational entropy to fundamental protein activities, such as binding of ligands, which is vital for protein function. Recent progress has illuminated the possibility of employing NMR relaxation based methods to quantitatively determine the role of changes in conformational entropy in molecular recognition by proteins [24]. This approach rests on using the fast internal motion of proteins as a proxy for conformational entropy [17]. The initial approach which as introduced previously (Figure 1.9) was to effectively take an inventory of the change in motion at a limited number of sites and interpret this within the context of a simple physical model, such as the harmonic oscillator [16] or diffusion within an infinite square well [15]. This raised several obvious issues, such as the effects of correlated motion, the operation of a more complex potential energy function, the completeness of the oscillator count, and others [17]. More recently this issue was sidestepped by using an empirical calibration of the dynamical proxy for conformational entropy. Rather than attempt a model-dependent interpretation of an inventory of changes in local dynamics, an empirical scaling between the experimental measure of local dynamics and local disorder (entropy) was attempted (Figure 1.10) [14]. The idea rests of the assumption that the experimentally accessible methyl-bearing side chains are numerous enough to provide sufficient coverage of internal motions and are sufficiently coupled to the neighbouring side chains to report on the total entropy [24]. With an appropriate experimental system, this leads to a rather simple relationship between what can be measured (protein motion and total binding entropy) or confidently calculated (solvent entropy) and what is desired (conformational

entropy) [14]. Measuring the change in motion of methyl bearing side chains on the nanosecond time scale using classical NMR relaxation methods [33–37], it is possible to obtain a quantitative measure of the underlying change in conformational dynamics [14]. This methodology was first applied on calcium-saturated calmodulin to a series of calmodulin-binding domains [14].The results demonstrated that changes in conformational entropy were important to ligand binding and also confirmed that conformational entropy contributed to the tuning of free energy of binding in this system [13, 14]. Recently, a second example demonstrated by Tzeng & Kaladimos, where a collection of mutants of catabolite activator protein (CAP) was used to study the thermodynamic origins of high affinity binding od DNA to the cAMP-activated protein [27]. The advantage of the CAP system is that the same binding ligand (DNA) was used, the structure does not change appreciably upon binding the DNA, and a array of structurally benign mutations remote from the binding interface were available. This eliminated most of the uncertainties associated with the analysis of the calmodulin complexes described previously. The resulting empirical calibration of CAP was excellent and led to a convincing analysis of the role of conformational entropy in binding of ligands to the CAP protein and its potential role in allosteric regulation [27]. This was a remarkable result as it demonstrates that for two completely independent protein system binding to different ligands, the NMR measured conformational entropy is excellently correlated to that from ITC and that the magnitude of protein conformational entropy is of the same order as that of the enthalpic contribution to free energy. These studies establish that

the protein conformational entropy is a significant contributor to the free energy of ligand binding and may even play a vital role in tuning the affinity of proteins to different ligands. However, a key difference between the CAP study and the calmodulin study was that the degree of correlation between the NMR measured protein conformational and from ITC was different by two fold. This raised several questions pertaining to the validity of using methyl groups as reporters of global protein conformational entropy, whether such a relation is indeed system specific and if so, what are the consequences of same.

Here I attempt to understand the microscopic origins of the apparent success of the empirical "entropy meter" based on a dynamical proxy measured through NMR relaxation methods. Extensive MD simulations in explicit water are employed to examine the ability of measure of fast internal motions of methyl groups in proteins to adequately represent the conformational entropy fo proteins. In addition, MD simulations provide an excellent platform to analyse the contributions of correlated motions of side chains to the measure of protein conformational entropy. However, extending the results from MD simulations to experimental measurements impose the condition that the calculated variables are correlated well to the NMR relaxation measured observables, namely order parameter (Figure 1.5). Interestingly, despite significant improvements in computational methodlogies towards long-time scale simulations [38, 39], there has been very little benchmarking of MD simulations with experimental measured of methyl-bearing

side-chain motions derived from NMR relaxation [40]. With one exception [41], the correspondence between Lipari-Szabo (L-S) methyl group order parameter derived from MD and experiment has been poor and definitely insufficient to promote further studies dependent on such correspondence. Therefore, prior to dissecting the origins of the entropic contribution coming from methyl bearing side-chains, I have compared the L-S order parameters obtained from the MD simulations performed here, to those obtained via NMR deuterium or carbon relaxation based methods. The MD simulations performed here have been using NAMD [42, 43] with CHARMM27 force field [44, 45]. I find that the agreement with experiment is excellent, a marked improvement over the limited number of previous studies. This provides a solid platform from which results from MD simulations can be extended to experiment and would potentially aid in understanding the origins of the "entropy meter" established using NMR relaxation methods.

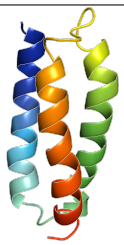
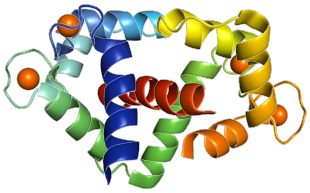
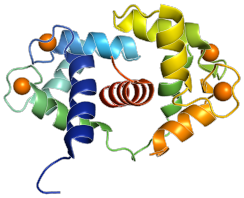
In summary, the results of the analysis of the simulations reveal that the experimentally accessible motions of methyl bearing side chains are coupled sufficiently to the motions of other side chains and hence serve as excellent reporters of conformational entropy. The total side chain conformational entropy can be accurately recapitulated using only measures of side-chain methyl motions provided there is a good distribution of methyl containing amino acids in the protein. Interesting, MD simulations reveal that the conformational entropy varies mostly through the changes in populations of the rotameric states rather than by variation in the effective potential defining each rotameric well.

This provides comprehensive evidence for the linear correlation between NMR measured conformational entropy versus the total binding entropy measured using ITC, for both the model-dependent approach and model-independent approach even though the model-dependent approach uses a potential, the conformational entropy contribution is predominantly from the changes in population of rotameric well. Finally, a statistical analysis reveals that the correlated motions are sufficiently limited to have limited impact on the relationship between the NMR measure of methyl dynamics and the derived entropy. These, in total, validate the use of conformational dynamics derived experimentally from NMR relaxation as a proxy from which to obtain quantitative estimates of changes in protein conformational entropy.

2.2 Results

MD simulations of several different proteins containing different structural motifs, ranging in size from 73 to 168 residues were selected for study (Table 2.1). Some of the proteins studied here also contain bound metals (calmodulin) or a prosthetic group (cytochrome c2). Two of the proteins are also high-affinity complexes of small peptides with calmodulin. All have extensive experimental reference sets for dynamics derived from site-resolved solution NMR relaxation studies. The methyl group L-S squared order parmaters (O^2_{axis}) were taken from published deuterium methyl relaxation studies: $\alpha_3 D$

[46], adipocyte lipid binding protein (ADBP) [47], cytochrome c2 (Cyt c2) [48], calcium-saturated calmodulin (CaM) complexes with peptides corresponding to the calmodulin-binding domain of the smooth muscle myosin light chain kinase (smMLCKp) and the neuronal nitric oxide synthase (nNOSp) [13], hen egg white lysozyme (HEWL) [49], and ubiquitin [50]. The average dynamical properties of the proteins range from quite rigid (e.g. cytochrome c2 and lysozyme) to quite internally dynamics (e.g. $\alpha_3 D$). The molecular coordinates for the proteins used in the simulations were derived from structures deposited in the Protein Data Bank (PDB): ubiquitin [51], $\alpha_3 D$ [52], calcium-saturated calmodulin-smMLCK [53], calcium-saturated calmodulin-nNOS (unpublished), hen egg white lysozyme [54], adipocyte lipid binding protein [55], and cytochrome c_2 [56] (Table 2.2).

Protein	Structure
Ubiquitin	
$\alpha_3 D$	
CaM-smMLCK	
CaM-nNOS	

Set up variables for MD simulations: The MD simulations of the seven proteins listed above in Table 2.2 were carried out with NAMD2 [43] using CHARMM27 [44] all-atom parameter set and TIP3P [57] water potential. Hydrogen atoms were added to the PDB structure using VMD [58]. Individual proteins were centred in a TIP3P water box such that protein atoms were at least 8Å from the boundary except in the case of two small proteins, ubiquitin and $\alpha_3 D$, where the solvent layer was at least 5Å and 6Å, respectively. Simulations were performed using a time step of 2 femto-seconds. Bonds to hydrogen atoms were constrained using the SHAKE algorithm [59]. A switching distance of 10Å and cutoff of 12Å were used for non-bonded interactions, combined with particle mesh Ewald summation with 1Å grid

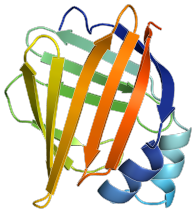
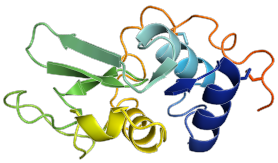
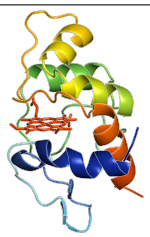
Protein	Structure
ADBP	
HEWL	
Cyt c2	

Table 2.1: Proteins used in MD simulations

Proteins	Residues	PDB code	Temperature (deg C)	waters	Length (ns)
Ubiquitin	76	1UBQ	20	4008	260
$\alpha_3 D$	73	2A3D	30	2532	160
CaM-smmLCK	167	1CDL	35	5228	1280
CaM-nNOS	168	2O60	35	4957	1120
ADBP	131	1LIB	20	4008	112
HEWL	129	1LZA	35	4133	240
Cyt c2	116	1C2R	30	5116	120

Table 2.2: Characteristics of the protein set used for MD simulations

spacing for long-range electrostatics. Simulations were run at constant temperature and pressure of 1 atm, controlled using extended Langevin thermostat. Simulation temperatures corresponded to those at which NMR relaxation experiments were

performed for each protein (Table 2.2). Following equilibration runs of atleast 1 ns, several 60 ns simulations were performed starting from the final coordinates of the earlier run but with different initial velocities. For three protein systems, ubiquitin, calmodulin-smMLCKp and calmodulin-nNOSp complexes, longer simulations were also run on the Anton supercomputer using the same force field and simulation conditions except for a non-bond cut off of 14Å.

2.2.1 Simulation versus Experiment: O^2_{axis}

As mentioned earlier, establishing the validity of the MD simulations is crucial for interpreting the conclusions and results derived from the simulations in the context of experimental data i.e. order parameters (O^2_{axis}) corresponding to the fast methyl side-chain motions. The first step towards this is to calculate the order parameters from MD simulations and compare it to those from NMR-relaxation data for each of the proteins. Briefly, the time scale of motions in proteins spans a wide range of time scales. The global macromolecular tumbling motions are in the nanosecond time scale, whereas the time scale of fast internal motions of methyl groups or other side chains are predominantly in the pico-second time scale. Due to this superposition of multiple time scale motions on every side-chain in proteins, the contributions from macromolecular tumbling are separated from internal motions to get accurate estimates or rather true measure of fast internal motions of side-chains. Hence, the interpretation of the NMR-relaxation data using L-S [18] involves the simple assumption that internal motions and macromolecular tumbling motions can be separated. This is accomplished by treating the the time scales of motions being atleast an order of magnitude different, these

motions can be separated or in other words treated independently. This leads to the calculation of macromolecular tumbling time from the main chain or proteins as these are predominantly rigid and are less sensitive to the effective correlation time (τ_f (= $\frac{1}{\tau_m}$ + $\frac{1}{\tau_e}$) composed of both macromolecular tumbling (τ_m) and the time scale of fast internal motion of side-chains (τ_e).

Order Parameters from MD simulation: The calculation of order parameter from MD simulations similarly involve extracting out the contributions from molecular tumbling. Therefore, L-S [18] squared generalized order parameters were calculated from MD simulations by overlaying snapshots of the protein from the trajectories using a standard rigid-body alignment ($C\alpha$) procedure. For each snapshot in the MD simulation, the unit vector along the methyl symmetry axis (C_{i-1} - C_i^{methyl}) was obtained in terms of its vector components in Cartesian axes x,y, and z. The O^2 parameter for a given methyl is then calculated using [60]:

$$O^2 = \frac{3}{2}\left\{< x^2 >^2 + < y^2 >^2 + < z^2 >^2 + 2 < xy >^2 + 2 < xz >^2 + 2 < zy >^2\right\} - \frac{1}{2} \quad (2.1)$$

where <> indicates the average over the trajectory. For comparison with experiment, the order parameter of the symmetry axis was derived from the experimentally measured order parameter for the methyl C-H bond by assuming free rotation around the symmetry axis and ideal tetrahedral geometry (θ = 109.5°) by [61]

$$O^2_{obs} = O^2_{rot}.O^2_{axis} = \frac{(3\cos^2\theta - 1)^2}{4}.O^2_{axis} = 0.1108.O^2_{axis} \quad (2.2)$$

Order Parameters from independent simulations on the same protein were first computed using equation 2.1, averaged and then compared to those obtained experimentally via NMR relaxation.

Prior to the calculation of order parameter, the convergence of the time correlation function ($C(t) = P_2(< \mu(t+T).\mu(t) >)$) was checked for select methyl groups spanning the range of observed order parameter values and by the variation in order parameter

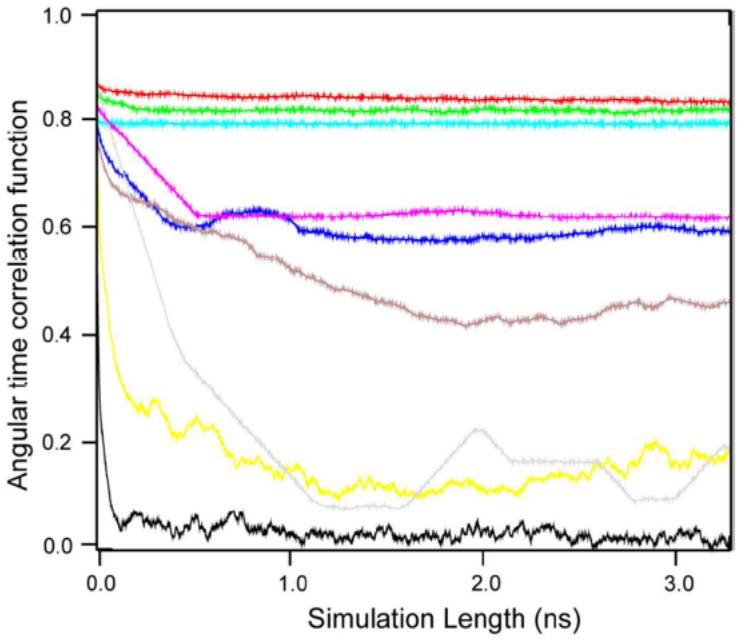


Figure 2.1: Convergence of methyl side-chain order parameters from MD simulations. The angular time correlation functions ($C(t) = P_2(< \mu(t+T).\mu(t) >)$), where $\mu(t)$ is the instantaneous unit axis vector and $P_2(x)$ is the second degree Legendre polynomial, is shown for a variety of methyl residues from highly dynamic to highly rigid. The convergence of the time correlation function indicates that adequate sampling is achieved even at short simulation time scales. The order parameter calculated from the plateau value of the above correlation functions correlate excellently to the average order parameter from the entire simulation length, using equation 2.1. The different residues of $\alpha_3 D$ shown above correspond to: (Red) -16 THR, (Green) - 21 LEU, (Blue) - 42 LEU, (Yellow) - 56 LEU, and (Black) - 1 MET. The different residues of HEWL shown above are: (Brown) - 44 MET, (Grey) - 43 THR, (Cyan) - 56 LEU, and (Magenta) - 58 ILE

calculated from different batches of individual trajectories (Figure 2.1). The methyl-by-methyl comparison for each protein revealed considerable site-to-site variation between the measured and simulated L-S order parameters with both over and underestimation being evident (Table 2.3 and Figure 2.3).

Summarized in Table 2.3 is the statistical comparison of the methyl order parameters

obtained from simulation and by experiment. It is clear that with the exception of cytochrome c2, the Pearson correlation coefficients (R^2) are all above 0.5, with the highest value of 0.85 for ubiquitin. It is important to stress here that such a high degree of correlation has not been observed before for such a wide range of proteins, especially for the side-chain methyl groups. This result presents itself as the first such benchmark for the comparison of methyl order parameters from simulation and experiment.

Proteins	**Expt** $< O^2_{axis} >$	**MD** $< O^2_{axis} >$	**R**^{2}	**Degree of correlation**
Ubiquitin	0.664	0.629	0.85	0.86
$\alpha_3 D$	0.451	0.571	0.76	1.25
CaM-smmLCKp	0.583	0.562	0.62	0.83
CaM-nNOSp	0.534	0.560	0.50	0.77
ADBP	0.633	0.619	0.75	0.88
HEWL	0.713	0.699	0.64	0.67
Cyt c2	0.767	0.670	0.47	0.68
Average	**0.620±0.11**	**0.616±0.05**	**0.65±0.014**	**0.85±0.013**

Table 2.3: Comparison of MD simulations and Experimental O^2_{axis} statistics

Even though considerable site-to-site variation (Figure 2.3) is observed for each of the

seven proteins, the overall average correlation of the O^2_{axis} values between the MD simulations and the NMR experiments is excellent (Figure 2.2). This proves to be very important in the context of entropy calculations since calculation of the total entropy for the protein involves contributions from all the side-chains.

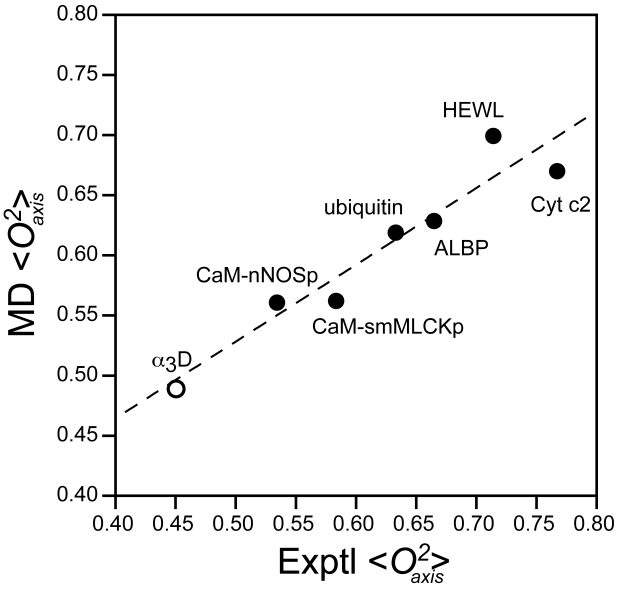


Figure 2.2: Comparison of the $< O^2_{axis} >$ values obtained from MD simulations versus those measured from NMR relaxation experiments for each of the seven proteins. The experimental average includes all available data. With the exception of $\alpha_3 D$, the MD average includes all the methyl groups as well. In the case of $\alpha_3 D$, the experimentally accessible sites are compared directly with the MD average of those respective sites to avoid an apparent artefact of limited experimental sampling. An excellent overall correlation is observed with a Pearson correlation (R^2) of 0.92 and a slope of 0.64±0.09 with an intercept of 0.21±0.05. Forcing the best fit line through the origin yields a slope of 0.92 with an almost similar Pearson correlation coefficient.

2.2.2 Rotamer Entropy from MD and O^2_{axis}

One of the fundamental aspects of the empirical "entropy meter" approach is the requirement for the correlation of the motional averaging of the methyl group symmetry

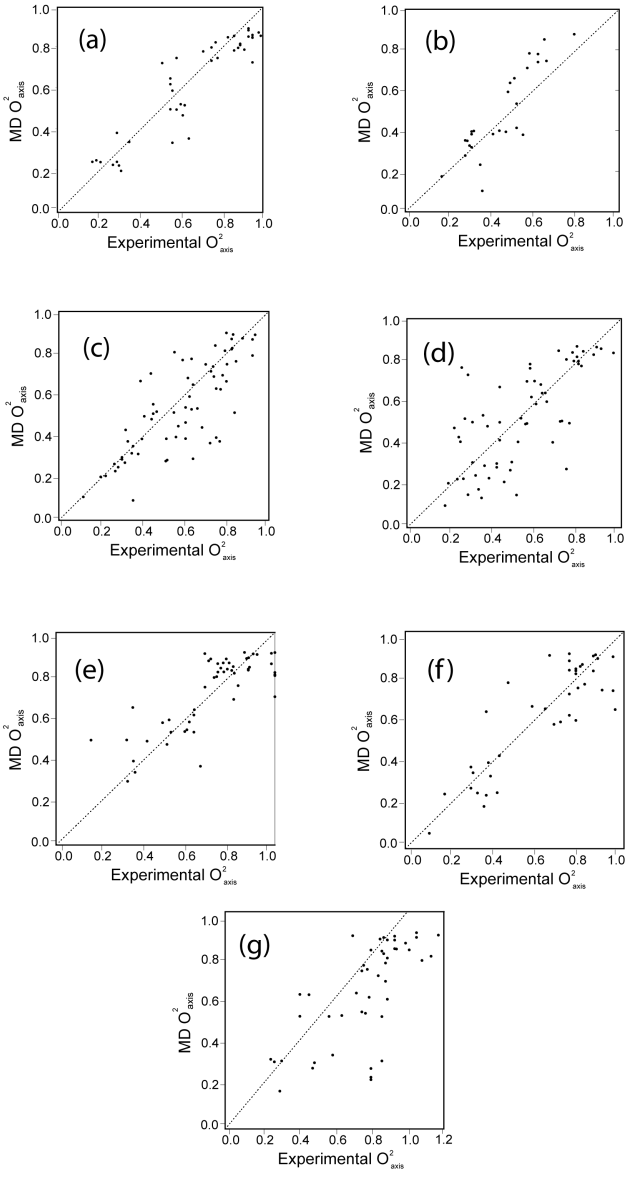


Figure 2.3: Comparison of the O^2_{axis} values obtained from MD simulations versus those measured from NMR relaxation experiments for each of the seven proteins. Considerable site-to-site variation is observed for all seven of the proteins: (A) Ubiquitin ($< r.m.s.d. >= 0.104$), (B) $\alpha_3 D$ ($< r.m.s.d. >= 0.116$), (C) Calmodulin-smMLCKp ($< r.m.s.d. >= 0.146$), (D) Calmodulin-nNOSp = ($< r.m.s.d. >= 0.172$), (E) HEWL ($< r.m.s.d. >= 0.124$), (F) ADBP ($< r.m.s.d. >= 0.125$), (G) Cytochrome c2 ($< r.m.s.d. >= 0.223$). The statistics for the best fit line are reported in Table 2.3.

axis ($< O^2_{axis} >$) and the underlying conformational entropy [14]. The MD simulation trajectories of the seven proteins were utilized in order to examine the validity of this requirement as well as this origins of the observed experimental correlation. The rotamer population distribution of all the methyl containing amino acids in all seven proteins were calculated from their respective simulations. The corresponding side-chain rotamer entropies were calculated as well.

Calculation of the Rotameric Entropy from MD simulations: Side chain rotamer (χ) angles for each amino acid were calculated from the MD trajectories and binned into one of three conformational states based on the dihedral angle: $gauche_+$ - $[0°, 180°]$, trans - $[120°, 240°]$, $gauce_-$ - $[240°, 360°]$. The rotamer probability distribution function (pdf) of all 3^{N_χ} rotamers of each side chain was constructed from the χ angle histograms, where N_χ is the number of side-chain χ angles. Only unique side-chain torsion angles were considered (e.g. the isopropyl group contributes to one giving Leucine 2 and Valine 1 side-chain torsion angles). The side chain rotamer entropy is then calculated as:

$$\frac{S}{k_B} = -\sum_i p_i ln p_i \tag{2.3}$$

where k_B is the Boltzmann constant, p_i is the probability of the i^{th} rotamer. $S_b = \frac{S}{k_B}$ is the entropy given in dimensionless units, and the normalized entropy is then defined as $\frac{S_b}{N_\chi}$.

The rotamer entropy of the different side chains calculated are normalized by their respective number of side chain torsion angles as the rotamer entropy represented by the motions of methyl groups are specific to the corresponding methyl containing side chain and that different methyl containing side chains contain different number of side chain torsion angles. The calculated normalized rotamer entropies for each methyl

group from each of the seven proteins when plotted against the corresponding calculated methyl L-S squared order parameters reveal an excellent linear correlation (Figure 2.4). More importantly, the distributions from the different proteins overlay each other, strongly suggesting that the empirical scaling of motion of the corresponding entropy should be universal.

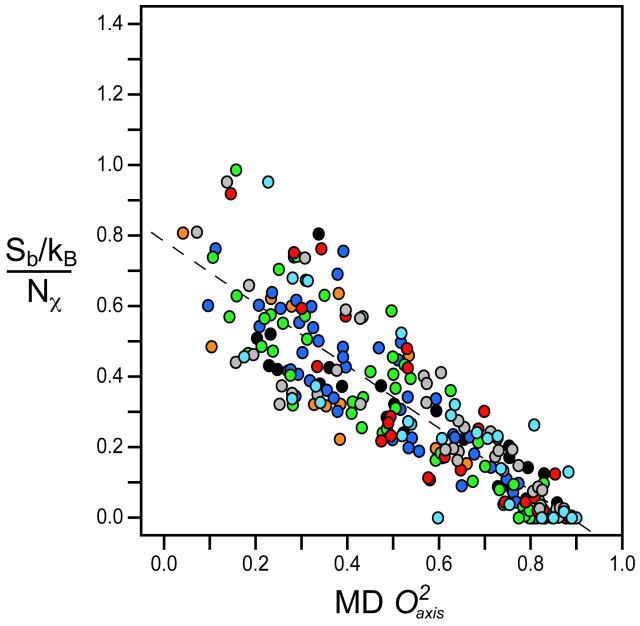


Figure 2.4: Correlation of the methyl rotamer entropy versus O^2_{axis} from MD simulations. Normalized entropy (Equation 2.2) is used for every methyl side-chain probe. The correlation was highly linear (R^2 = 0.77) with a slope of -0.88±0.03 and an intercept of 0.78±0.02. The different proteins are represented as follows (all solid circles): (black) ubiquitin, (orange) $\alpha_3 D$, (blue) calmodulin-smMLCKp, (turquoise) calmodulin-nNOSp, (red) HEWL, (light green) ADBP, (gray) cytochrome c2

Correlations of rotamer entropy and O^2_{axis} parameters for the individual proteins are shown below in Figure 2.5. Each of the seven proteins show an excellent linear correlation. It is important to note here that the rotamer entropy calculated here is for the

entire side-chain of the methyl containing amino acid while the order parameter reflects the motions of the methyl group about its symmetry axis, about the terminal rotamer χ angle. Even though the motions of methyl group occur about terminal χ angle, there could still be an influence of fluctuations about χ angles further away from the methyl group or in other words from the ones closer to the main chain, though this contribution is expected to be low. The linear correlation observed here indicates that the motions of the methyl group are indeed sensitive to the entire side-chain as expected. This further provides evidence to one of the aspects of the empirical "entropy meter" approach where this is an inherent assumption.

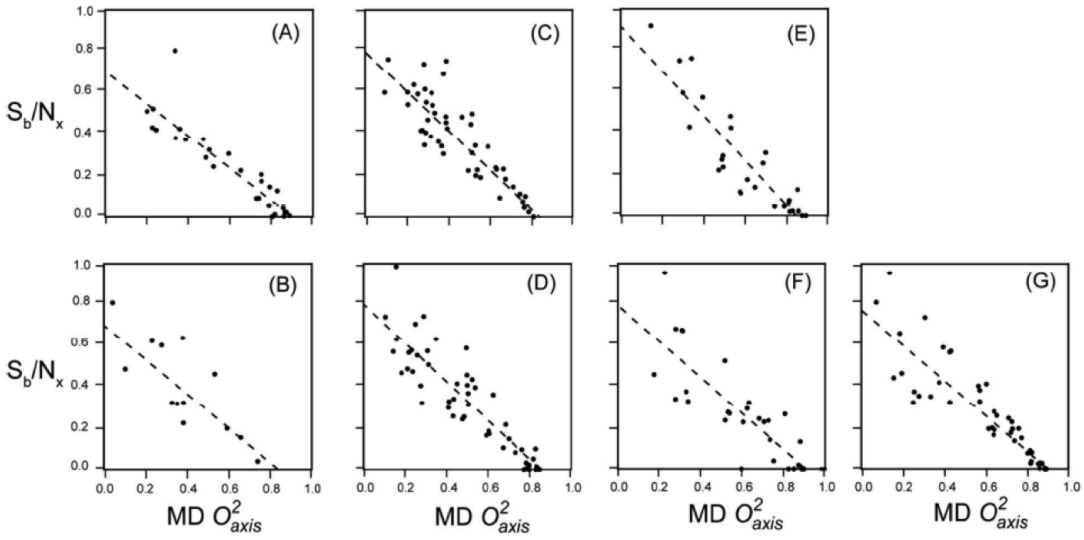


Figure 2.5: Correlation of the rotamer entropy normalized by the number of side-chain torsion angles to O^2_{axis} derived from MD simulations for methyl residues. Each of the panel (a-g) above represent the different proteins used: (A) Ubiquitin, (B) $\alpha_3 D$, (C) Calmodulin-smMLCKp, (D) Calmodulin-nNOSp, (E) HEWL, (F) ADBP, (G) Cytochrome c2.

The parameters of the best fit line are summarized in Table 2.4.

Proteins	Slope	R^2	$< \frac{S_b}{N_\chi} >$
Ubiquitin	-0.78	0.82	0.236
$\alpha_3 D$	-0.88	0.64	0.399
CaM-smmLCKp	-0.92	0.74	0.381
CaM-nNOSp	-0.95	0.80	0.335
ADBP	-0.84	0.79	0.276
HEWL	-1.10	0.81	0.257
Cyt c2	-0.88	0.68	0.382

Table 2.4: Correlation of normalized entropy (S_b) to MD O^2_{axis}

The excellent correlation of normalized rotamer entropy versus methyl group order parameter seen for the range of O^2_{axis} in $[0, 1]$ is intriguing as methyl groups with low order parameter i.e. flexible contribute to entropy in the same manner as groups which are rigid. In order to understand this correlation, it becomes important to dissect the different contributions to rotamer entropy. In principle, the contributions to rotamer entropy could not only arise from the distribution of populations between rotamer wells but also from the distribution within each rotamer well and their associated vibrational entropy [11, 62, 63] (Figure 2.6).

For each methyl group the fine distribution of torsion angles using a $3°$ bin size resolution was analysed to understand the extent of contributions to rotamer entropy. The width of the rotameric well, characterized by the root-mean-square-deviation (r.m.s.d) of its respective χ angles, weighted by their occupancy, provides an intuitive idea regarding the within well contributions. This mean well width as a function of O^2_{axis} hints at the variation in the contribution of within rotamer well fluctuations as a function of the extent of methyl side-chain motion (Figure 2.7).

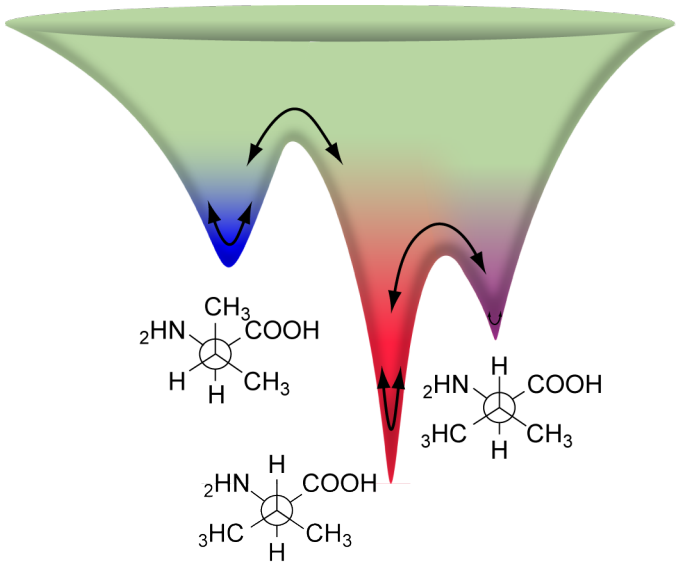


Figure 2.6: Illustrated here is a representative rotamer landscape for methyl containing side-chains. The arrows represent fluctuations between rotamer wells as well as within rotamer wells, both of which contribute to the rotamer entropy.

The within-well distribution of non-methionine rotamer angles is observed to be largely insensitive to the degree of fluctuations of the methyl groups, i.e. $O^2_{axis} < \sim 0.8$. However, at higher O^2_{axis} (> 0.8) some narrowing of the within-well distribution is observed. This suggests that when the methyl groups are very rigid, the extent of contribution from within-well fluctuations is different than when methyl groups are flexible. Methionine methyl groups, intriguingly, behave differently. The within-well widths of methionines from the different proteins have modest correlations with their corresponding methyl group fluctuations (R^2 = 0.30). This suggests that the energy well governing the directly connected torsion angle of the methionine methyl groups is somewhat variable. The rarity of methionine residues make it difficult to generalize any outcomes regarding this variable energy well and further exhaustive study of methionine residues is needed to

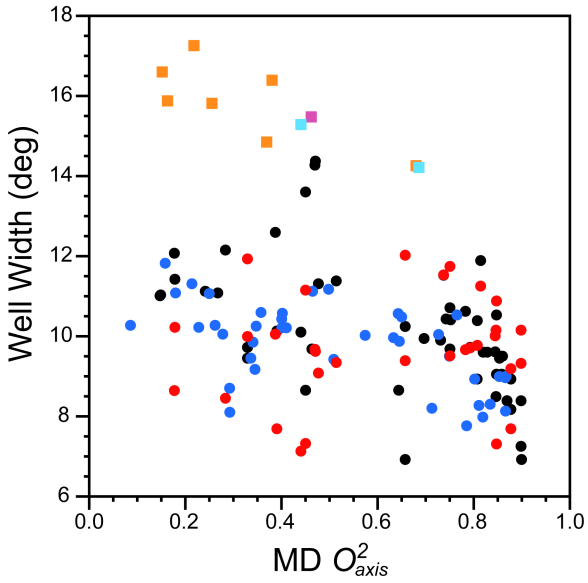


Figure 2.7: Correlation of within-well width distribution of side-chain torsion angles with the directly connected methyl group symmetry axis L-S squared order parameters calculated from MD simulations. The width of the distribution is given by the r.m.s.d protein and residue types indicated by solid circles (black) ubiquitin - (Leu, Ile, val, Thr), (blue) calmodulin-smMLCKp - (Leu, Ile, Val, Thr), and (red) HEWL - (Ile, Leu, Val, Thr). Methionines are indicated by solid square symbols (orange) from calmodulin complexes, (pink) from ubiquitin, and (turquoise) from HEWL.

address the same.

The excellent correlation observed for the rotamer entropy and the L-S squared order parameter provides sufficient evidence that methyl group serve as excellent reporters of conformational entropy of methyl containing amino acids in proteins. Also, the correlations being the same for seven different proteins, suggests an universal relationship between rotamer entropy and O^2_{axis}. Finally, the nature of the contribution of methyl group fluctuations to entropy seem to arise primarily from transitions between rotamer wells rather than fluctuations within a rotamer well.

2.2.3 Total protein conformational entropy

One of the critical aspects of the "entropy meter" formulation is that methyl groups not only serve as reporters of their own conformational entropy but that they are coupled sufficiently to their surroundings that they report on the whole protein as well [14]. The excellent correlation observed in Figure 2.4 suggests that methyl groups are indeed excellent reporters of the conformational entropy of their respective methyl containing amino acid. This however does not yet answer whether they can serve as reporters of the total conformational entropy of the protein. Prior to addressing this issue, it is important to keep in mind that internal motions in proteins are highly correlated owing to their dense packing. Therefore, any measure of conformational entropy must account for such correlations as their absence will lead to an overestimate. This problem is tackled in two steps, first, the total protein conformational entropy arising from rotamer fluctuations of all side-chains, methyls plus non-methyls was calculated. Secondly, the effect of both inter-residue correlations as well as intra-residue correlations are quantified using maximum information spanning tree (MIST) [64, 65] method and these are subsequently used to correct the values of the total protein conformational entropy.

Quantifying the effect of correlated motions on entropy using MIST: In order to quantify the effect of correlated motions of side-chains on their entropy, one- , two-, and three-dimensional (1D, 2D, 3D) population distribution functions (pdfs) for every side-chain torsion angle, doublet, or triplet combinations, respectively, were constructed from the trajectories. The torsion angles were again binned into $gauche_+$, $gauche_-$, and trans configurations. The 1D (i.e. uncorrelated) entropy (S_{1D}) was obtained by summing $-P(\chi)lnP(\chi)$ over each 1D pdf, $P(\chi)$, and then summing over all residues in the protein. The total side-chain entropy including the effect

of correlated motions of side-chains, was then estimated using the maximum information spanning tree (MIST) method. A second-order MIST approximation was obtained by computing pairwise second-order mutual information terms (I_2) from the 1D and 2D pdfs as,

$$I_2(j,k) = \sum_{l=1}^{3}\sum_{m=1}^{3} P(\chi_l^j,\chi_m^k) ln\left\{\frac{P(\chi_l^j,\chi_m^k)}{P(\chi_l^j).P(\chi_m^k)}\right\} \tag{2.4}$$

where the 2D pdf $P(\chi_l^j,\chi_m^k)$ express the joint probability that side-chain torsion angles j, k are in the l^{th}, and m^{th} conformations respectively. $P(\chi_l^j)$ and $P(\chi_m^k)$ are the corresponding 1D pdfs. Given all the I_2s, the second order MIST expansion is then obtained by selecting a spanning set of I_2s to maximize:

$$I_2^{MIST} = \sum_{i=2}^{n} I_2(j,i) \tag{2.5}$$

where n is the total number of side-chain torsion angles in the protein. To construct this spanning set, for each i, the index j $\in [1,2,..i-1]$ which gives the largest $I_2(i,j)$. Similarly, a third order mutual information term (I_3) can be constructed as follows:

$$I_3(j,k) = \sum_{o=1}^{3}\sum_{l=1}^{3}\sum_{m=1}^{3} P(\chi_o^i,\chi_l^j,\chi_m^k) ln\left\{\frac{P(\chi_o^i,\chi_l^j,\chi_m^k)}{P(\chi_o^i,\chi_l^j).P(\chi_m^k)}\right\} \tag{2.6}$$

Equation 2.6 expresses the change in mutual information between i, and j by including the k^{th} torsion angle. A third-order MIST approximation is built up analogously by choosing a spanning set of I_3s to maximize:

$$I_3^{MIST} = I_2(1,2) + \sum_{k=3}^{n} I_3(i,j,k) \tag{2.7}$$

where for each k, the indices i, and j $\in [1,2,..k-1]$, which gives the largest $I_3(i,j,k)$. Then an upper bound estimate for the total entropy including correlations at order O can be formulated as [64]:

$$S_{MIST-O}^{total} = S_{1D} - I_O^{MIST}(O = 2,3) \tag{2.8}$$

Averaged over the entire protein, there is an excellent correlation between the dynamical proxy (O^2_{axis}) and the total protein conformational entropy (Figure 2.8) from methyl side-chains alone. This however is a direct result of the observation that methyl group fluctuations are correlated excellently to their respective amino acid entropy. More interestingly, the total protein conformational entropy arising for all the side-chains is also excellently correlated to the methyl group fluctuations (Figure 2.8). Thus averaged over a protein, the methyl order parameters report well on the total side-chain entropy. This can be succinctly expressed as a simple equation:

$$\frac{\sum S_b}{\sum N_\chi} = 0.91 - 0.74. < O^2_{axis} > \tag{2.9}$$

Equation 2.9 however does not contain any corrections arising from correlated motions which has the effect of lowering the entropy measure. However, this above equation, which has been calculated entirely from intra side-chain rotamer pdfs account for correlations within a particular side-chain but does not account for correlations between different side-chains. Upon calculation of the contributions from interresidue side-chain correlations, the degree of reduction in entropy for all the proteins are proportional (inset Figure 2.8).

The correction from third-order correlations account for (17%) is not significantly higher than the second order contributions (11%) indicating that the bulk of the correction due to all orders of correlation is captured by these lower order, calculable effects. Quantifying contributions arising from even higher order correlations becomes computational

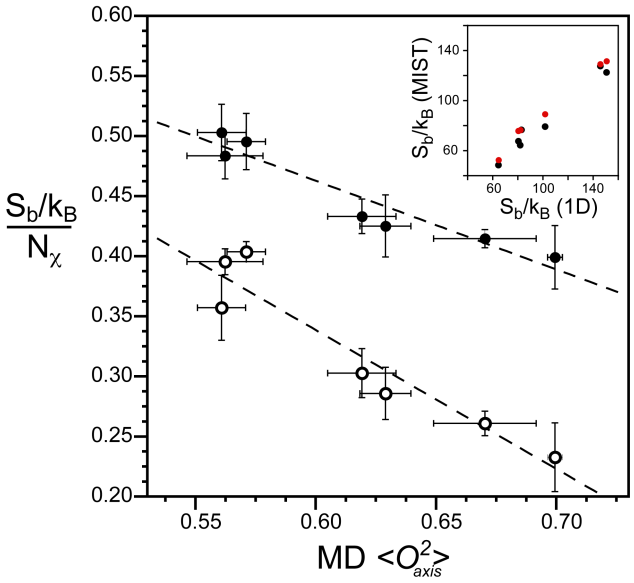


Figure 2.8: Dynamic proxy of methyl groups is an excellent reported of both methyl and total side-chain rotameric entropy. (open circle) The normalized methyl rotameric entropy for each protein is calculated as the summation of S_b for individual methyl-bearing amino acids divided by the number of associated rotamer χ angles. (black solid circle) The total rotameric entropy for each protein calculated as the summation over all residues normalized by the respective total number of rotamer χ angles. A very high linear correlation is observed for both methyl entropy (slope = -1.16±0.17, R^2 = 0.90) and total rotamer entropy (slope = -0.74±0.10, R^2 = 0.91). The inset shows the scatter plot of uncorrelated entropy with the entropy corrected for correlated motions using second-order (red solid circle) and third-order (black solid circle) MIST estimates of inter-residue correlations.

intractable. The effect of such correlations on equation 2.9 is such that the adjusted equation (equation 2.10) now represents an estimate of the total side-chain entropy of a protein, including both methyl- and non-methy containing residues, intra-, and inter residue correlation effects.

$$S^{total}_{side-chain} = 0.83 \sum N_\chi (0.91 - 0.74 < O^2_{axis} >) \tag{2.10}$$

From equation 2.10 it becomes possible to quantify the residual protein conformational entropy of a protein given its average dynamics property, in this case, the methyl L-S squared order parameter (O^2_{axis}).

2.3 Reformulation of the "Entropy Meter"

The original formulation of the "entropy meter" involving calmodulin-complexes raised several issues, some of which have been addressed in the previous section in this thesis [14, 24]. One of the most challenging issues raised involved the generality of the entropy meter formulation. In the previous section, MD simulations reveal that all seven proteins studied exhibited the same degree of correlation between the rotamer entropy of methyl side-chains and their corresponding methyl group fluctuations suggesting that this relationship indeed could be universal. However, recent experimental work by Tzeng & Kaladimos on the catabolite activator protein (CAP) binding to DNA (Figure 2.9) suggested that although the use of methyl group dynamics as a proxy for conformational entropy holds true, the degree of this correlation for different systems could be different.

In other words, the degree of correlation observed between the methyl group dynamics and the conformational entropy was different for calmodulin-complexes (slope = -0.008 kcal $mol^{-1}\ K^{-1}$) and CAP-complexes (slope = -0.002 kcal $mol^{-1}\ K^{-1}$) [24]. This discrepancy in the difference in calibration could either be due to the fact that different systems

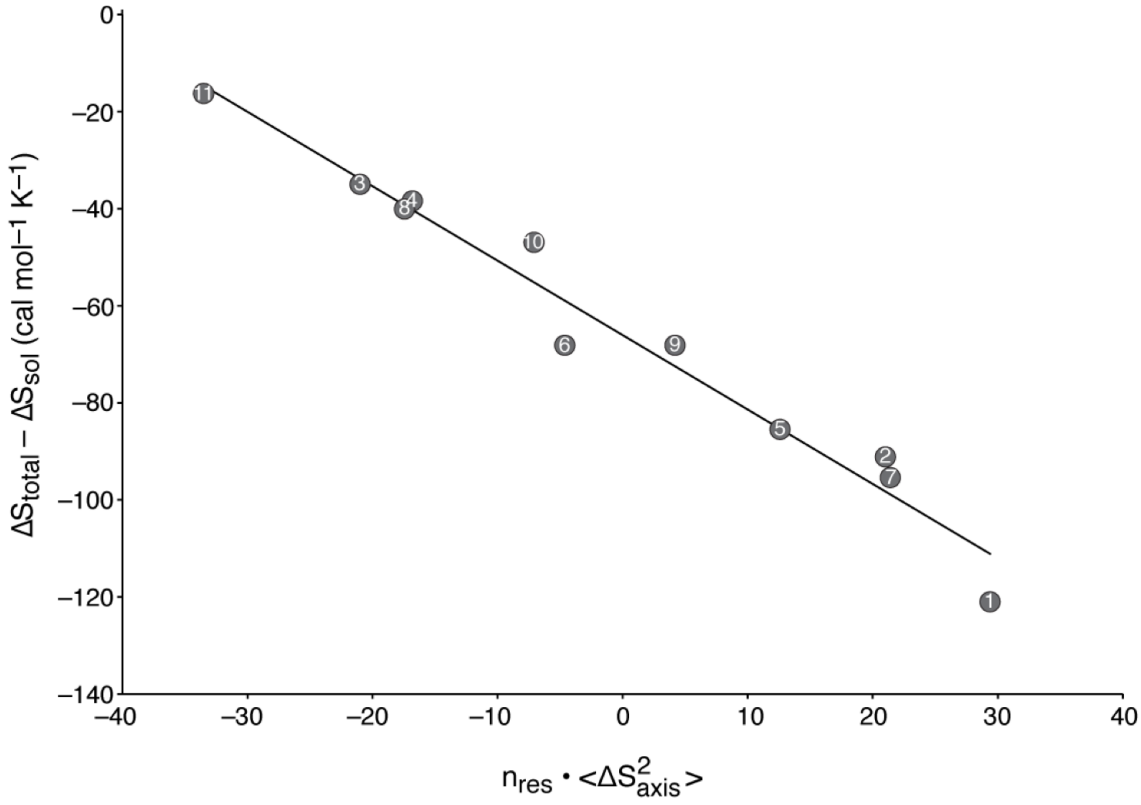


Figure 2.9: Empirical calibration of the NMR-measured methyl group fluctuations in CAP-DNA complexes to the protein conformational entropy assessed from ITC after adjusting for the solvent entropy contributions (Taken from [27]).

indeed exhibit different calibration. However, this does not agree with the MD simulations of the seven different proteins, all of which exhibit a similar degree of correlation. Secondly, this discrepancy could be due to the inherent differences in methyl group reporting on conformational entropy in the two systems. However, again the MD simulations suggest that the all seven proteins have excellent correlation of their total protein conformational entropy to their respective average methyl group fluctuations which indicate the methyl groups act as excellent proxy. *What is the source of this apparent*

discrepancy?

In the calibration performed for the calmodulin-complexes and the CAP-complexes a simple residue weighted scheme was utilized to project the entropy to the whole of protein molecule. This approach assumed that additional sources of entropy, such as rotational-translational entropy or undocumented entropy were constant across a family of protein-ligand interactions. The linear correlation assumed [15, 16, 66] for the empirical calibration is the simplest possible model with the least number of variables and one which results in least loss in information which is particularly prevalent in a model-dependent interpretations. These assumptions led to the original formulation of the "entropy meter" (equation 1.16) [14]. The computational results obtained here suggest a slight refinement of the empirical calibration. Due to the excellent correlation observed for the measures of angular disorder (O^2_{axis}) and the conformational entropy normalized by the number of involved torsion angles (Figure 2.8), the empirical calibration can now be reformulated with the use of torsion angle weighting instead of a simple residue weighted analysis (Equation 2.11).

$$\Delta S_{total} - \Delta S_{solvent} = N_{\chi}^{prot}.[< O^2_{axis} >^{prot}_{complex} - < O^2_{axis} >^{prot}_{free}] + N_{\chi}^{lig}.[< O^2_{axis} >^{lig}_{complex} - < O^2_{axis} >^{lig}_{free}] + \Delta S_{RT} \quad (2.11)$$

where N_{χ}^{prot} and N_{χ}^{lig} are the total number of side-chain torsion angles in the protein and ligand, respectively. Using equation 2.11 as the new empirical calibration formulation and applying it to both calmodulin complexes as well as CAP complexes yields a single linear calibration line resolving the discrepancy observed in the earlier reports (Figure 2.10). It is however important to note here that the changes in rotational-translational and other entropy contributions between the two protein systems is still assumed to be invariant.

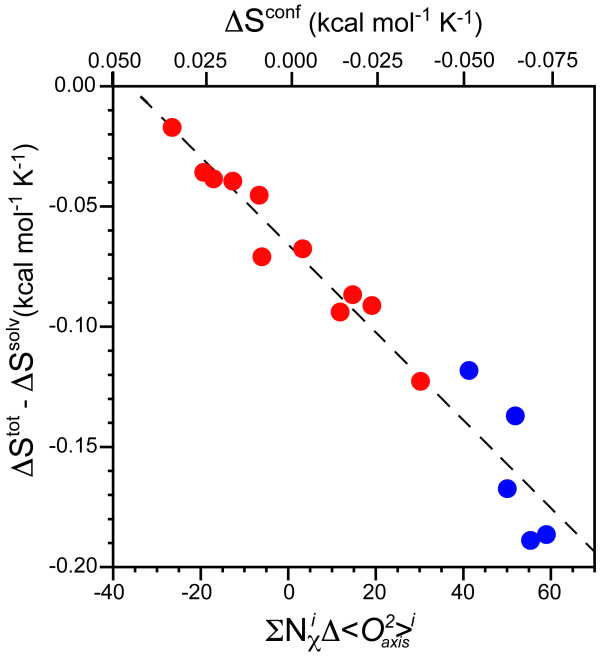


Figure 2.10: Calibration of the experimental dynamic proxy for protein conformational entropy. The points corresponding to both wild-type and mutants of CAP binding to DNA (red) [27] and CaM binding to its partners (blue) [14] follow the same empirical relationship when correlated using the average change in methyl order parameter weighted by the total number of torsion angles. An excellent linear correlation is observed (R^2 = 0.95) suggesting a common relationship between conformational entropy of proteins and their respective change in dynamics as represented by the methyl groups. A slope of -0.0018±0.0001 kcal mol^{-1} K^{-1} $\sum N_{\chi}^{-1}$ provides a convertion between the dynamical proxy and conformational entropy.

Given the linear relationship revealed by MD simulations, the experimental "entropy meter" should represent the true conformational entropy change as the effects of correlated motion are encapsulated within the empirical calibration. This is of course assuming the absence of significant protein-solvent coupling. The conformational entropy calculated from MD derived O^2_{axis} alone would not be accurate yet as this assumes that sufficient simulations have been performed and that these are reliable enough to be compared to experimental measurements. The empirical entropy meter calibration based on the experimentally measured methyl-group dynamical proxy however would be more accurate in relating the side-chain motion to the underlying conformational entropy. This is now essentially recapitulated in a remarkably simple relationship represented as follows:

$$\frac{\Delta S^{total}_{side-chain}}{k_B} = 0.92 \sum N_\chi \Delta < O^2_{axis} > \tag{2.12}$$

Equation 2.12 is derived from the degree of correlation or slope (-0.0018 kcal mol^{-1} K^{-1}) of the new entropy meter (Figure 2.10) converted to units of k_B.

2.4 Summary and Outlook

The original entropy meter formulation by Marlow and co workers [14] and the subsequent follow up by Tzeng & Kaladimos [27] raised several questions regarding the fundamental nature and origins of the entropy meter. Some of the key issues raised were:

- Can methyl group report on the entropy of their respective methyl containing amino acid?
- Can methyl groups report on the total conformational entropy of the protein?
- What is the origin of the methyl side-chain conformational entropy in proteins?
- Fundamental nature of the entropy meter & its generality to different protein systems?

Using MD simulations, these issues have been addressed in detail. Since the pioneering introduction of MD simulations for comprehensive studies of internal protein motion several decades ago [67], the capabilities and accuracy of such simulations have expanded greatly with increases in computational power and significant improvements in various aspects of the force fields employed [39]. Perhaps surprisingly, the accuracy of MD simulations has not been extensively tested in the ps-ns time regime against the sizable database of side chain order parameters obtained by solution NMR relaxation methods [38, 40]. Using the NAMD implementation of the CHARMM27 potential in a largely "out of the box" fashion, reasonable "site-to-site" agreement with experiment. These results compare favorably with evaluations of various AMBER potentials employing calbindin and ubiquitin as test proteins [68, 69]. MD simulations reveal that methyl groups are excellent reporters of the conformational entropy of their respective methyl side-chain (Figure 2.5 and Figure 2.6). This is perhaps not unexpected given

that the motion of terminal methyl groups arises from motion about all torsions connecting it to the backbone. Indeed, this time of motion is mirrored in the product of the corresponding generalized order parameters [18, 70]. This indicates that in the context of the total conformational entropy of proteins, methyl groups serve as excellent reporters of not only their own but also that of the total conformational entropy of proteins owing to their excellent coupling with nearby side-chains (Figure 2.9). The origins of the reported side-chain rotamer entropy is also dissected into its different contributions and MD simulations reveal that population changes due to rotamer transitions are the dominant contributors where as fluctuations within a rotameric well are not. These result suggested an alteration of the formulation of entropy meter. This reformulation of the entropy meter resulted in reconciling different calibrations observed for calmodulin bound to its different peptide targets versus that for catabolite activator protein and mutants of catabolite activator protein bound to DNA. Further evidence for this generality comes from a recent and parallel analysis of apo HIV protease simulations [71]. Glass et al. [71] also find a linear dependence of methyl group entropy on O^2 over the range 0.1 - 0.8, with a slope, when expressed in dimensionless units, very similar to that of Figure 2.4. It is important to note in this regard that the motions detected by the NMR relaxation discussed here are restricted to time scales faster than overall tumbling of the macromolecule [18], which is on the order of 10 ns or less for proteins examined in this study. Interestingly, the MD simulations exceed this by an order of magnitude. Thus, the high correlation suggests that the ensemble of states that is experimentally

accessible accurately reflects that sampled on longer time scales. Of note here is that all methyl-bearing side chains, except for threonine, show extensive rotamer interconversions during the simulation. In principle, it is possible that longer time scale motion may interconvert states that are not similarly sampled in the time regime relevant to the generalized order parameter obtained by classical NMR relaxation [72]. Moreover, the calculation of the effect of correlated motion on the quantification of conformational entropy by MIST algorithm [64] indicate that only a small linear scaling of the uncorrected entropy is required. This picture of a rather restricted local effects of correlation is consistent with other recent studies of conformational entropy in proteins [73–75].

One of the fundamental tenants of the methyl based entropy meter formulation, which has been alluded to previously, is the requirement for sufficient coverage by methyl side-chains in a protein. Without sufficient coverage, a reliable proxy for conformational entropy cannot be attained. Secondly, a closer look at the binding interfaces of protein:protein interactions, as well as protein:DNA/RNA/small molecule interactions reveal that population of methyl groups within these interfaces is minimal compared to other side-chains, especially charged amino acids like arginine, lysine, glutamic acid, glutamine. However, an even more important and ubiquitous class of amino acids, namely, aromatic amino acids heavily populate binding interfaces due to the ability to contribute to not only van der Waal interactions but also hydrogen bonding as well as their unique ability to form stacking interactions with other aromatic amino acids. Their

flexible nature in the interaction landscape favours their presence in the binding interfaces. Agonisingly, thermodynamic information, specifically the conformational entropy contributions, from aromatic amino acids remains unexplored. This is attributed to the complications that arise in the NMR-relaxation measurements of aromatic amino acids. These results promote a more exhaustive examination of the quantitative role of conformational entropy in the free energy governing protein-ligand associations. This may be particularly important in the context of proteins with unnatural man-made pharmaceuticals where the influence of conformational entropy remains largely unknown [76].

Exhaustive sampling of the conformational entropy contributions from side-chains other than methyl require the development of key techniques which would aid in the NMR-relaxation measurements. In the next chapter, several issues pertaining to the development of the NMR-relaxation of aromatic amino acid side-chains is discussed. Furthermore, a novel scheme aiding in significantly simplifying the NMR-relaxation measurements of aromatic amino acid side-chains is presented.

Eventually, combining methyl- based conformational entropy measurements and aromatic side-chain based conformational entropy measurements will result in a global view of the changes that occur in the protein thermodynamic landscape during association events with ligands. This will be further discussed by looking at a very high affinity protein:protein interaction system. New insights from this study will be presented and its implications for opening new avenues of rational drug design discussed.

Chapter 3

Aromatic amino acid side-chain relaxation

3.1 Introduction

NMR-relaxation has proven to be a versatile probe of the link between fast internal motions that occur in proteins over a wide range of time scales and their relevance to function [14, 17, 77, 78]. The motions of the backbone or methyl side-chain in proteins are most commonly studied using ^{15}N relaxation or ^{13}C, ^{2}H relaxation respectively [17, 78]. This is due to the restrictions arising from the requirements of isotopic labelling and unfavourable relaxation properties of some of the sites within proteins. Aromatic amino

acids are one such example. Aromatic amino acids are plagued by unfavourable relaxation properties, predominantly, scalar coupling to nearby nuclei as well as their strong chemical shift anisotropy which heavily influences NMR-relaxation measurements and their subsequent interpretation. Needless to say that there have been several attempts at addressing this issue for methyl groups as well as aromatic amino acids. Examples include the use of $[3-^{13}C]$-pyruvate [79], $[2-^{13}C]$-glycerol, or $[1,2-^{13}C]$-glycerol [80, 81] or even mixtures of singly ^{13}C-enriched acetates [82] as carbon precursors to generate isolated ^{13}C spins. In some cases even more selective spin enrichment schemes are sometimes required in order to suppress deleterious spin interactions and invoke more complex biosynthetic precursors [35, 37, 79]. Such complex isotope labelling methods involve chemical sythesis as well, which though expensive, has proven viable [83]. With the exception of $[1-^{13C}]$-glucose, $[2-^{13C}]$-glucose [84] and SAIL isotope labelling schemes [83], generating isolated ^{13}C spins in aromatic amino acids has proven difficult. However, these labelling schemes have yet to be critically tested for their relaxation properties. Furthermore, even though these generate isolated ^{13}C spins in the aromatic ring, the aromatic ring remains protonated and in some cases even have residual ^{13}C label at other positions. Cumulatively, these result in through bond scalar coupling with remote protons (2-3 $\%$ contribution to relaxation of ^{13}C nuclei), ^{4}J ($\sim$ 5 Hz) bond meta coupling with proton (results in modulation of intensity in NMR-relaxation experiment), two bond ^{13}C - ^{13}C J coupling interaction, as well as remote ^{13}C scalar coupling interactions which are prevelant in samples using glucose as precursor.

In order to obtain very clean and interpretable NMR-relaxation data for aromatic amino acid side-chains, it is absolutely necessary to control and if possible eliminate some or all of these extraneous interactions [84–86]. Such clean data is vital for understanding the role played by internal motions of aromatic residues in protein function, in the context of molecular recognition, and catalysis [29, 87–92]. Over the wide time scale of internal motions, the relatively slow time scale of motions of aromatic amino acid side chains have been previously documented in very few cases, these are manifested in line broadening or population exchange [93]. Such measurements have predominantly been observed though the aromatic protons rather than the ^{13}C spins [93]. For probing the fast sub nanosecond motions, aromatic residues present a difficult situation. Previously published work on aromatic ^{13}C spin labelling even though reported to singly label the Cδ carbon of the aromatic ring, when tested, resulted in scrambling to other positions in the aromatic ring apart from other side-chains in the protein. This as mentioned above confounds relaxation analysis. The ideal system for NMR-relaxation measurements involve the presence of a single isolated ^{13}C - ^{1}H spin pair in the aromatic ring which is other per-deuterated i.e. all other positions are ^{12}C spins with deuterium (^{2}H) attached. This is preferable as there are no through bond scalar couplings with ^{13}C spins as well as the absence of remote scalar coupling with ^{1}H spins. Though anticipated to be less of an issue for methyl ^{13}C relaxation studies [35, 50], the presence of remote ^{1}H spins does present a complication in the analysis of aromatic ^{13}C relaxation in proteins. The sole phenomenon responsible for relaxation of such an isolated spin

pair is the direct bonded dipolar coupling with ^{1}C spin and the chemical shift anisotropy of the ^{13}C spin nuclei in the aromatic ring [17, 19].

Labelling strategies based on glucose will not provide such specificity due to the abundance of pathways in which glucose is utilized [2]. To overcome this limitation, this problem is tackled by analysing the root of the biochemical pathways involved in the synthesis of aromatic amino acids. The early committed step in the synthesis of aromatic amino acid involves the condensation reaction of a three carbon moiety (phosphoenol pyruvate) and a four carbon moiety (erythrose 4-phosphate) [2, 94, 95]. It is interesting to note here that both these carbon moieties are generated from glucose in the typical glycolysis cycle. This provides evidence as to why the use of glucose most often results in scrambling of the carbon label within the aromatic ring system. There are two possibilities in the ^{13}C isotope labelling of aromatic residues:

- Use of phosphoenol pyruvate to introduce ^{13}C isotope label
- Use of erythrose 4-phosphate to introduce ^{13}C isotope label

The use of pyruvate to introduce ^{13}C spin label is avoided since pyruvate like glucose is involved in the synthesis of many carbon precursors within the cell. This introduces the possibility of isotope scrambling into other amino acids as well as other positions of the aromatic ring. Hence, erythrose 4-phoshphate, the four carbon moiety is chosen as the best candidate. Upon following the carbon flow pathway in the glycolysis cycle as well as the condensation reach of phosphoenol pyruvate with erythrose 4-phosphate

(Figure 3.1), the C4 position of the erythrose 4-phosphate results in one of the $C\delta$ (say C3 position) positions of the aromatic ring while the other $C\delta$ position (C5 position) in the aromatic ring originates from phosphoenol pyruvate. Also, the $C\epsilon$ position of the aromatic ring proceeds via a $C = O$ intermediate, which is reduced to $C - H$ (for more detailed information, refer to [2]). Hence, the use of singly labelled $[4 -^{13} C]$-erythrose (erythrose 4-phosphate is not commercially available) along with completely deuterated, and ^{12}C labelled pyruvate in deuterated medium would result in an isolated ^{13}C label at one of the $C\delta$ position with a proton attached. Also, the other $C\delta$ position (i.e. C5) of the ring would be unlabelled as well as attached to a ^{2}H. Furthermore, the $C\epsilon$ positions would be ^{12}C attached to a deuterium (^{2}H) due to the $C = O$ intermediate reduced in a deuterated medium. The $C\eta$ position of the ring in the case of tyrosine has a hydroxyl group which does not contribute to the relaxation of the ^{13}C spin label at the $C\delta$ position. However, in the case of phenylalanine and tryptophan, this $C\eta$ position is ^{12}C attached to a proton (^{1}H). This proton would contribute to the relaxation of the ^{13}C spin label but the contribution is estimated to be <1% (Figure 3.1).

This labelling pattern therefore largely eliminates the potential complications of extraneous intraring scalar or dipolar interactions (^{1}H or ^{13}C) with the isolated ^{13}C spin. The presence of remote ^{2}H deuterium spins eliminates the remote scalar coupling contributions. Additionally, the use of ^{13}C labelled erythrose results in reduced scrambling compared to glucose. This reduced scrambling is due to the fact that erythrose is predominantly involved in the synthesis of aromatic amino acids, including histidines as

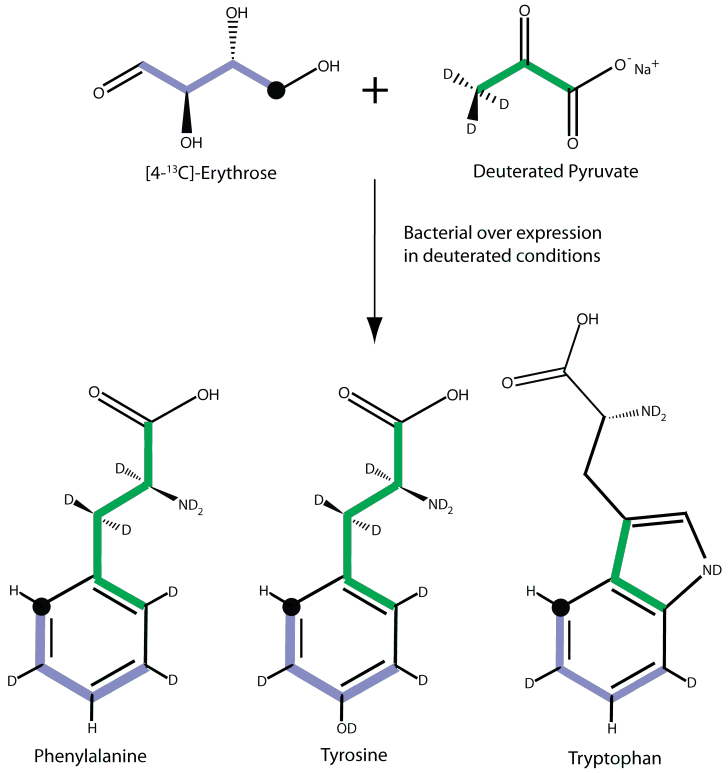


Figure 3.1: Carbon inflow into the aromatic pathway through the condensation reaction of erythrose 4-phosphate and phosphoenol pyruvate. Carbons originating from pyruvate and erythrose are shown in green and purple, respectively. The position of the ^{13}C label is indicated by the • symbol.

well as in the synthesis of lipids. Also, due to the low steady state levels of erythrose in the cell, scrambling is significantly reduced.

3.2 Establishing ^{13}C labelling in aromatic amino acids

The previous section details the outline of the labelling scheme proposed for site-specific ^{13}C isotope labelling of aromatic amino acid side-chains. However, in order to

verify that the experimentally detected labelling pattern and the theorized or expected labelling pattern match, extensive NMR experimental results are discussed in this section. There are several issues addressed here, namely, the efficiency of labelling i.e. the percentage labelling achieved at the site-specific $^{13}C_{\delta}$ position, the nature of such percentage labelling observed i.e. the fraction of $^{13}C_{\delta}$ versus $^{12}C_{\delta}$. Also discussed are the deuteration at other sites both in the surrounding milieu of protein as well as the other aromatic ring positions. Finally, evidence for the ^{13}C site-specific labelling in trytophan is also discussed. Two different proteins are used here as test examples, *Calmodulin* [96] and *Flavodoxin C55A* [97].

Expression of Vertebrate Calmodulin: Vertebrate calmodulin is expressed in *Escherichia Coli* using BL21(DE3) cells carrying expression plasmid encoding vertebrate calmodulin [96]. These cells are pre-adapted to growth in D_20 before carrying out expression and purification. Calmodulin is expressed at $37°C$ using BL21(DE3) cells grown in M9 minimal media/$99\%D_20$ supplemented with 1.0 gram/L ^{15}N-labeled ammonium chloride and 2.0 gram/L deuterated ^{12}C pyruvate as sole nitrogen and carbon source, respectively. The cells are grown to an OD_{600} of 0.55 at which point 1.3 gram/L of $[4\ -^{13}\ C]$-erythrose pD ($pH^{\star}$) 7.8 was added to the growth cultures. The cultures are allowed to grow for a further 60-75 minutes and subsequently induced with 1 millimolar (mM) isopropyl β-D-1-thiogalatopyranoside (IPTG). Induction of cell cultures are carried out at $27°C$ for 17 hours. The final OD_{600} of the two cultures were 1.2 and 1.1, respectively. Calcium saturated calmodulin is purified using established methods. This yields 60 milligram/L of calmodulin. The use of this above mentioned isotope labeling scheme results in 67% ^{13}C enrichment at $C\delta$ aromatic carbons sites. Uniformly ^{15N} enriched and ^{13}C aromatic-labeled calmodulin (1 mM) is prepared in 10 mM MES (2-(N-morpholino) ethanesulphonic acid) buffer pH 6.5, 50mM KCl, 5mM $CaCl_2$, 0.02% (w/v) azide. ^{15}N relaxation experiments were carried out in the above buffer with $10\%D_20$, whereas for ^{13}C relaxation experiments, the protein sample is lyophilized and re-dissolved to a concentration of 1 mM in the same buffer mixture containing nominally 100% D_20 (pD 6.9). All NMR experiments are carried out at $35°C$.

One dimensional (1D) ^{13}C filtered and unfiltered ^{1}H NMR spectra revealed complete

deuteration of the protein except at single $C\delta$ positions in tyrosine and phenylalanine residues and the $C\eta$ position of phenylalanine. No protonation at any other carbon sites is observed (Figure 3.2).

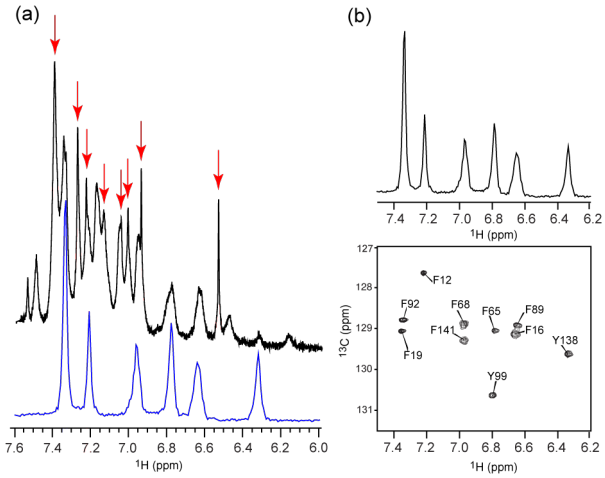


Figure 3.2: (a) Comparison of the aromatic regions of ^{13}C-filtered and decoupled 1D ^{1}H (lower, blue) and the unfiltered ^{1}H spectrum (upper, black) of calcium-saturated calmodulin expressed using erythrose labelling scheme indicating the selective introduction of a ^{1}H - ^{13}C pair at a single $C\delta$ carbon of phenylalanine and tyrosine. Also observed at ^{1}H resonances attached to $^{12}C_{\eta}$ of phenylalanine (red arrows). These resonances have narrow linewidth due to the absence of scalar coupling. (b) 2D ^{13}C HSQC spectrum of the aromatic region and the corresponding 1D ^{1}H (^{13}C filtered) spectrum. No significant ^{13}C labelling of other amino acids is observed.

The enrichment efficiency is further determined by comparing ^{13}C filtered ^{1}H spectra with and without ^{13}C decoupling. Based on the ratio of split peaks due to the intrinsic scalar coupling of ^{13}C spins with ^{1}H spins with that of the degenerate resonance of ^{1}H when attached to ^{12}C, the percentage labeling efficiency is determined to be around 67% (Figure 3.3).

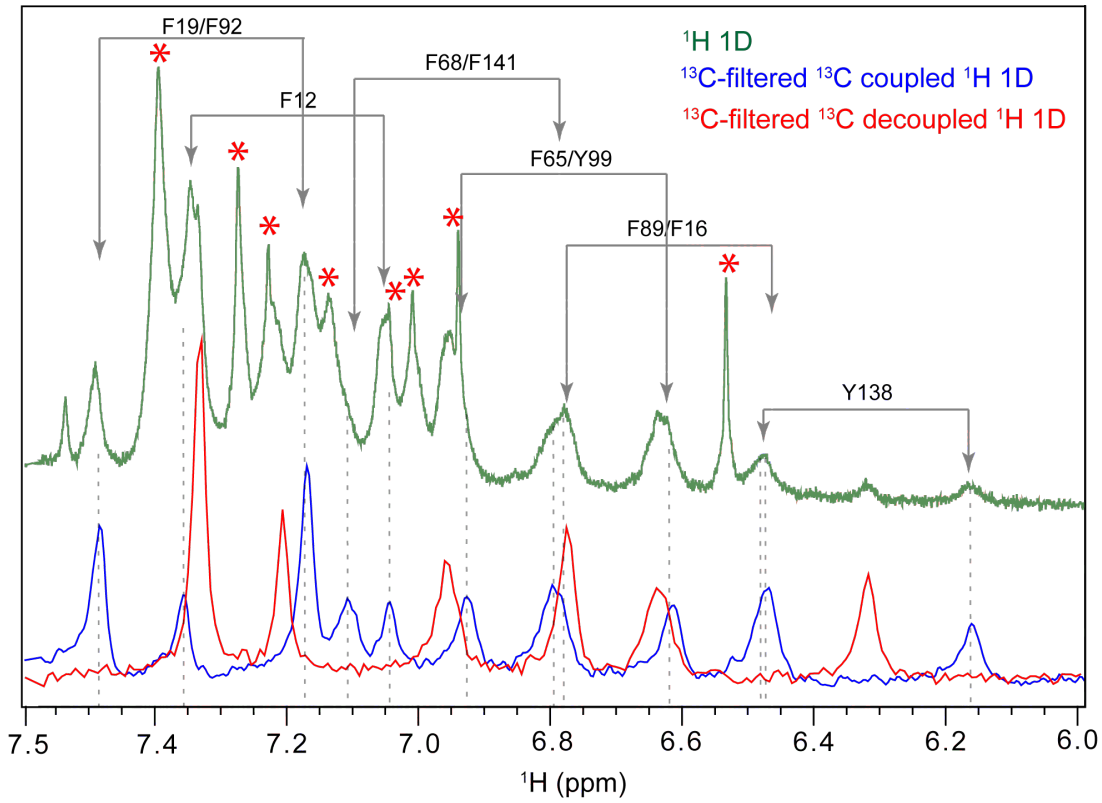


Figure 3.3: Analysis of ^{1}H 1D spectrum of calcium saturated calmodulin. The pairs of peaks indicated by the arrows in the ^{1}H 1D spectrum (green) correspond to splitting of ^{1}H resonance by its corresponding $^{13}C-^{1}H$ scalar coupling constant ($J_{CH} \sim 152-157Hz$), the ^{1}H resonance in the centre of these pairs correspond to the proton attached to a ^{12}C nucleus. For the spin pairs that are resolved in the ^{1}H 1D spectrum, Y138, F19/92 (two resonances overlapped), and F12, the ratio of the integrated intensities of these peaks used to estimate the ^{13}C labelling percentage yields 67%. Also shown are the ^{13}C-coupled (blue) and ^{13}C-decoupled (red) ^{1}H 1D, ^{13}C-filtered spectrum as evidence that the peaks observed in the ^{1}H 1D (^{13}C unfiltered) spectrum do indeed correspond to the $^{13}C_{\delta}$ -^{1}H of the aromatic rings. Also observed in the ^{1}H 1D unfiltered spectrum are the η-hydrogen resonances (*) predicted to be present in phenylalanine (a total of 8 are observed for all 8 phenylalanines in calmodulin). The η-hydrogen resonances exhibit narrow linewidths due to the absence of any ^{3}J ^{1}H - ^{1}H scalar coupling.

In order to establish that this 67% ^{13}C labelling is not due to the dilution of $^{13}C_{\delta}-^{1}H$ with $^{13}C_{\delta}-^{2}H$, 1D ^{13}C spectra with and without ^{1}H coupling was collected (Figure 3.4). This further established that the under labelling of ^{13}C is a consequence of using deuterated

pyruvate ($^{12}C,^{2}H$) : [4 $-^{13}C$]-erythrose in a 3:2 ratio. The motivation for using such a ratio was dictated by the need to suppress scrambling of ^{13}C to other amino acids or other positions in the aromatic rings.

3.3 Fast dynamics of aromatic amino acid side-chain in calmodulin

The limited amount of information available on the fast internal motions of aromatic amino acids has hampered our understanding of the fundamental aspects governing their motions. The results of small MD simulations and tryptophan side-chain ^{15}N NMR-relaxation has resulted in the notion that aromatic amino acid side-chains are predominantly rigid. In order to gain new insight into the internal motions of aromatic amino acids and to simultaneously challenge the notion that aromatic amino acids are predominantly rigid, the fast internal motions of aromatic amino acids in vertebrate calmodulin are measured. The presence of aromatic residues near methionines in calmodulin have previously been hypothesized to be correlated to the large scale changes observed in methionine methyl side-chain motions [98]. The recent direct measurements of CH/π interactions [99] between methyl groups and aromatic rings further hint at a plausible correlation of their dynamic characteristics. This of course remains a hypothesis. Calmodulin is chosen here due the abundance of aromatic amino acids as well to understand the potential functional role in the recognition of the diverse set of calmodulin

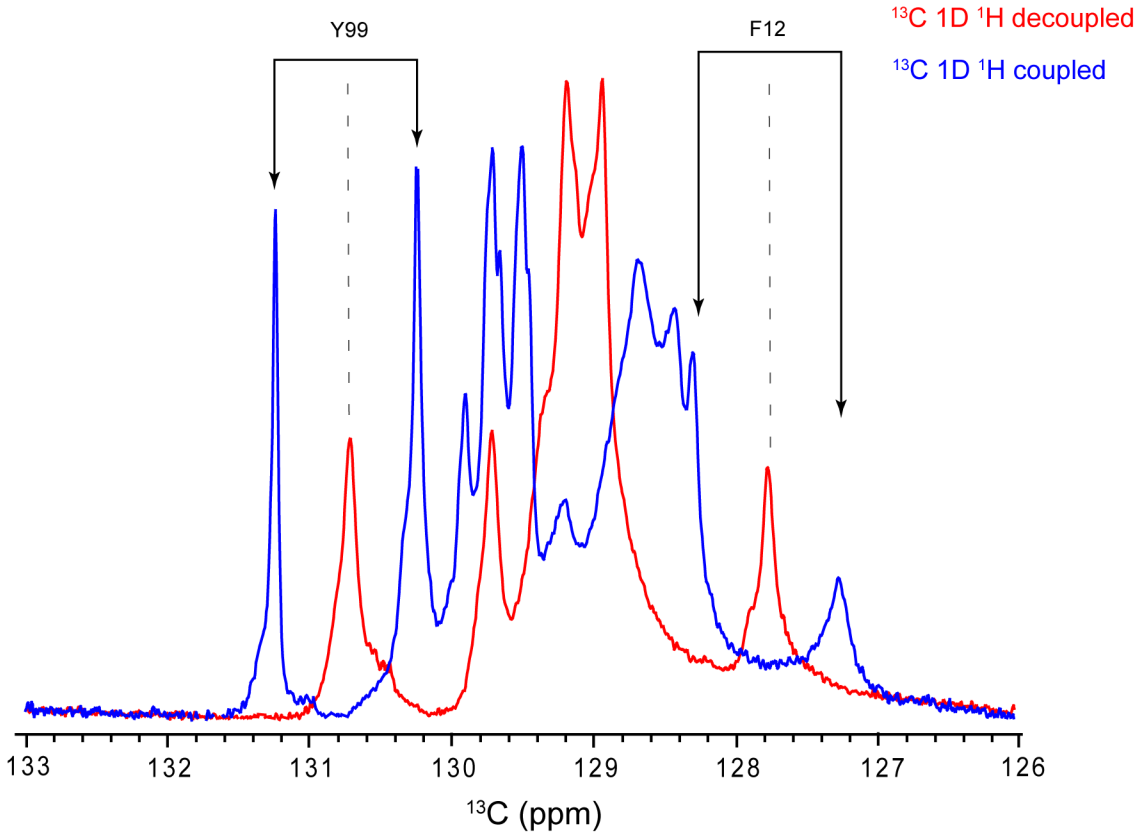


Figure 3.4: Analysis of ^{13}C 1D spectrum of calcium saturated calmodulin. Shown Above are the ^{1}H coupled (blue) and ^{1}H decoupled (red) ^{13}C 1D spectrum. In the coupled spectrum, each $^{13}C-^{1}H$ results in the splitting of ^{13}C resonance by $^{1}J_{C-H}$ ($\sim$ 155 Hz). However the presence of any $^{13}C-^{2}H$ bond remains unaffected by the ^{1}H decoupling due to the absence of ^{1}H. This $^{13}C-^{2}H$ resonance, if present, would appear at the ^{13}C chemical shift offset by -0.3 p.p.m compared to that observed in the ^{1}H decoupled ^{13}C spectrum (dashed lines as shown). The offset is due to the isotope shift (^{2}H versus ^{1}H). Two aromatic residues (Y99, and F12) are shown. These display resolved ^{13}C chemical shift in the 1D spectra with little overlap. No measurable $^{13}C-^{2}H$ intensity is observed, implying that the $^{13}C_{\delta}$ in the aromatic amino acids in the expressed protein during growth on deuterated pyruvate and $[4-^{13}C]$-erythrose in D_2O exists as $^{13}C_{\delta}-^{1}H$ (> 99%). The ^{13}C 1D spectra was collected with 16384 complex points and signal averaged using 16384 scans. The sweep-width employed in ^{13}C was 200 p.p.m with carried on carbon set at 100 p.p.m. Both ^{1}H decoupled and ^{1}H coupled ^{13}C 1D spectra were collected without ^{1}H or ^{2}H saturation and standard Waltz-16 ^{1}H decoupling was used in the decoupled spectra with proton carrier at 7.0 p.p.m. Irrespective of either a 1.5 seconds or 8 seconds recycle delay, no $^{13}C-^{2}H$ peak is observed even with very high signal averaging.

Typtophan labelling in Flavodoxin: On account of the fact that calmodulin does not contain any tryptophans and to confirm the site specific ^{13}C labelling of tryptophans, flavodoxin C55A [97] was expressed in deuterated pyruvate, $[4-^{13}C]$-erythrose in D_2O. Flavodoxin has four *Trp*, eight *Tyr*, and eight *Phe* residues along with one *His* residue . The twenty-two anticipated aromatic $^{13}C-^{1}H$ correlations were seen in the ^{13}C heteronuclear single-quantum coherence spectrum (Figure 3.5). No other significant ^{13}C labelling was observed confirming that the desired labelling pattern for all four aromatic amino acid side-chains is observed in the context of a per-deuterated background.

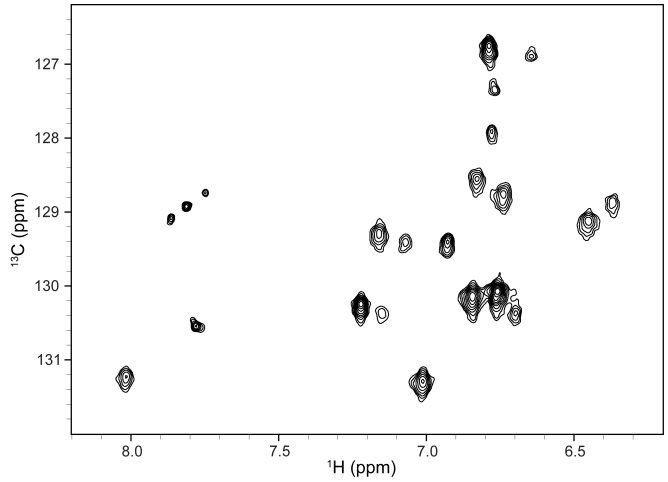


Figure 3.5: ^{13}C heteronuclear single-quantum spectrum of aromatic region of oxidized flavodoxin C55A expressed during growth on $[4-^{13}C]$-erythrose and deuterated pyruvate in D_2O. All 22 anticipated correlations are observed.

binding partners. ^{13}C R_1 and $R_{1\rho}$ relaxation in calcium-saturated calmodulin prepared using this labelling strategy (Figure 3.1) is measured at three magnetic fields (11.7, 14.0, and 17.6 Tesla) using standard two-dimensional (2D) sampling pulse sequences [37, 84]. For comparison, similar measurements are also performed on calmodulin grown using the $[1-^{13}C]$-glucose labelling scheme [84, 86]. (Figure 3.6).

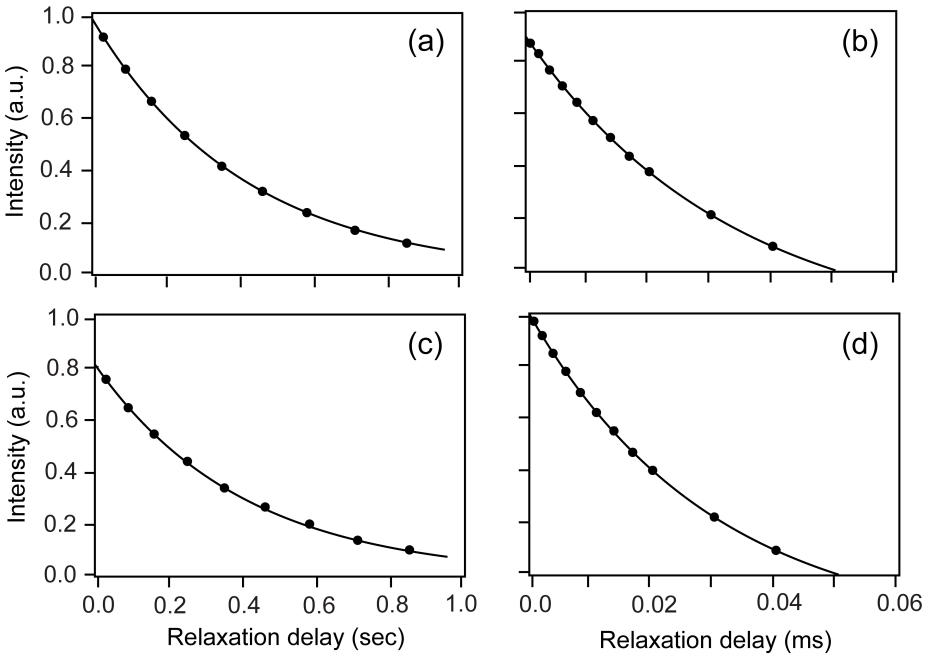


Figure 3.6: Aromatic ^{13}C R_1, and $R_{1\rho}$ relaxation in calcium-saturated calmodulin prepared using (a,b) $[4-^{13}C]$-erythrose/deuterated pyruvate/D_2O strategy and (c,d) $[1-^{13}C]$-glucose strategy. Relaxation at the $^{13}C_\delta$ of F92 is shown.

The global macromolecular tumbling of calmodulin is characterized using the ^{15}N relaxation experiments [21] collected on the same sample but in H_2O buffer [20, 33]. The anisotropy of the tumbling is characterized using the crystral structure of calcium-saturated calmodulin (PDB entry 3CLN) and assessing the two globular domains separately [100]. Macromolecular tumbling characterized using backbone relaxation employed a N-H bond length of 1.04 Å [101] and a simple uniform ^{15}N CSA tensor breadth of -170 p.p.m [102]. The N- and C- terminal domains of calmodulin are treated separately for tumbling analysis to avoid biases in order parameter. The N-terminal domain is determined to be tumbling with an effective tumbling of 8.96±0.14 ns while, the C-terminal domain tumbles at 8.05±0.10 ns. This calculated macromolecular tumbling is

then subsequently used for the calculation of Lipari-Szabo model-free squared generalized order parameter (O^2) [18] and effective correlation times (τ_e) using a grid search approach [103]. An effective bond length of 1.09 Å and residue-specific chemical shift anisotropy (CSA) tensors with axially symmetric and anisotropic CSA values for Phe, and Tyr, respectively, according to Ye et al [19, 104]. The appropriate use of chemical shift anisotropy is vital for the analysis aromatic relaxation since CSA contributes substantially for the relaxation of aromatic nuclei. Also, contributions from $^{13}C-^{2}H$ as well as $^{13}C-^{13}C$ dipolar coupling [105] due to natural abundance is calculated to contribute < 0.05% of the $^{13}C-^{1}H$ direct bond dipolar coupling interaction. Relaxation data obtained from the sample grown using erythrose labelling scheme (Figure 3.1) fit well to the simple model-free spectral density [18]. The obtained O^2 values ranged from 0.47 to 0.96, indicating a rich spectrum of aromatic ring motions within calcium-saturated calmodulin on the subnanosecond time scale (Table 3.1). All the analysis was carried out using an updated version (Relxn2A [50, 103]) of the Wand lab in-house software which I have programmed using Python (www.python.org).

Similarly, the primary ^{13}C R_1 and $R_{1\rho}$ relaxation profiles from calcium-saturated calmodulin grown in $[1-^{13}C]$-glucose [84] fitted well to single-exponential decays (Figure 3.6). However, in contrast, the obtained relaxation rates largely gave relatively poor fits to L-S model-free [18] interpretation (5-10% vs <1% residual error). Inclusion of remote ^{1}H spin dipolar interactions as well as dipolar interactions with $^{13}C_\epsilon$ failed to recover the excellent statistics of the relaxation data derived from the more optimal $[4-^{13}C]$-erythrose

Residue	O^2	τ_e
F12	0.49±0.03	236±21
F16	0.95±0.01	127±12
F19	0.47±0.01	232±19
F65	0.70±0.04	176±23
F68	0.96±0.01	292±17
F89	0.94±0.02	900±39
F92	0.72±0.03	624±23
Y99	0.79±0.02	240±21
Y138	0.89±0.02	604±36
F141	0.95±0.01	101±11

Table 3.1: Lipari-Szabo Model-Free Parameters for Aromatic Ring Motion in Calcium-Saturated Calmodulin grown in $[4-^{13}C]$-erythrose

scheme. It seems likely that the unaccounted dipolar relaxation, intraring one-bond $^{13}C-^{13}C$ J coupling, and intraring two-bond $^{13}C_\delta-^{1}H_\epsilon$ J coupling effects contaminated the measurement and interpretation of the ^{13}C relaxation in structured proteins in protonated backgorund. This is consistent with the presence of additional peaks in the $^{13}C-^{1}H$ HSQC spectrum of calmodulin derived from $[1-^{13}C]$-glucose labelling scheme (Figure 3.7). These additional peaks arise from the partial ^{13}C labelling (8-15%) at C_ϵ of the aromatic ring and are a consequence of the scrambling of the ^{13}C label when glucose is used as a carbon precursor. These considerations provide a plausible scenario where the relaxation profiles fit reasonably well to single-exponential decays but fail to be fit reasonably by the Lipari-Szabo model-free formalism.

The results for the L-S model-free fit of the $[1-^{13}C]$-glucose summarized in (Table 3.2) indicate that even after attempting to include possible extraneous interactions, the order parameters are markedly different from that obtained by the cleaner $[4-^{13}C]$-erythrose.

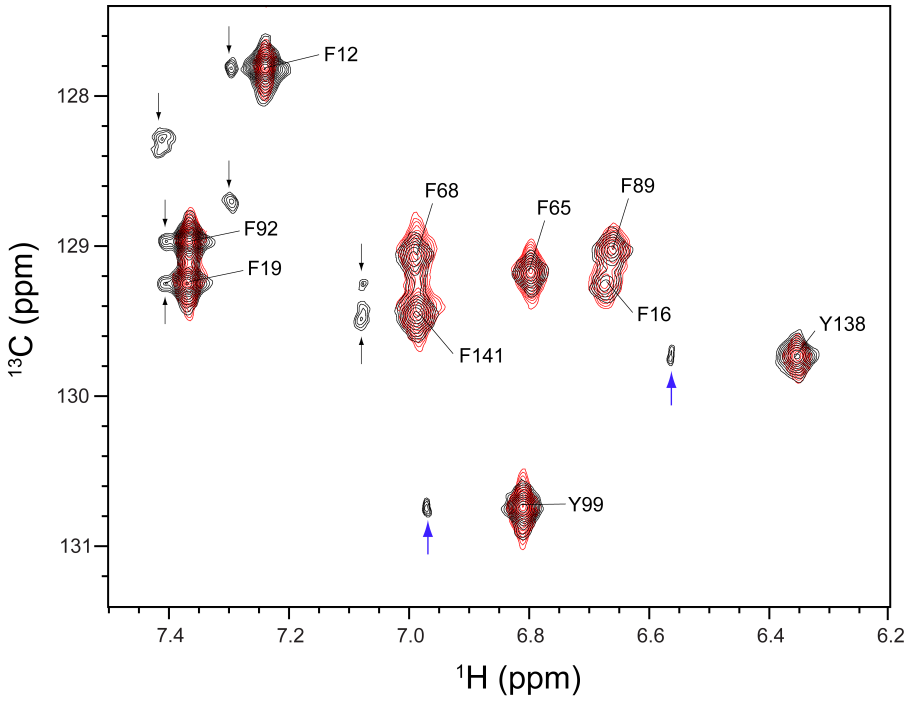


Figure 3.7: Overlay of the aromatic region of ^{13}C HSQC spectrum of calcium-saturated calmodulin expressed during growth on $[4-^{13}C]$-erythrose/deuterated pyruvate/D_2O (red), and $[1-^{13}C]$-glucose (black). The additional $^{1}H_\epsilon-^{13}C_\epsilon$ correlation peaks (8-15%) present in the $[1-^{13}C]$-glucose sample, the blue arrows indicate $^{1}H_\epsilon-^{13}C_\delta$ 2-bond J_{C-H} correlations ($\sim$ 1% ratio of integrated intensities) observed for tyrosine. Due to the dispersive nature of the ^{13}C frequency, the $^{1}H_\epsilon-^{13}C_\epsilon$ correlations of tyrosine cancel out and are in the noise plane. A full ^{13}C carbon sweep width 2D acquisition results in the presence of two tyrosine $^{1}H_\epsilon-^{13}C_\epsilon$ correlations as expected (not shown).

3.4 Summary and Discussion

With the recent shift in emphasis on the effect of protein dynamics on protein:protein, protein:ligand interactions [13, 14], it is critical to paint an exhaustive picture of the motional landscape of proteins, i.e. backbone, and side-chains. The recent methyl dynamics mediated measure of protein conformational entropy has paved the way to reliably

Residue	$\mathbf{O^{2a}}$	τ_e^a
F12	0.55±0.05	284±30
F16	0.98±0.02	492±56
F19	0.52±0.02	284±49
F65	0.91±0.05	908±81
F68	0.98±0.02	688±150
F89	0.98±0.01	724±141
F92	0.50±0.06	276±32
Y99	0.98±0.03	996±165
Y138	0.98±0.03	1200±108
F141	0.98±0.02	800±135
Residue	$\mathbf{O^{2b}}$	τ_e^b
F12	0.55±0.05	276±32
F16	0.98±0.02	436±50
F19	0.52±0.02	284±49
F65	0.91±0.05	868±85
F68	0.98±0.02	664±150
F89	0.98±0.01	680±150
F92	0.49±0.06	273±30
Y99	0.97±0.03	984±210
Y138	0.97±0.03	1260±105
F141	0.98±0.02	780±135
Residue	$\mathbf{O^{2c}}$	τ_e^c
F12	0.54±0.06	300±32
F16	0.98±0.02	1068±124
F19	0.52±0.02	284±49
F65	0.90±0.05	948±92
F68	0.98±0.02	1476±156
F89	0.98±0.01	660±81
F92	0.48±0.07	264±27
Y99	0.96±0.02	884±100
Y138	0.96±0.02	1432±156
F141	0.98±0.02	772±85

Table 3.2: Lipari-Szabo Model-Free Parameters for Aromatic Ring Motion in Calcium-Saturated Calmodulin grown in $[1-^{13}C]$-glucose

[a] Fit without including any extraneous interactions

[b] Fit with the inclusion of extraneous ^1H dipolar interactions

[c] Fit with the inclusion of extraneous ^1H and ^{13}C dipolar interactions

estimate the change in conformational entropy of proteins upon ligand binding [14]. Although proving to be a tremendous leap forward, there are still aspects of the motional landscape which remain poorly understood. Particularly relevant here is the observation that the methyl residues in proteins largely do not sample the binding interface extensively. A true picture of the change in conformational entropy has to include the changes occuring at the binding interfaces. Where methyl residues largely fail, aromatic residues are pivotal. Aromatic amino acids are usually abundant in binding interfaces as they not only contribute to van der Waals interaction but are also involved in hydrogen bonding and stacking interactions [28]. Attempts at utilizing aromatic amino acids as probes for conformational entropy changes has been unsuccessful so far due to the complicated relaxation properties associated with them [84]. Here, I introduce a simple yet powerful method to study the fast relaxation properties of aromatic amino acids using site specific isotopic enrichment (^{13}C from $[4-^{13}C]$-erythrose) of the aromatic ring accompanied by perdeuteration of the ring hydrogens (using deuterated pyruvate). This isotopic labeling scheme accomplished by providing these carbon sources to bacteria during protein over expression eliminates all of the complications (extraneous dipolar coupling with remote hydrogens, through bond scalar coupling with ring hydrogens and ^{13}C carbons) traditionally associated with the measurement of fast motions of aromatic side-chains.

Contrary to the idea that aromatic side-chains are fairly rigid with restricted motion, I

find that aromatic rings exhibit a diverse range of dynamics. Measurement of fast motions of the aromatic side-chain in calcium saturated calmodulin yielded a wide range of aromatic side-chain motions within calmodulin. This study establishes standard procedures which would enable routine measurements of fast motions of aromatic side-chains.

What is the significance of this diverse range of aromatic side-chain motions observed in calcium saturated calmodulin?: Previous measurements of fast methyl motions in calmodulin in both free and ligand bound states revealed a remarkably diverse changes in the degree of motions, especially those of methionine [14, 66, 98, 106]. Calmodulin has several methionines which are very close to the binding interface of the peptide ligand. The consequence of these changes, particularly for methionine, has been puzzling. The only previous investigation into this revealed that all the methionines in calmodulin are all very close to aromatic residues [98]. Although this has so far proven to be speculative, with the above proposed scheme to measure fast motions of aromatic side-chains, it could prove very useful in understanding the nature of the changes in motions occuring in calmodulin upon complex formation. Further work involving measurement of the fast motions of aromatic side-chain in calmodulin bound to its peptide ligand would shed more light on this issue. Preliminary results on the motions of aromatic rings in calmodulin bound to smMLCK peptide is presented in the next chapter (4).

Historically, the nature of the methyl side-chain motions have been thoroughly investigated [17]. The origins of the fundamental nature of methyl group motions is now reasonable well understood and are attributed to the rotameric transitions about its symmetry axis [107]. This has proven to be very useful in understanding the dynamic transitions occuring with the interior of protein and has helped provide insights into the fluid nature of proteins and the consequence of the same on ligand binding [14, 27]. This could prove very important in attempting to design targeted drugs as these provide an additional layer or regulation and control. However, the fundamental nature of aromatic side-chain motions has not yet been thoroughly studied with the exception of a few studies.

In the next chapter, I address known issues regarding the motions of aromatic side-chains as well as provide compelling evidence for a particular type of ring motions, ring flips, which occur in the fast time scale (picoseconds to nanoseconds). These type of ring motions, ring flips, have largely been associated with slow time scale motion (microsecond to millisecond). The impact of this observation that aromatic rings can and do indeed flip on the very fast time scale is also discussed.

Chapter 4

A sharp thermal transition of fast aromatic ring dynamics in ubiquitin

4.1 Introduction

Aromatic amino acid side chains have a rich structural role within proteins [29, 87–89] and are often central to their biological function, particularly in the context of molecular recognition [90, 91] and catalysis [92]. The importance of aromatic ring dynamics has so far largely been limited to the few cases where NMR spectroscopy has been

used to study their motions in the millisecond to second time scale. These early studies largely relied on the manifestation of such slow motions through line-broadening or population exchange phenomena [93]. Moreover, such studies interpreted these slow motions of the aromatic side chain as ring flipping about the χ_2 torsion angle. due to the underlying symmetry considerations. In the previous chapter, I have demonstrated for the first time through experimental measurements that the motions of the aromatic side chain in sub-nanosecond time scale could be diverse revealing a previously unknown dynamical nature of their motions [26]. However, the fundamental nature of the motions of aromatic side-chains i.e. the thermodynamics governing the underlying fast motions still remains poorly understood. Previous studies on the temperature [108] and pressure [109, 110] dependence of the slow motions of the aromatic ring suggested a jump-like rather than continuous diffusive motion within the environment of protein milieu. These jump-like motions were also characterized to have high enthalpy barrier on account of the slow flip rates observed. The experimental measurement of these slow flip rates have been restricted to the handful of cases that show distinct peaks for the symmetry related nuclei [111], for example, when observing the carbon nuclei in the aromatic rings, two sets of peaks are observed for both C_δ, and C_ϵ. These, however, constitute a very small fraction of the total number of cases observed. As a consequence, aromatic rings which do not show line broadening or non degenerate symmetry related resonances were generally assumed to be flipping at very fast rate, faster than the chemical shift difference of the interchanging nuclei, i.e. $> 10^3 s^{-1}$. Absent from

this view point is the insight into more restricted (librational) motion within a rotamer well. There has been a recent renaissance of interest in the details of aromatic ring motion with the introduction of new experimental strategies that have broadened the spectrum of insight available. Particularly noteworthy is the use of supercooled water to access temperatures where aromatic ring dynamics with low activation enthalpies can be more favorably studied using line broadening and chemical exchange phenomena [108]. More recently, a number of isotopic labeling strategies have been introduced to enable relaxation studies to probe the picosecond to nanosecond and microsecond to millisecond timescales [26, 112, 113]. These isotropic labeling strategies are designed to ameliorate confounding spin-spin interactions [26] that have previously limited measurement of ^{13}C-relaxation in aromatic ring systems to natural abundance [93]. One of the fundamental goals of this study is to provide insight into the nature of aromatic motions within the protein. As has previously been elucidated in the case of motions of the methyl bearing side chains where a principal knowledge of the nature of their motions has allowed for a quantitative relationship with the underlying thermodynamics, in particular conformational entropy of proteins [107]. While the studies [14, 27, 107] so far have utilized methyl groups to provide a quantitative estimate for protein conformational entropy (addressed in more detail in Chapter 7, the contributions of aromatic amino acids remain poorly understood. Furthermore, the interfaces of protein-protein interactions display an abundance of aromatic amino acids as these contribute to not only van der Waals interaction but also hydrogen bonding and stacking interactions

[28]. The role played by aromatic amino acids in high-affinity interactions ($K_d < 10^{-9}M$) also remains uncharacterized.

4.2 Effect of temperature on fast motions of aromatic rings

Temperature dependence of fast motions has previously been used to understand its thermodynamic landscape, in particular, heat capacity and activation volumes for motional transitions [12, 114]. Here, I have used ^{13}C-relaxation to characterize the fast internal motion of phenylalanine and tyrosine ring systems in the protein ubiquitin. Ubiquitin [51] is a small 76-residue protein essential to the eukaryotic ATP-dependent protein degradation pathway [115]. For such a small protein, ubiquitin displas a surprising range of secondary structure that includes a five-stranded mixed β-sheet, α-, and 3_{10}-helices and a number of tight turns. The protein also contains three aromatic residues: Phe-4, Phe-45 and Tyr-59. The aromatic rings of these residues are largely buried ($\sim$ 25% solvent accessibility) (Figure 4.1). The focus on this relatively simple yet rich system enables the detailed characterization of the fundamentals governing the motions of aromatic side chains.

For classical relaxation phenomena used to probe fast sub-nanosecond motions, aromatic residues present a difficult situation. In addition to the concern about the isolation of the spin interaction of interest from extraneous contributions, aromatic ring systems suffer from extensive homo- and heteronuclear scalar interactions that can

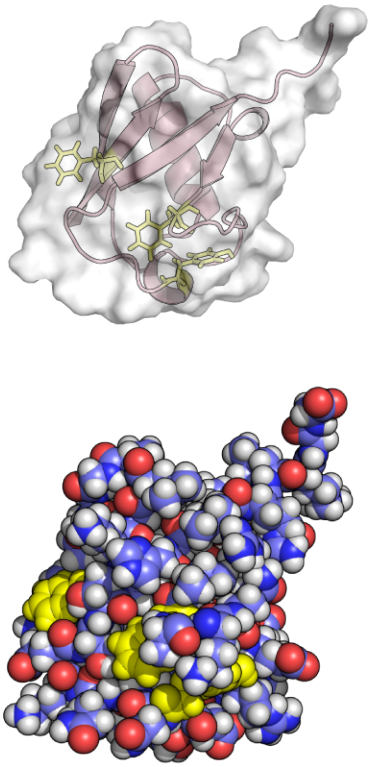


Figure 4.1: Aromatic side-chains in ubiquitin are largely buried and are located relatively far away from each other (10Å away). Shown above are two different representations of the aromatic amino acids on ubiquitin. Aromatic amino acids are colored in yellow in both representations.

also complicate relaxation data. To isolate a $^{1}H-^{13}C$ pair in the aromatic ring in an otherwise perdeuterated background I have used the biosynthetic scheme based on the $^{13}C_4 - erythrose$ [26] described in more detail in Chapter 3. The protein was also uniformly labeled with ^{15}N. I have carried out the ^{13}C-relaxation experiments at 283, 293, 303, 308, 313, 323 and 328 K. I have also carried out relaxation measurements at elevated pressures of 1200 bar and 2500 bar at 283, 303 and 323 K using previously

established methods pioneered by the Wand lab [116]. The model-free treatment of Lipari & Szabo [18] was utilized for the analysis of the backbone and side chain aromatic relaxation data. Summarized in the table below (Table 4.1) are the parameters obtained for the macromolecular tumbling at the different temperatures. For clarity, only the isotropic values are shown. The calculation of order parameters of the aromatic side chains however utilized the tumbling values corresponding to the appropriate tumbling models.

Table 4.1: Temperature dependence of the macromolecular tumbling time of ubiquitin

Temperature (deg C)	τ_m^{iso}(ns)
10[a]	7.60±0.05
20[b]	6.00±0.05
30[a]	3.90±0.03
35[b]	3.82±0.06
40[a]	3.73±0.04
45[b]	3.67±0.07
50[a]	3.60±0.04
55[b]	3.53±0.06

[a] Measured rotational correlation times
[b] Macromolecular rotational correlation times obtained via interpolation

The effect of temperature on the molecular tumbling of ubiquitin is to change from a completely anisotropic tumbling model (283 K to 308 K) to axially symmetric tumbling (above 308 K) at a pressure of 1 bar. Similar analysis for higher pressures are summarized in the next section. Using the measured and estimated values of macromolecular tumbling times, the motions of the aromatic side chain (Order Parameter:

O^2) i.e. $^1H-^{13}C_5$ bond were calculated to a high precision. The motions measured here using classical NMR relaxation measurements are on a time scale faster than the macromolecular tulmbling, which is on the order of 3 ns. I find that at low temperatures, the O^2 parameters i.e. the motions of the C_5-H bond vectors of Phe-4, Phe-45 and Tyr-59 are high, with that of Tyr-59 close to the theoretical limit of 1.0 indicating nearly complete rigidity within the molecular frame. There is little temperature dependence of the order parameters up to about 303 K where a sharp cooperative transition to lower value begins. I find that all three aromatic rings display this transition that is complete by 330 K. The fitted mid-point of this transition is 312±1 K, which is well below the thermal unfolding transition observed by calorimetry (Figure 4.2) [117]. Indeed, previous differential scanning calorimetry study of ubiquitin only consists of a barest hint in the pre-transition region [117].

All three aromatic rings reach a significantly low plateau value for their order parameters following the transition. In this regime, the order parameters seem to be largely temperature independent. While this transition seems characteristic of a two-state transition, van't Hoff analysis of the same results in an erroneous enthalpy change of ~30 kcal mol^{-1} (Figure 4.3). Although large 'activation' enthalpies for aromatic amino acid side chains motions have been observed previously, these studies were largely focussed on the slow time scale ring flip motions. Hence, the large enthalpy change obtained using van't Hoff analysis is puzzling.

However, analysis of the temperature dependence of the effective correlation times

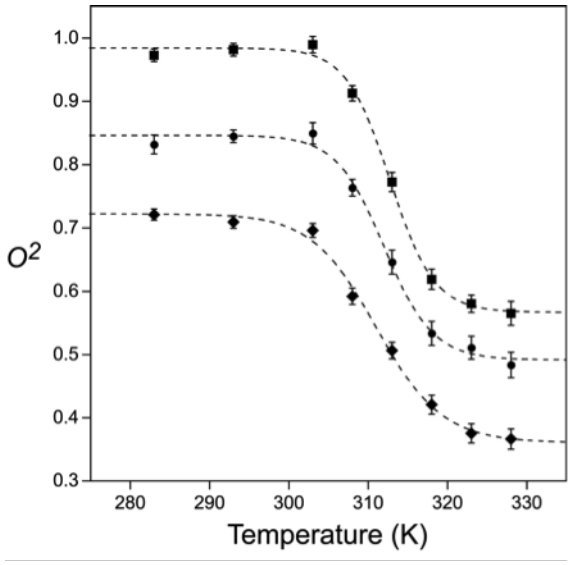


Figure 4.2: Lipari-Szabo model-free squared generalized order parameters of the $C_5 - H$ ring bond vectors of Tyr-59 (■), Phe-4 (◆) and Phe-45 (●) determined by ^{13}C-relaxation. The dashed lines are fitted sigmoidal curves.

revealed the problems associated with the assumption of a two-state dynamical transition. During the transition the fitted τ_e values *increase* substantially (Figure 4.4). This indicates that increase in temperature results in the changes in the qualitatively slower motions seen at low temperatures.

The effective correlation time of Tyr-59 undergoes an extremely sharp transition with an apparent midpoint of 305±2 K. Phe-4 and Phe-45 however both show gradual transition in effective correlation times with a similar midpoint temperature ($\sim 314 \pm 2$ K). The rigid nature of the aromatic ring allows for relatively straightforward modeling of the influence of available motions about the connected χ_1 and χ_2 side chain torsion angles on the obtained O^2 parameter [118]. Steric considerations suggest that motion about χ_1 can be largely neglected. At temperatures below the dynamical transition the high

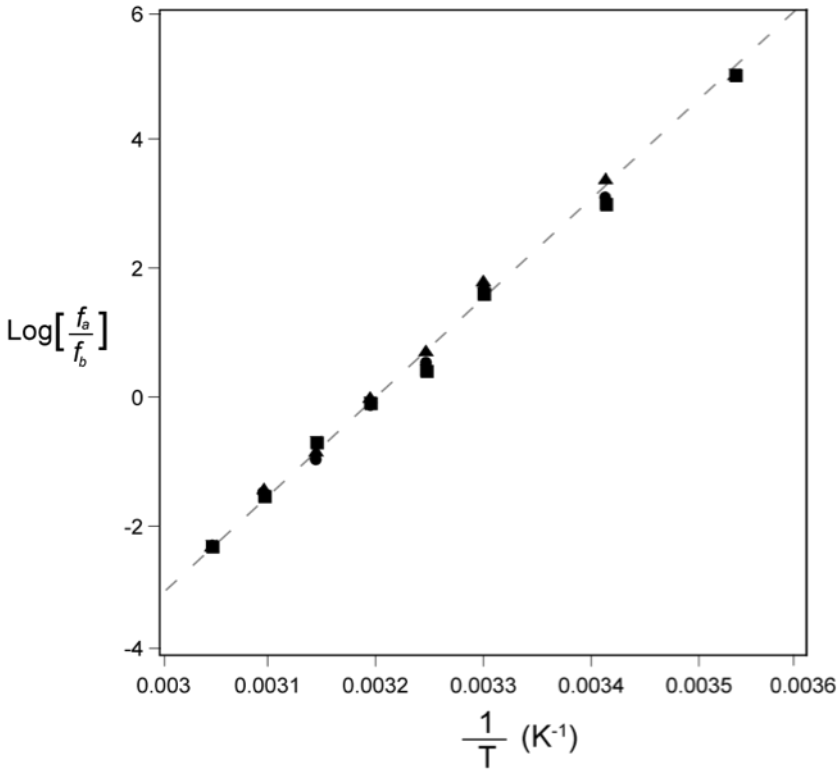


Figure 4.3: Two-state like dynamical transition results in a large activation enthalpy of ~30 kcal mol^{-1} for all three aromatic rings in ubiquitin.

values of O^2 strongly suggest the absence of aromatic ring rotation or flipping and that the ring is undergoing liberational motion. The restricted diffusion model of Witterbort & Szabo [119] predicts liberational angles of 9, 39 and 27 degrees for Tyr-59, Phe-4 and Phe-45, respectively. The insensitivity of motions below this dynamical transition temperature indicates an essentially freely diffusive motion. The dynamical transition most likely results from the onset of qualitatively different motion, namely the rotation of the aromatic ring, as this is the only reasonably physical mechanism for lowering the O^2 parameter to 0.5 and below beyond the transition temperature [118]. However, the

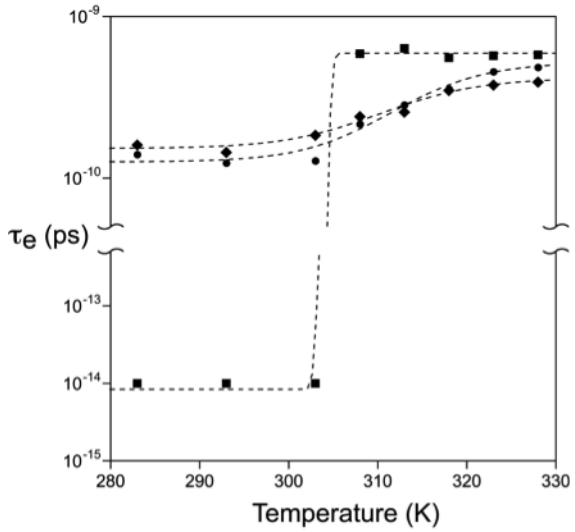


Figure 4.4: Semi-log plot of the Lipari-Szabo effective correlation times of the motion of the $C_5 - H$ ring bond vectors of Tyr-59 (■), Phe-4 (◆) and Phe-45 (●) determined by ^{13}C-relaxation. The fitted errors are less than the symbol size as indicated. The dashed lines are fitted sigmoidal curves.

experimental measurements cannot resolve the different kinds of motions experienced by the aromatic side chains in the ps-ns time scale as the motions faster than the macromolecular tumbling are averaged. To understand the qualitative nature of these motions, molecular dynamics (MD) simulations were performed at 283, 293, 303, 308, 313, 323 K and the motions of the aromatic side chains were analyzed. MD simulations were performed as described in Chapter 2 [107].

MD simulations qualitatively reproduced the effect of temperature on O^2 values (Figure 4.5), but with minor differences. Below 300 K, the order parameters of the three aromatic groups are high. With increasing temperatures above 300 K, the O^2 for both Phe-4 and Phe-45 residues decrease sharply and in concert, to a low value. However,

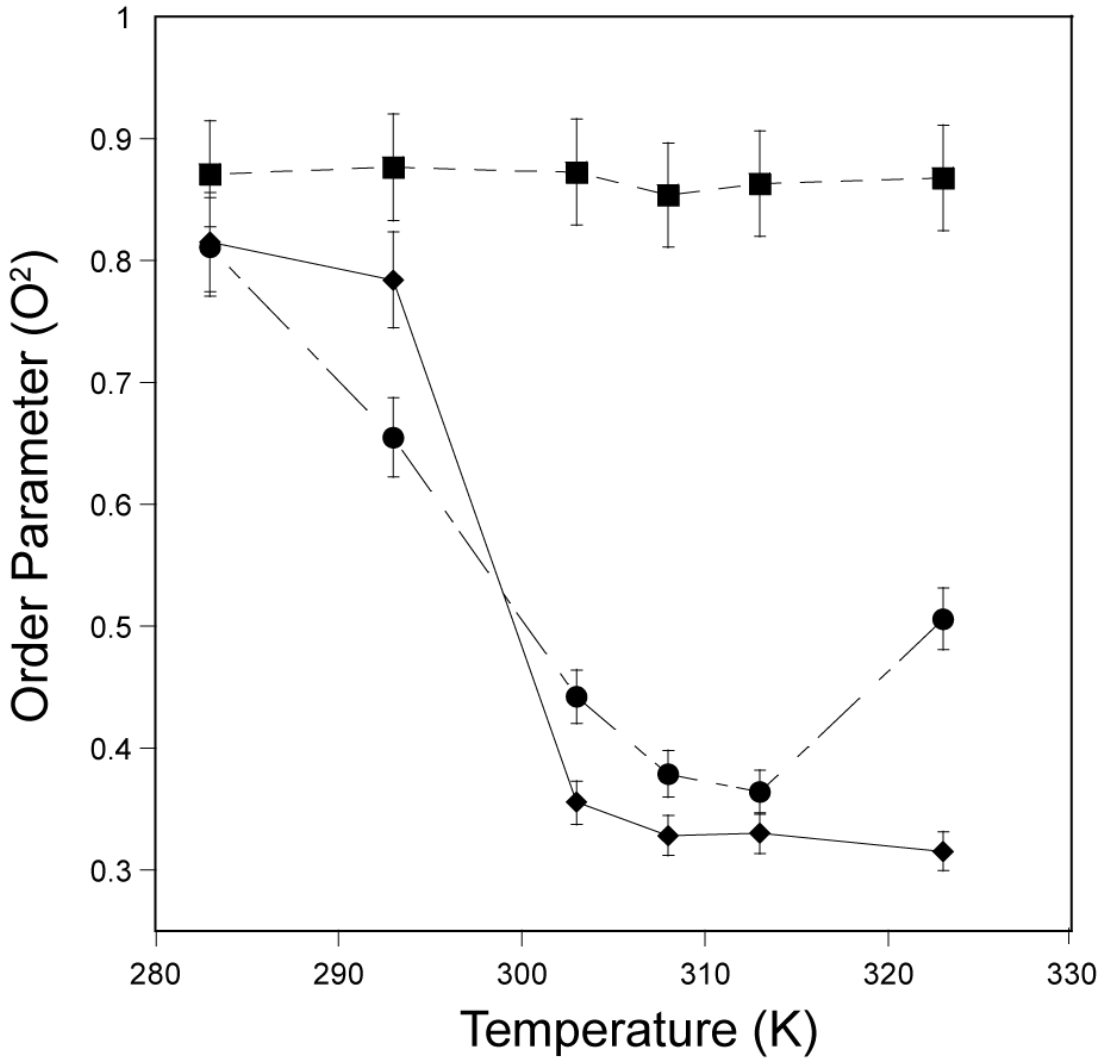


Figure 4.5: Lipari-Szabo model-free squared generalized order parameters of the $C_5 - H$ ring bond vectors of Tyr-59 (■), Phe-4 (◆) and Phe-45 (●) derived from 120 ns of MD simulations at 1 bar.

the O^2 for Tyr-59 does not show any temperature dependence from the MD simulations in contrast to the experimental measurements. This is likely attributed to the nature of the force fields used in MD simulations. These force fields are primarily intended for folding proteins [45] and that they are biased towards the presence of hydrogen bonds due to their favorable energetic contributions. Tyr-59 in ubiquitin acts both as a hydrogen bond donor and acceptor. Thus the MD simulations most likely preserve this hydrogen bonding pattern across the different temperatures resulting in the insensitivity of the O^2 from MD to temperature. The temperature dependence of Phe-4 and Phe-45 is almost identical and in agreement with the experiment with differences only seen in the midpoint of transition being 10 K lower than the experiment measurement. MD simulations revealed that the two Phe residues experienced liberational motions of $\sim 15^\circ$ around the χ_2 torsion angle below 300 K. Further analysis of the aromatic side chain population distribution as a function of the χ_2 torsion angle revealed the onset of ring flips with increasing frequency above 300 K (Figure 4.6).

Evident from Figure 4.6 is that no ring flips are seen below 300 K. The width of the population distribution at all temperatures are indifferent indicating that the liberational motions are largely insensitive to temperatures and that the transition observed is due to the onset of ring flip motions. Even though it is now clear that the dynamical transition behaviour is due to ring flip motions, it is still not clear why the effective correlation times increase with increase temperatures or in other words why is the assumption of a two-state like transition erroneous? MD simulations (Figure 4.7) revealed a multiexponential

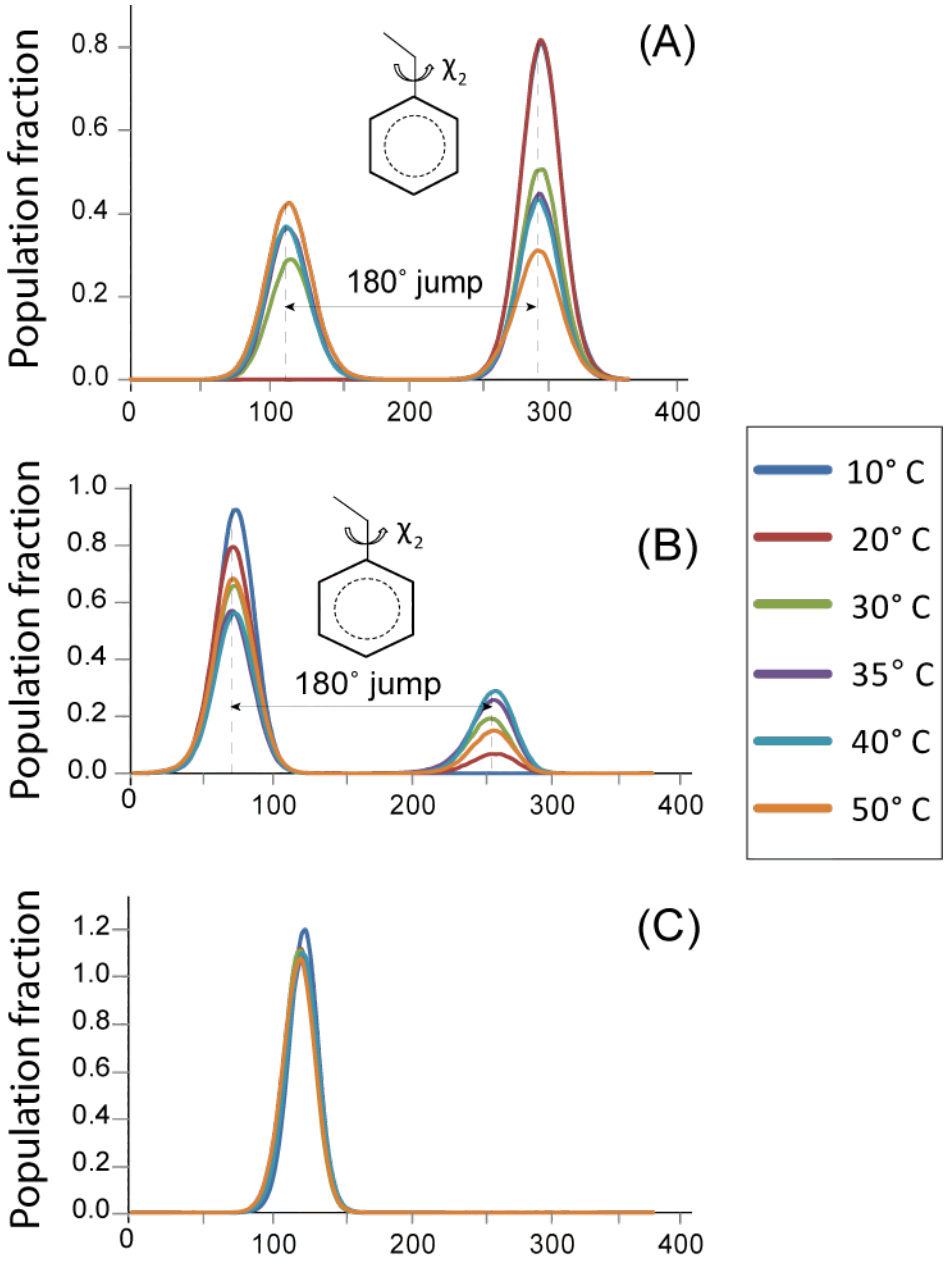


Figure 4.6: The torsion angle population distribution of the three aromatic amino acid side chains revealed ring flip motions occuring above 300 K for the two Phe residues. Tyr-59 however does not show ring flip motions in the MD simulations due to reasons described previously.

behaviour of the time correlation function [98] above the transition temperature of 300 K indicating the onset of ring flip. Moreover the time correlation for liberational motion is really fast with $\tau < 1ps$. Ring flipping is several orders of magnitude slower than liberational motions but only contributes significantly to relaxation at temperatures above 300 K. This provides an explanation for the temperature dependence of the internal correlation times of the aromatic ring $C_5 - H$ bond vector.

4.3 Examining the cooperative nature of the aromatic ring motions

Since the transition of all three aromatic rings in ubiquitin occur simultaneously and have the same midpoint of transition, the cooperative nature of the temperature dependence was examined from the MD simulations. For each flip of either Phe-4 or Phe-45, the time delay to the next closest flip of the other residue was tabulated. The mean delay time ($< \delta t >$) over the entire simulation was statistically indistinguishable from that expected from two completely independent stochastic events suggesting *no Phe-Phe correlation in ring flip dynamics*.

In a solid state view of a well-packed natively folded protein, increasing temperature promotes increased amplitude and frequency of packing defects, which would subsequently permit more frequent ring flipping. However, examination of both the overall packing efficiency and the packing efficiency of Phe-4 and Phe-45 side chains showed

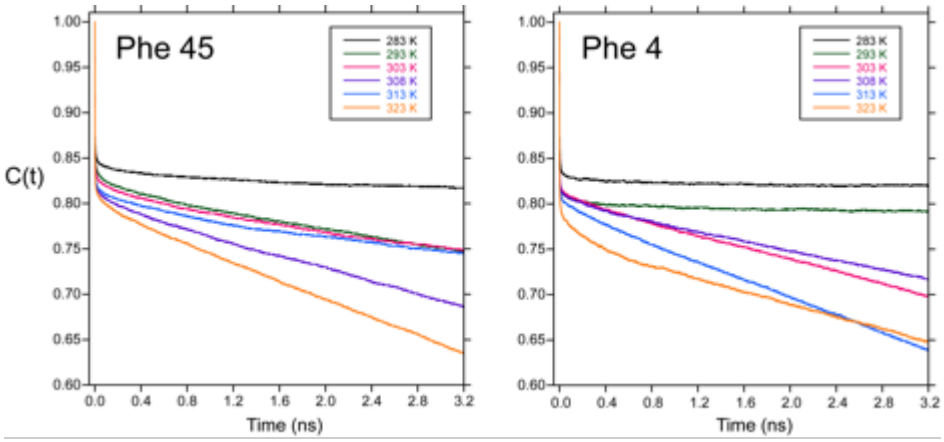


Figure 4.7: Time correlation function $C(t) =< P_2(u(t).u(t+\delta t)) >_t$ for angular motion of $C_5 - H$ bond vector u at various temperatures. P_2 is the second degree Legendre polynomial. T = 283, 293, 303, 308, 313, 318 and 323 K (indigo, blue, green, yellow, orange and red, respectively. Phe-4 is indicated by Panel A and Phe-45 by Panel B

no support for this model of ring-flip activation: *The temperature range used here had no significant effect on average packing or fluctuations in packing* (Figure 4.8).

To provide a more quantitative description of the changes in packing, the van der Waals contribution to the pressure virial [120], a more realistic measure of packing, was calculated according to:

$$v_{ij} = r_{ij}.U_{ij} \tag{4.1}$$

Where r_{ij}, U_{ij} are the distance and interaction potential, respectively between atoms i and j. The repulsive and attractive parts of the van der Waals terms of the pressure virial (Equation 4.1) revealed no significant change over the temperature range of 293 K to 303 K where the ring-flipping becomes activated (Figure 4.9).

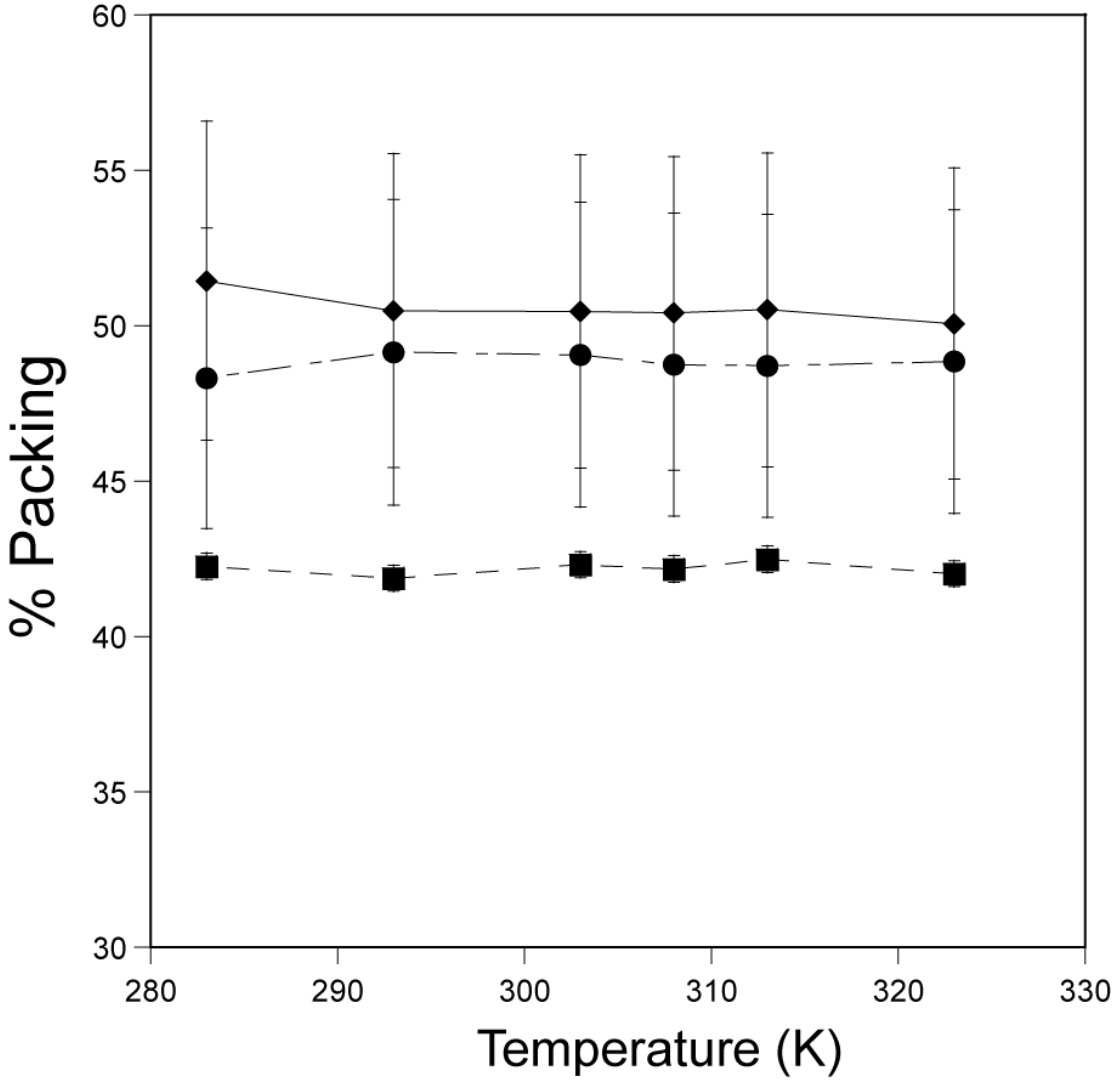


Figure 4.8: Percentage packing is defined as the ratio of the van der Waals volume to the Voronoi volume for Phe-4 (◆), Phe-45 (●) and Tyr-59 (■). Fluctuations in packing over the simulation are indicated by vertical bars: positive direction for Phe-45 and negative direction for Phe-4 are of the same magnitude but are omitted for clarity.

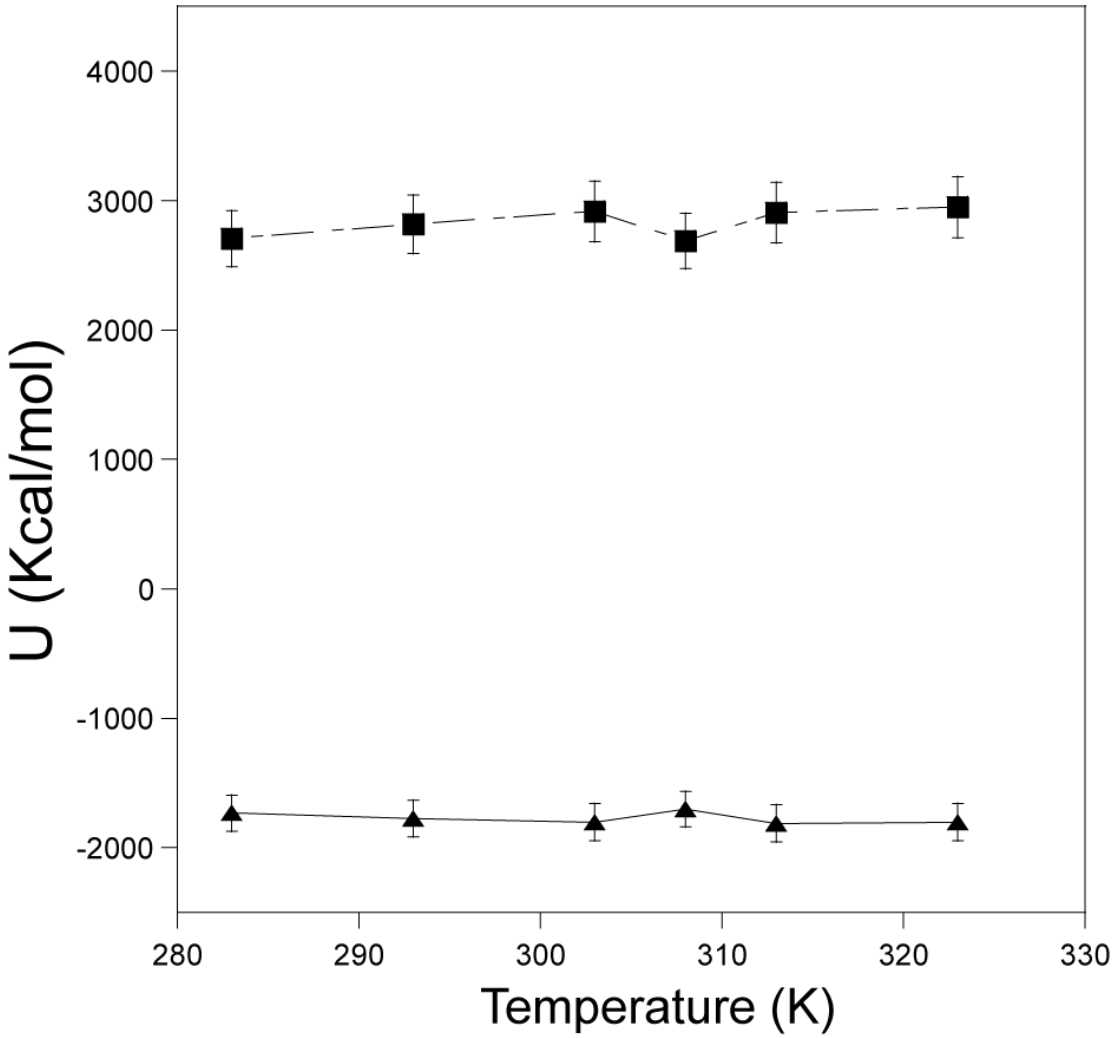


Figure 4.9: The repulsive and attractive terms of the van der Waals pressure virial are indicated by ■ and ▲, respectively.

The experimental measurement and MD simulations together support the model of aromatic rings undergoing independent stochastic ring flip (ps-ns) in a 'liquid-like' environment. This is starkly different to previous experimental characterization of aromatic ring flip in the slow time scales (μ s-ms) and suggest that aromatic amino acids possess sufficient degree of freedom within the protein milieu and that ring flip motions are not restricted to slow time scales.

4.4 Pressure induced fluctuations in aromatic dynamics

Application of hydrostatic pressure has been used previously to investigate the volumetric properties of slow ring flipping processes in proteins [109, 110]. In the few cases examined, the associated activation volumes are large and range from $\sim$ 30-50 mL mol^{-1}. In distinct contrast, the application of hydrostatic pressure results in very minor changes to the fast ring motion observed in ubiquitin (Figure 4.10). Below the thermal transition, the order parameter vary little and correspond to reduction in effective restricted diffusion angles ($< 3°$). The corresponding effective correlation times are similarly insensitive to the applied pressure. Above the thermal transition, where fast ring flipping or rotation is present, the pressure sensitivity of the order parameters of all three rings is significant.

It is important here to remember that the order parameter is a measure of equilibrium fluctuations [18]. Thus its sensitivity to pressure indicates the presence of a significant

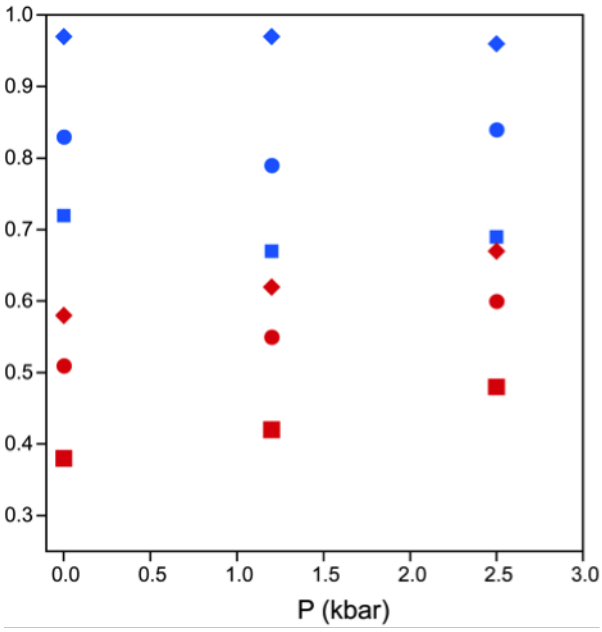


Figure 4.10: The pressure sensitivity of aromatic ring motions in ubiquitin above and below the thermal transition temperature for ring flipping is shown above. Lipari-Szabo model-free squared generalized order parameters of the $C_5 - H$ ring bond vectors of Tyr-59 (■), Phe-4 (◆) and Phe-45 (●) determined by ^{13}C-relaxation at 283 K (blue symbols) and 323 K (red symbols).

contribution of large amplitude continuous diffusion to the high temperature motion. A pure jump-like ring flip motion interconverting two identical (for Phe) and nearly identical (for Tyr) states would be predicted to be insensitive to pressure.The associated effective correlation times for ring motion above the dynamical transition temperature are also quite responsive to hydrostatic pressure. However, the interpretation of the effective correlation is fraught with qualifications [18]. The effective correlation time depends both on the microscopic diffusion or jump constants **and** the spatial nature of the motion. Furthermore, for the diffusion motion evident here, the internal correlation function is potentially defined as an infinite sum of exponentials [18, 70] and therefore

cannot be associated with a simple rate constant. Nevertheless, the increase in effective correlation times is consistent with the thermally induced motions being somewhat slower than the more restricted motion seen below the transition temperature.

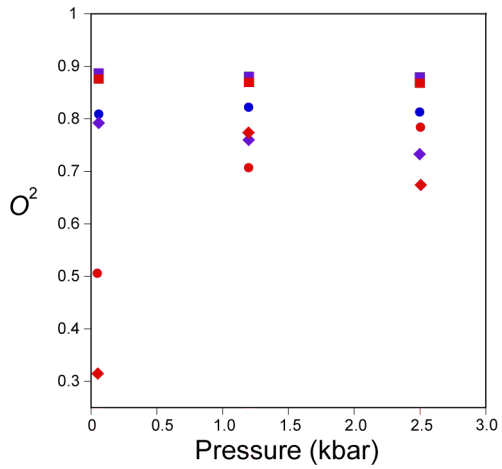


Figure 4.11: The pressure sensitivity of aromatic ring motions in ubiquitin above and below the thermal transition temperature for ring flipping is shown above. Lipari-Szabo model-free squared generalized order parameters of the $C_5 - H$ ring bond vectors of Tyr-59 (■), Phe-4 (◆) and Phe-45 (●) determined by ^{13}C-relaxation at 293 K (blue symbols) and 323 K (red symbols).

MD simulations reproduce the qualitative effects of pressure on the Phe order parameters well (Figure 4.11). Below the transition temperature, the order parameters are high and pressure has no significant effect . Above the transition temperature, the order parameters increase sharply as the pressure is increased to 2500 bar. This increase is not accompanied by any significant change in packing of the protein, globally or around the Phe residues, nor by any changes in the volume fluctuations.

Consideration of the pressure sensitivity of the order parameters indicates that the thermally induced dynamical transition involves the onset of a qualitatively distinct large amplitude but diffusive rotation of the aromatic rings. This view is in contrast to that seen in slowly interconverting systems where ring rotation is characterized by discrete jumps, high activation volumes and large activation enthalpies [109, 110]. The diffusive large scale motions identified here clearly indicate that the protein interior is more liquid-like than perhaps previously appreciated. Previously in collaboration with Dr. Yinan Fu in the Wand lab, I had characterized the pressure dependence of the motions of the methyl groups in proteins [116]. This study revealed a spatial clustering of the pressure dependence of methyl motions and quite surprisingly revealed that the motions of a subset of methyl containing amino acids exhibited non-linear pressure dependence. In contrast, the measurements here indicate that the aromatic amino acids show a simple linear pressure dependence. Intriguingly, the methyl residues which show non-linear pressure dependence are localized near the aromatic residues in ubiquitin (Figure 4.12). This raises the question about the role played by aromatic amino acids in the non-linear pressure response of methyl bearing amino acids side chains.

While the correlation between the motions of the aromatic amino acids and the proximal methyl are hard to understand in great detail from just one study, it remains suggestive of the liquid-like environment of protein interior. This would prove particularly important in the context of the role played protein conformational entropy in modulating protein-ligand binding and activity which are addressed in the next two chapters.

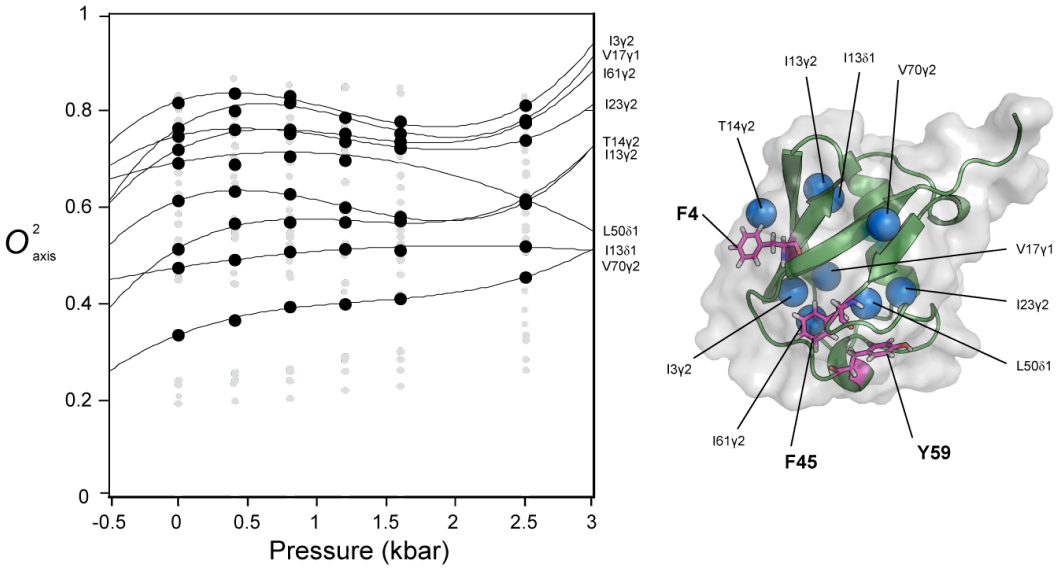


Figure 4.12: The methyl probes exhibiting non-linear pressure dependence of fast motions are located in close proximity (< 5Å) to the aromatic residues in ubiquitin. (Left Panel) The residues exhibiting non-linear pressure dependence are shown as black spheres and ones with simple linear pressure dependence are indicated as grey spheres. (Right Panel) Relative localization of the methyl groups (shown as spheres) exhibiting non-linear pressure dependence and the aromatic amino acids in ubiquitin (shown as sticks).

4.5 Summary and Outlook

Aromatic amino acids are known to play important role in molecular recognition and protein function [87, 90–92]. These however have so far been predominantly understood from a structural view point wherein aromatic amino acids contribute to the van der Waals stacking interactions and hydrogen bond interactions [28]. Previous NMR-based measurements of aromatic motions have revealed largely restricted motion with ring flipping occuring in the slow time scales (μs-ms) [93]. Using the recently introduced ^{13}C isotopic enrichment schemes (Chapter 3) aromatic amino acids for the first

time have been shown exhibit rich dynamic behavior in the ps-ns timescale [26]. Here the temperature dependence of the fast motions of the aromatic rings in ubiquitin have been investigated to shed light on the origins of the fast motions of aromatic rings. The temperature dependence of the motions of aromatic rings reveal a sharp transition with a midpoint close to the physiological temperature (310 K). At low temperature, motions of the aromatic ring consist only of restricted liberational motions representative of the high order paramter values measured where as beyond the thermal transition temperature the onset of ring flipping results in the decrease in NMR-measured order parameter. This is further corroborated my MD simulations which suggest that the effective correlation times for liberational motions are very low ($< 1ps$) whereas those associated with ring flip is slow resulting in the increase in effective internal correlation time with increasing temperatures. The pressure sensitivity of the motions of the aromatic ring above the thermal transition temperature suggests the presence of large amplitude continuous diffusion rather than a pure jump-like motion as reported in previous studies. The diffusive motion evident here is more representative of the motions of aromatic amino acids as jump-like motions of the aromatic ring have been seen by NMR in only a handful of cases [109, 110]. Furthermore, all three aromatic amino acids in ubiquitin transition close to the physiological temperature. The significance and generality of this is currently being investigated with NMR-relaxation measurements of aromatic ring motions in calcium-saturated calmodulin bound to a peptide fragment of the smooth muscle myosin light chain kinase [53]. Since vertebrate calmodulin is also active at

the same physiological temperature as ubiquitin (310 K), the observation of a thermal transition in calmodulin would shed more light on the significance of the transition temperature being so close to the physiological temperature. Moreover, future studies will also be carried out on the motions of aromatic rings in extremophiles [121] which would expand the scope and generality of such a transition.

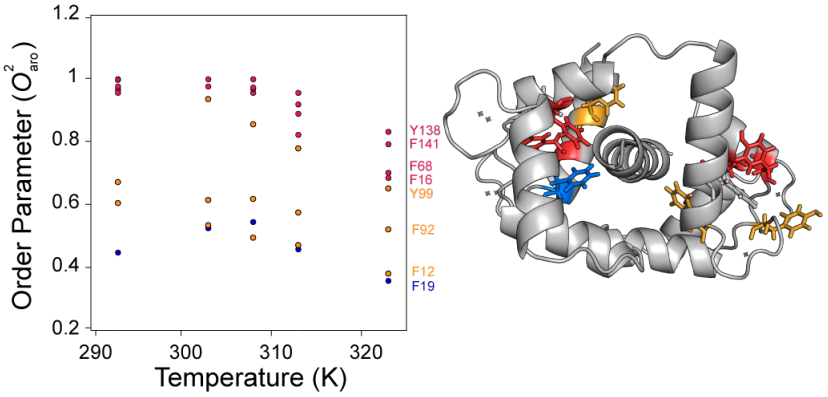


Figure 4.13: (Left panel) Lipari-Szabo model-free squared generalized order parameters of the $C_5 - H$ ring bond vectors of calmodulin determined by ^{13}C-relaxation. (Right Panel) The spatial location of the aromatic amino acids in the $Ca^{2+} - Calmodulin - smMLCK$ complex. Colored in (red) are residues which show a transition-like behavior, (orange) display linear response and (blue) displays a complicated temperature response with increasing rigidity between 293 K - 308 K and subsequently increased dynamics above 308 K.

The temperature response of the aromatic amino acid side chains in calmodulin is more complicated than ubiquitin and the preliminary data above (Figure 4.13), while containing hints of a transition-like behavior for some of the aromatic amino acids, is not conclusive. Further studies are currently being performed to understand this highly heterogeneous and complicated temperature response of the dynamics of aromatic amino acid side chains in $Ca^{2+} - Calmodulin - smMLCK$ complex.

The motions of methyl bearing amino acids have been used to quantify protein conformational entropy (Chapter 2) and has been shown to play a vital role in modulating the binding affinity of protein-ligand interactions [14, 24]. However, the role played by aromatic amino acids remains largely unknown. In Chapter 3 & Chapter 4 I have presented intriguing insights into the nature of aromatic side chain motions which has helped shed light on the underlying thermodynamics. Taken together these pave way for a detailed understanding of the intricate relationship between the motions of aromatic amino acids within proteins and potential their role in modulating/regulating protein function. In particular, the interfaces of very high-affinity protein-ligand complexes consist of many aromatic residues due to their role in both stacking and hydrogen bonding interactions [28]. Therefore NMR-measurements of methyl dynamics alone would not suffice to understand the thermodynamics governing very high-affinity interactions. In the next chapter, the contributions of both methyl- and aromatic side chain dynamics to the very high-affinity barnase-barstar [122] interaction are investigated. Furthermore, a detailed analysis of the contributions of different entropic components to the Gibbs free energy of binding is presented shedding light on a previously uncharacterized role for conformational entropy in very high-affinity binding interactions. The contributions of motions of the methyl bearing amino acid side chains together with those of the aromatic side chains shed light on how nature accomplishes such high affinity and highly selective binding interaction.

Chapter 5

Conformational entropy in very high-affinity interactions

Sections of this chapter pertaining to histamine-binding protein were contributed by Jackwee Lim, Postdoctoral fellow in the Wand Lab.

5.1 Introduction

The current emphasis of pharmaceutical drug design has primarily relied on structure based optimization of the binding affinity to target proteins of interest [123]. The emergence of vast database of molecular structures of proteins and their complexes has allowed for a detailed analysis of their structural origins including the structural role of

solvent water [6–8]. The view that has emerged is one where the design of pharmaceutical drugs have largely relied on gaining favourable changes in solvent entropy and subsequent optimization to gain higher enthalpy of binding [123]. While this strategy has resulted in the design of high affinity drugs, their numbers have been restricted to a handful [123]. Moreover, due to the nature of protein-small molecule ligand/drug interaction, i.e. the large hydrophobic interaction contribution resulting in favourable changes in solvent entropy, their specificity has proven difficult to control. An understanding of the biophysical origins of high affinity interactions is hence essential, especially with the emergence of pharmacological intervention strategies based on the disruption of protein-protein interactions [124, 125].

Specific molecular interactions are a hallmark of complex cellular processes and often involve very high affinities. A dominant view of high affinity interactions involving proteins is that they are largely driven by favourable interactions at the interface of binding and additionally by an increase in solvent entropy due to the creation of a dry or solvent depleted interface [7, 126–128]. High affinity protein-protein interactions are critical to the stability of the complex biochemistry and are wide-spread in the proteome. Moreover, these are very important drug targets, most notable perhaps being antibody-antigen interactions and cell-surface signalling interactions ($K_d \sim 10^{-12}M$) [129–131]. A prominent issue is the role of conformational entropy in the formation of very high affinity protein-ligand complexes. However, the contributions of protein conformational entropy have been difficult to measure experimentally. It has been recognized for some

time that the motion between different states of a protein reflects its residual entropy and that NMR spectroscopy could provide a means to connect measures of motion to measures of this entropy [11, 17]. The quantitative conversion of fast internal motion measured by NMR relaxation methods to measures of conformational entropy without undue model-dependence or assumptions has recently been used to elucidate the role of the latter in tuning the binding affinity of protein-ligand interactions of moderately high affinity ($K_d = 10^{-6} - 10^{-9}M$) [14, 27, 107]. In these few initial examples, conformational entropy largely contributes unfavourably to ligand binding. However, the role of conformational entropy in very high affinity complexes ($K_d < 10^{-9}M$) still remains completely unclear.

Dissecting the thermodynamics of binding (Equation 5.1) for very high affinity protein-ligand interactions would prove valuable in understanding the relative contribution of the different thermodynamic parameters contributing to binding.

$$\Delta G_{binding} = \Delta H_{binding} - T\left[\Delta S_{solvent} + \Delta S^{protein+ligand}_{conformational} + \Delta S^{protein+ligand}_{RT}\right] \tag{5.1}$$

In this chapter, a detailed analysis of the contributions of not only protein conformational entropy but also that of changes in solvent entropy is presented for multiple high affinity protein-ligand complexes. On the other extreme, the extensive characterization of the role played by entropy of binding in weak affinity interactions is also presented. Taken

together, a general paradigm for achieving very high affinity protein-ligand interaction and its implications are discussed.

5.2 Thermodynamics of protein-ligand systems: barnase-barstar; HBP(D24R)-histamine

In this study, four protein-ligand complexes are examined to understand in quantitative detail the contributions of both conformational entropy and solvent entropy to the Gibbs free energy of binding. The complexes studied here span a wide range of binding affinities and would aid in understanding the nature of conformational entropy contributions to the different binding interactions.

5.2.1 Very high affinity: barnase-barstar

To initially address this issue, the complex between barnase and its natural high affinity protein inhibitor barstar ($K_d = 10^{-14}M$) [122, 126] and the interaction of a small molecule ligand histamine with the histamine-binding protein (HBP(D24R)) is examined ($K_d = 2.5x10^{-9}$ M) [132, 133]. Barnase is a 110-residue extracellular ribonuclease of *Bacillus amyloliquefaciens* [134] and is inhibited by the 89-residue barstar [135] to suppress its potentially lethal ribonuclease activity inside the cell. The barnase-barstar interaction,

one of the tightest known interactions in nature exhibits little structural perturbation upon formation of the complex (Figure 5.1) [126].

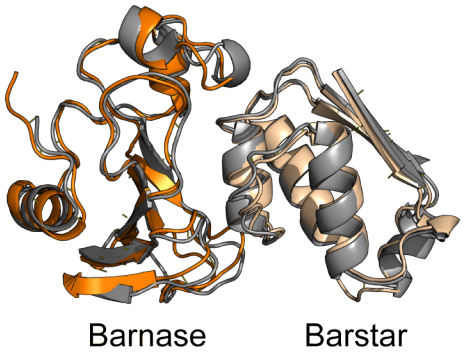


Figure 5.1: Overlay of the free and bound states of barnase and barstar, respectively. The average C_α r.m.s.d of the superposition is 0.9Å. PDB code barnase-barstar complex: 1BRS, PDB code barnase free: 1BNR, PDB code barstar free: 1BTA

The thermodynamics of barnase binding to barstar has been extensively characterized previously by isothermal calorimetry (ITC) experiments (Table 5.1) [126]. This has established that the binding of barnase to barstar is driven by a very favourable enthalpy with little net contribution by the total entropy of binding [126, 136]. Structural analysis of barnase and barstar in the free and bound states has revealed that the interface is "wet" (Figure 5.2), which raises the possibility that the solvent binding entropy may be unfavourable [137] yet the barnase-barstar complex is of very high affinity ($K_d = 10^{-14}$ M).

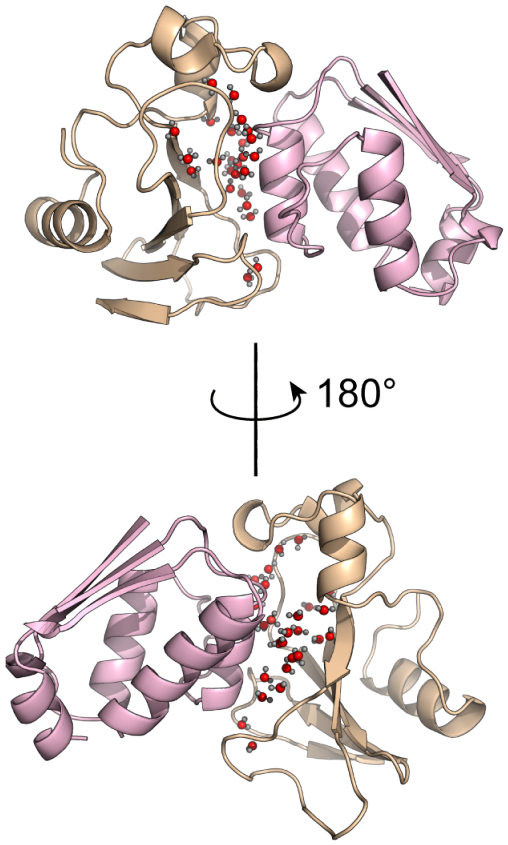


Figure 5.2: The structure of barnase-barstar consists of several water molecules at the interface of binding involved in hydrogen bonding interactions. PDB code: 1BRS

5.2.2 Very high affinity: histamine-binding protein:histamine

The histamine-binding protein studied here is a member of the lipocalin family isolated from the cattle-feeding brown-ear tick *Rhipicephalus appendiculatus* and is known to bind two histamine molecules, one with very high affinity ($K_d = 7x10^{-9}$ M) and the other much more moderate ($K_d = 6x10^{-8}$ M) [132, 133]. Histamine-binding protein (HBP) is a 171-residue beta-barrel protein which binds and sequesters histamine, a primary mediator of inflammatory reactions upon tissue damage. Wild type HBP is

also known to bind serotonin, another small molecule ligand, at the secondary binding site [138]. Not surprisingly, serotonin is a known mediator of inflammatory reactions in rodents. The binding affinity of wild type HBP for serotonin at the secondary binding site is higher than for histamine [138]. The histamine binding protein from *Rhipicephalus appendiculatus* is one of the most evolved members of the lipocalin family and their ability to bind multiple small molecule ligands simultaneously could potentially reflect host adaptation [132]. However, investigations into this allosteric mechanism is beyond the scope of the current study involving the role played by conformational entropy in very high affinity interactions. In order to simplify the interpretation of the role played by conformational entropy, arising from this 2:1 stoichiometry, the secondary histamine binding site is mutated (D24R) to retain just the high affinity histamine binding site. The structure of apo HBP and HBP(D24R) bound to histamine is almost identical with minimal structural perturbations (Figure 5.3) [133]. The binding affinity determined by isothermal calorimetry (ITC) of HBP(D24R) binding to histamine is still very high affinity with a K_d ($= 2.5x10^{-9}$ M) similar to that of the wild type HBP-histamine interaction (Figure 5.3, Table 5.1).

Histamine binding is mediated by an extensive network of electrostatic and hydrogen bond interactions. Four water molecules mediate some of the hydrogen bond interactions (Figure 5.4) hinting at a potential unfavourable solvent entropy contribution [133] to binding albeit much smaller than for barnase-barstar which displays approximately 40 water molecules at the interface of binding.

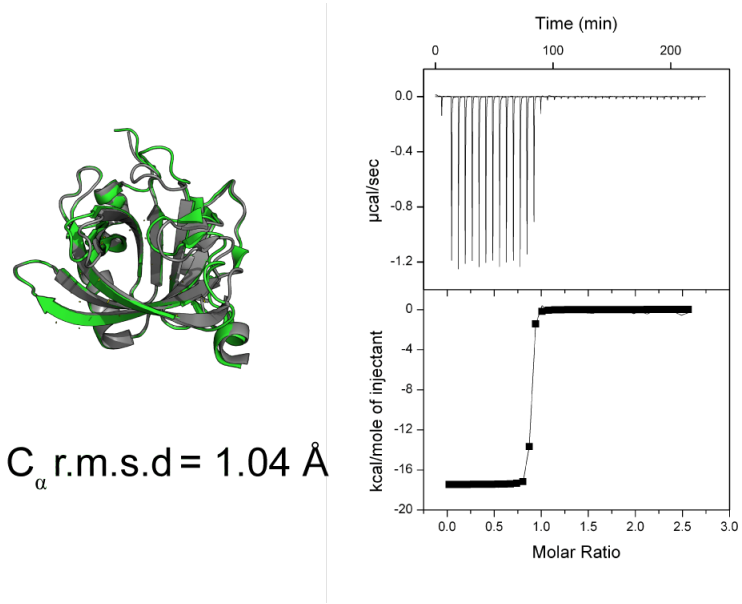


Figure 5.3: Left panel represents the overlay of free and bound structures of histamine binding protein bound to histamine. Right panel corresponds to the isothermal calorimetry binding curves for histamine binding to HBP(D24R). PDB code HBP free: 3GAQ, PDB code HBP-histamine: 3G7X

5.2.3 Moderate affinity: barnase-dCGAC

The natural substrate for barnase is ribonucleic acids (RNA). However since barnase is known to actively cleave RNA strands upon binding, DNA has been used as a model substrate to understand the binding of barnase to nucleic acids [139]. Previously, barnase has been shown to bind both 4mer single stranded DNA as well as dinucleotides [139]. Here, the binding of barnase to 4mer DNA CGAC is investigated. Barnase binds to dCGAC with moderate affinity ($K_d = 5x10^{-5}M$) and is almost completely driven by entropy with little or no enthalpy changes (Figure 5.5). Moreover, barnase experiences very little structural perturbations upon binding dCGAC (Figure 5.5) [139].

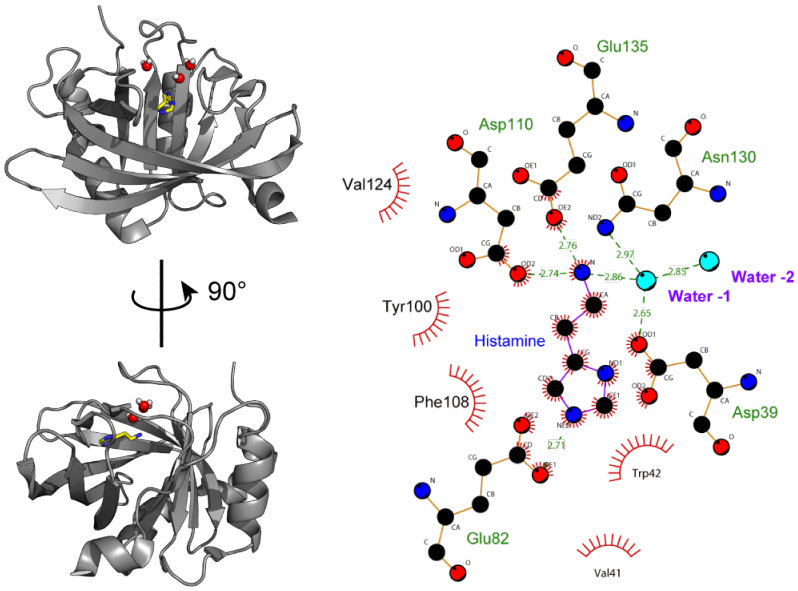


Figure 5.4: On the left, two orientations of the HBP(D24R)-histamine 3D structures are shown. On the right, the network of interactions both electrostatic and hydrogen bond mediated by two of the four water molecules are shown (generated by LigPlot).

The binding of barnase-dCGAC is accompanied by minimal heat change (Figure 5.5) as observed via ITC experiments. Therefore, the dissociation constant was measured using NMR chemical shift titration (Figure 5.6). The dissociation constant values obtained via ITC and NMR titration agree well.

Protein conformational entropy could be playing a favourable role in both enzyme-inhibitor, barnase-barstar interaction (negligible total entropy of binding with potential unfavourable solvent entropy change) and enzyme-substrate, barnase-dCGAC complex (completely entropy driven).

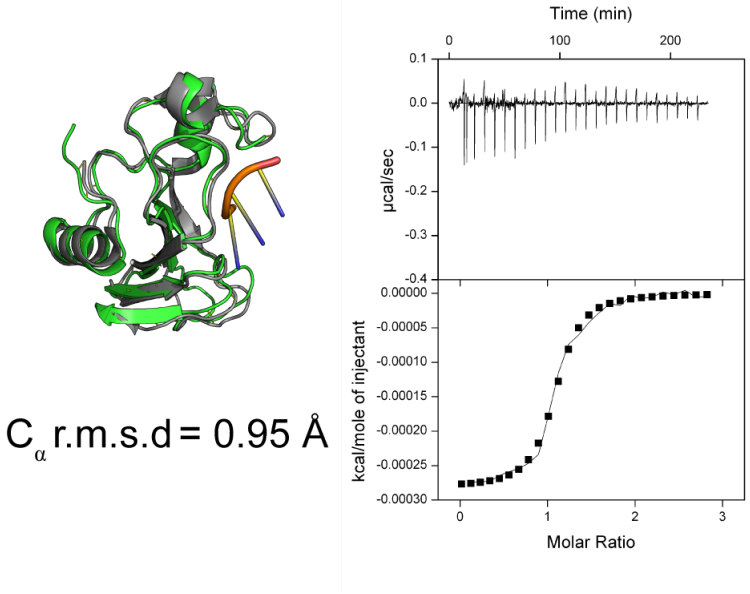


Figure 5.5: The overlay of free barnase and barnase-dCGAC three dimensional structures indicate minimal structural perturbations. The ITC measurements of barnase-dCGAC interaction suggest negligible heat change making K_d estimation via ITC unreliable. Independent K_d measurement was performed using NMR chemical shift perturbation experiment. PDB code barnase free: 1BNR, PDB code barnase-dCGAC complex: 1BRN

5.2.4 Weak binder: HBP(D24R)-serotonin

While serotonin has been shown to bind wild type HBP at the secondary binding site [138], serotonin binds to HBP(D24R) at a remote binding site distal ($\sim$ 10 $\mathring{\text{A}}$ away)to both the primary and secondary binding histamine binding sites. NMR chemical shift titrations (Figure 5.7) and ITC experiments (contributed by Dr. Jackwee Lim) show that the binding is driven completely by entropy with no enthalpic contribution. The binding affinity for serotonin to HBP(D24R) is estimated to be 1.89 mM.

Intriguingly, serotonin binds to the same remote distal binding site of HBP(D24R) in the

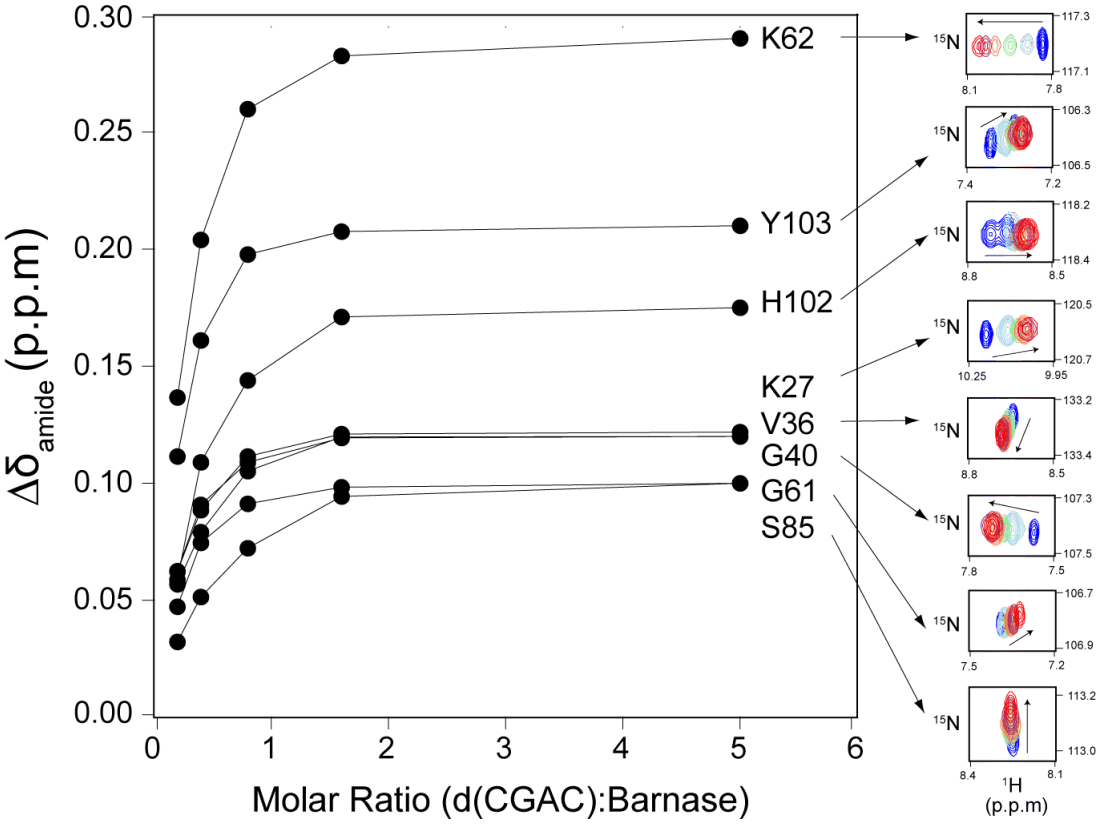


Figure 5.6: The dissociation constant of barnase-dCGAC interaction was estimated by fitting the NMR backbone chemical shift perturbations incurred upon 4mer DNA CGAC titration. The residues which are incur the maximum chemical shift changes are shown on the right. The global average K_d values obtained via fitting match very closely to the K_d value obtained from the residues shown here. The K_d values obtained here agree very well with previously published estimates. The concentration of barnase used was 0.25 mM.

presence of histamine; albeit at a lower binding affinity ($K_d \sim 5mM$). In both cases, the binding interaction of serotonin is completely entropy driven (Figure 5.7). *What role does conformational entropy play in modulating the binding affinity of serotonin?*.

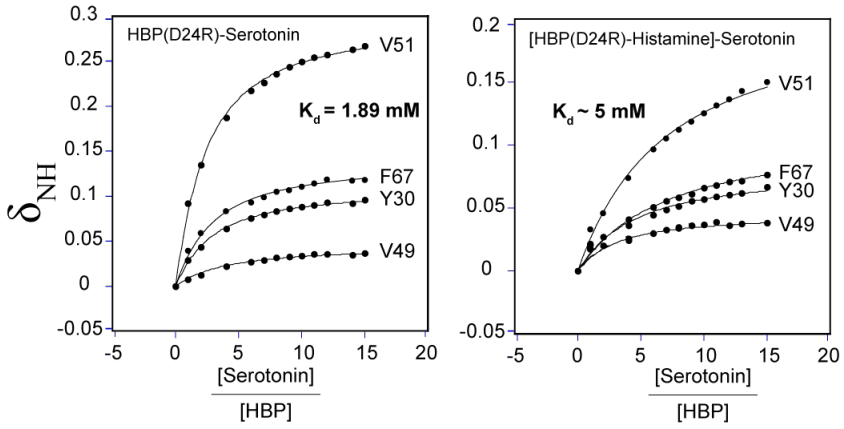


Figure 5.7: The dissociation constant of HBP(D24R)-serotonin interaction was estimated by fitting the NMR backbone chemical shift perturbations incurred upon serotonin titration in the absence of histamine (left panel) and to histamine prebound HBP(D24R) (right panel). The residues indicated incur the maximum chemical shift change upon serotonin binding and are used for fitting the dissociation constant.

5.2.5 Summary

The thermodynamics of binding for the different protein-ligand complexes are listed in Table 5.1. While ITC experiments can provide quantitative measures of the enthalpy of binding and the total entropy of binding, the different contributions to the total entropy of binding (conformational, solvent and rotational/translational) cannot be resolved.

Table 5.1: Thermodynamics of binding of the different protein-ligand complexes

Protein:Ligand	$K_d(M)$	$\Delta H(kcalmol^{-1})$	$-T\Delta S\,(kcalmol^{-1})$
Barnase:Barstar	$1x10^{-14}$	-19	-0.1
Barnase:dCGAC	$5x10^{-5}$	-1.0	-4.87
HBP(D24R)-Histamine	$2.5x10^{-9}$	-17.5	5.75
HBP(D24R)-Serotonin	$1.89x10^{-3}$	-	-3.72
(HBP(D24R)-Histamine)-Serotonin	$4.5x10^{-3}$	-	-3.20

In this study, NMR-relaxation methods probing the motions of the methyl bearing amino

acids are used as a proxy to provide quantitative estimates of the change in protein conformational entropy upon ligand binding (more details about the method can be found in Chapter 2 and Chapter 7). In addition, the high affinity barnase-barstar interaction and the barnase-dCGAC interaction consist of an abundance of aromatic amino acids at the interface. With the new ^{13}C isotope labelling scheme introduced for aromatic amino acids (Chapter 3) [26], the influence of the motions of the aromatic amino acids are probed. Together the motions of the methyl- and aromatic amino acid side chains will provide a comprehensive view of the dynamical landscape governing high affinity interactions. Additionally, using the recently introduced reverse micelle encapsulation of proteins [140, 141], the changes in hydration (i.e. change in solvent dynamics) between the free and bound states of barnase and barstar are studied to provide a quantitative estimate for the change in solvent entropy.

5.3 Results

5.3.1 NMR chemical shift assignments

Previous NMR studies have already established the chemical shift assignments of free barnase and free barstar (both $^{15}N-^{1}H$ backbone and $^{13}C-^{1}H$ side chain methyl). The backbone and side chain methyl chemical shift assignments of barnase in the barnase-barstar complex has also been previously reported [142, 143]. In this study, we have re-characterized the chemical shift assignments of the backbone, side chain methyl

and side chain aromatic amino acids of free barnase, free barstar, barnase and barstar in the barnase-barstar complex as well as barnase bound to d(CGAC) (Figure 5.8, 5.9, 5.10, 5.11). The backbone, side chain methyl- and aromatic amino acid chemical shifts of barnase in the barnase-dCGAC bound state were determined by tracking the chemical shift changes upon dCGAC titration.

Similarly, in the case of HBP(D24R) the backbone amide chemical shifts were previously published and continuing from these the side chain methyl groups were assigned in both the apo and the histamine bound states (not shown).

5.3.2 Increase in conformational entropy upon barnase-barstar high affinity interaction

The NMR-relaxation measurements [17] of the backbone amide reveal that on average the changes in motion of the polypeptide backbone of barnase and barstar upon formation of the complex is small and reveals a general restriction of motion of the amide N-H bonds upon complex formation. The response of the individual sites is however somewhat heterogeneous (Figure 5.12).

While the changes in motion of the backbone is small, the motions of the methyl-bearing amino acid side chains of both barnase and barstar increase upon complex formation ($< O^2_{axis} >= -0.096 \pm 0.012$; n=90).

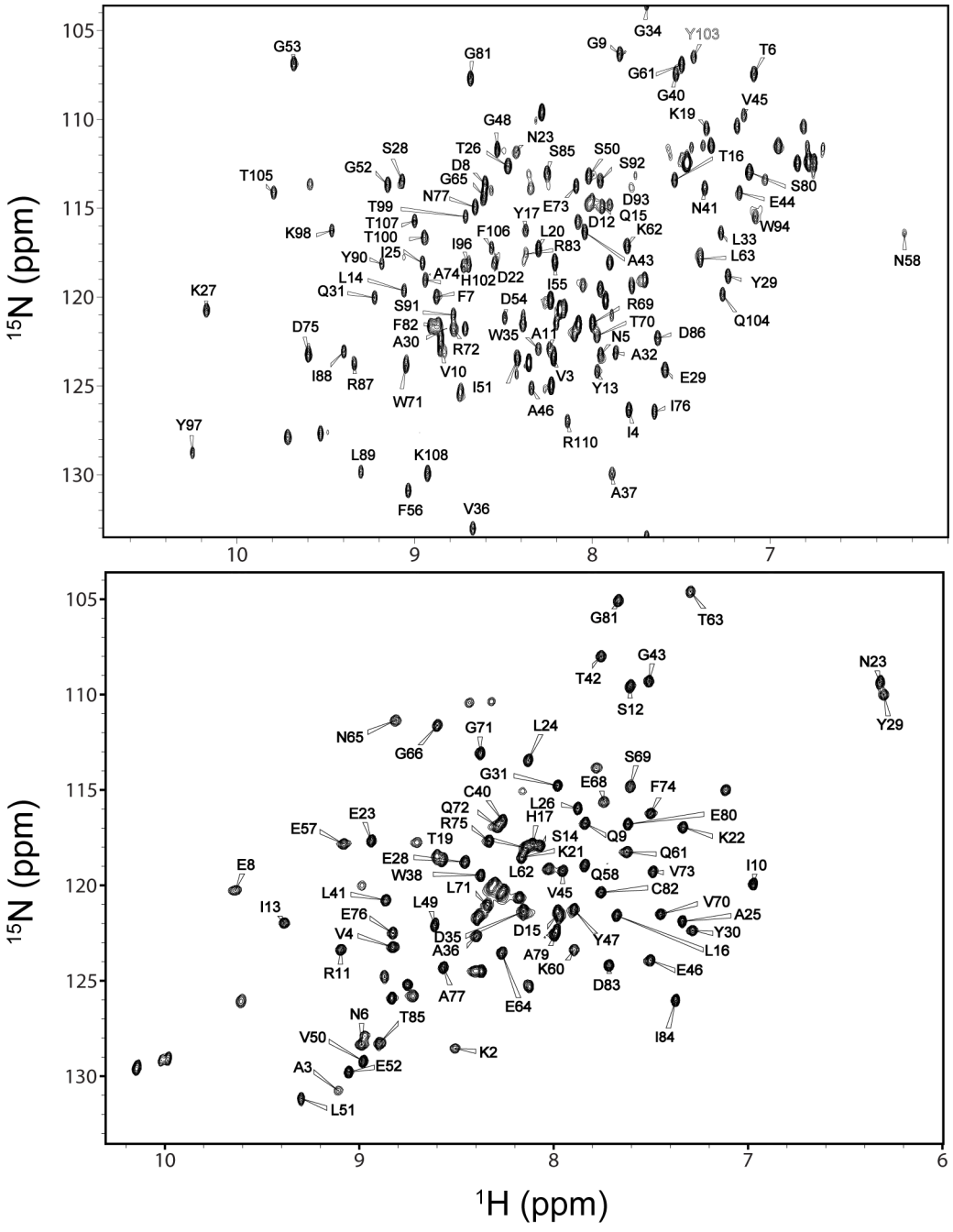


Figure 5.8: The ^{15}N backbone amide chemical shifts obtained for free barnase (top panel) and free barstar (bottom panel) using triple resonance experiments match with previously published assignments.

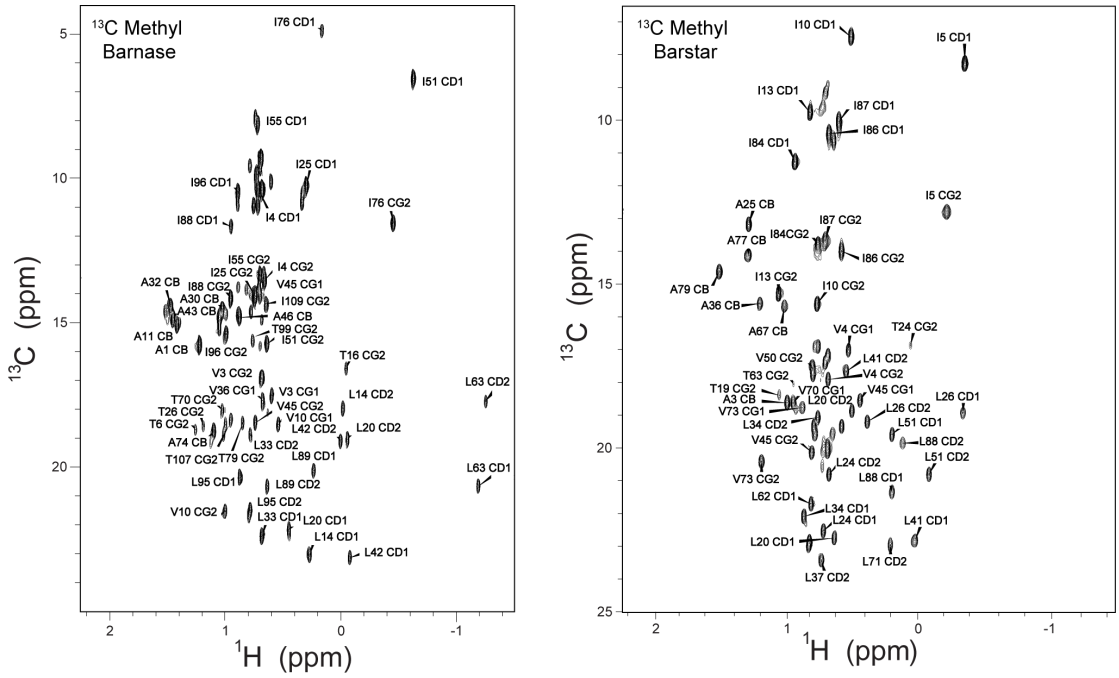


Figure 5.9: The ^{13}C-methyl chemical shifts obtained for free barnase (left panel) and free barstar (right panel) match with previously published assignments.

Comparison of ^{13}C-relaxation measured O^2_{axis} with previous ^{2}H-relaxation based measurements for free barnase and complexed barnase. The agreement between the methyl dynamics measurement for complexed barnase in this study to that measured previously is poor. This disagreement is a result of inaccurate characterization of macromolecular tumbling in the previous work where the molecular tumbling time was estimated based on the shape and hydrodynamic radius of the complex. It is however important to note here that the correlation of measured methyl dynamics in free barnase agrees reasonably well with the previous work because the molecular tumbling time for the free barnase was characterized using standard backbone NMR relaxation experiments. However, the measured methyl dynamics in complexed barnase differs significantly due to the inaccurate macromolecular tumbling estimate (Figure 5.13). Hence, the observed increase in side chain dynamics differs from that described in the previous study of free and complexed barnase.

The use of the dynamical proxy (Chapter 2) to convert the average change in motion

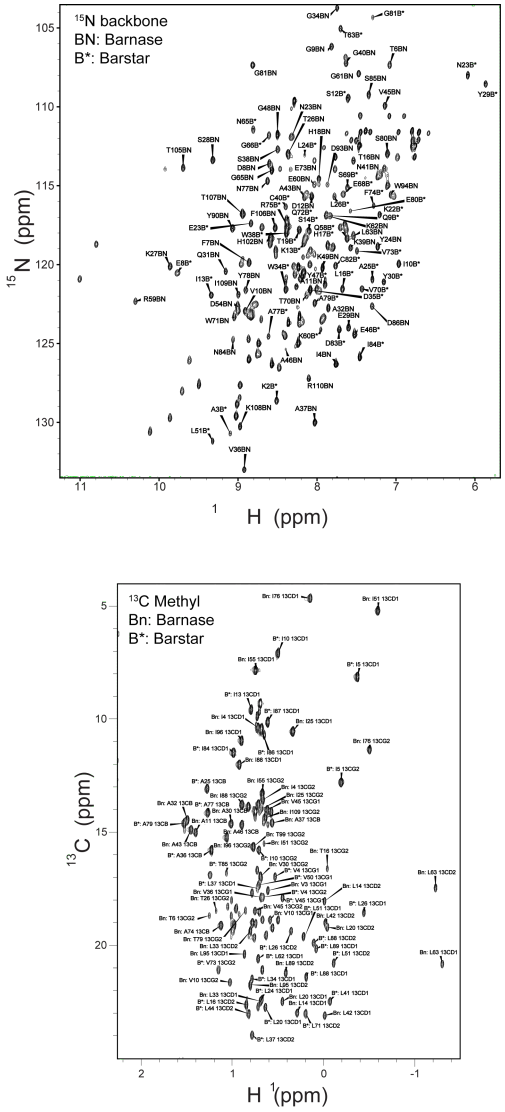


Figure 5.10: The ^{15}N backbone amide chemical shifts assignment (top panel) of barnase and barstar were assigned in the complexed state using triple resonance experiments. (bottom panel) The ^{13}C-methyl chemical shifts assignments obtained for barnase in the complex matched with previously published results. The ^{13}C-methyl chemical shift of barstar in the complex were assigned. Residues which were overlapped were not used for any of the analysis.

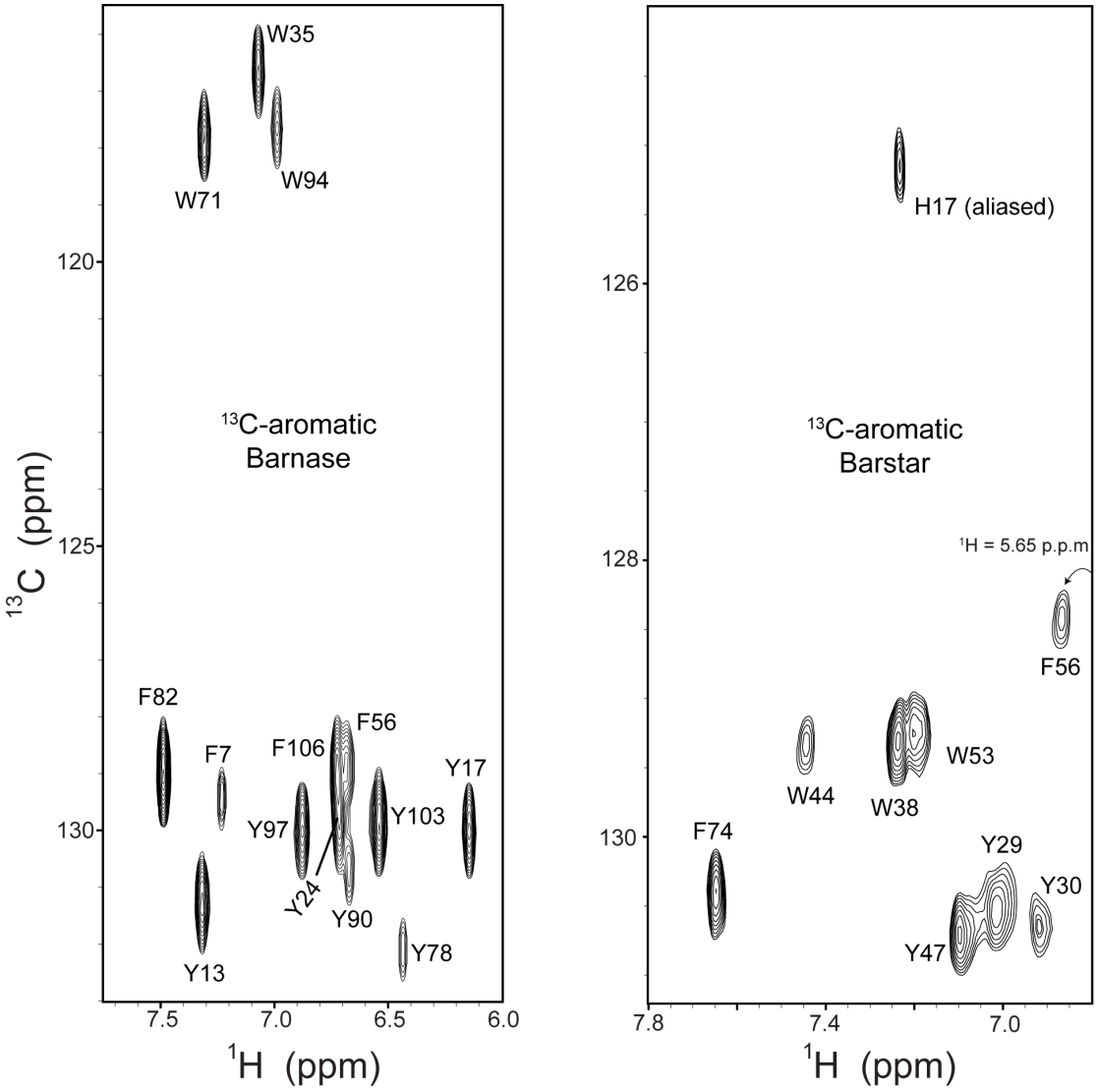


Figure 5.11: The ^{13}C-aromatic chemical shifts of CD1 of Phe & Tyr and CE3 of Trp obtained for free barnase (left panel) and free barstar (right panel).

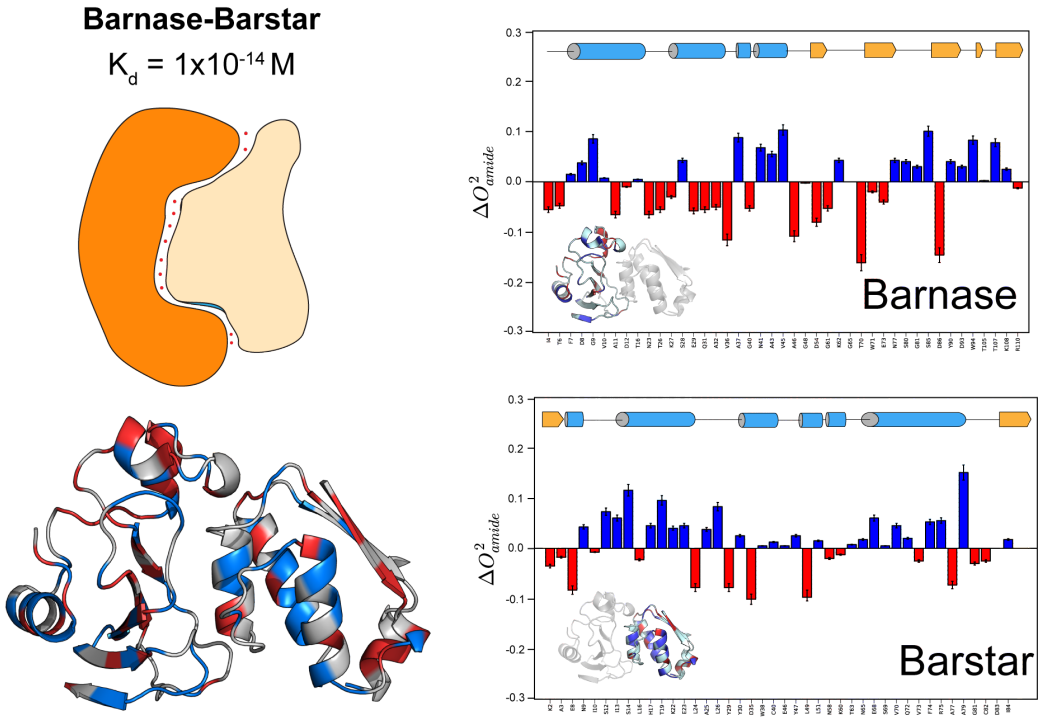


Figure 5.12: The changes in motion of the $^{15}N-^{1}H$ amide bond are mapped onto the structure of the barnase-barstar complex (left panel) and indicate that the changes are on average small. The changes in motion of the individual sites in both barnase and barstar are indicated as a bar graph on the right. Some of the backbone amides in both barnase and barstar do not show any change in order parameter, these are not indicated on the bar graph plot. PDB code barnase-barstar complex: 1BRS

of methyl-bearing side chains to changes in conformational entropy [14, 107] indicates that the conformational entropy ($-T\Delta S_{conf}$) of barnase and barstar is increased by 18.5 kcal mol^{-1} upon complex formation (i.e. $-T\Delta S_{conf}$ = -18.5 kcal mol^{-1}) and therefore contributes favourably to the high affinity binding free energy. The distribution of the dynamical response of side chains to binding is also complex. Methyl-bearing residues of barnase at the interface experience small changes upon binding ($< O^2_{axis}|_{interface} >=$

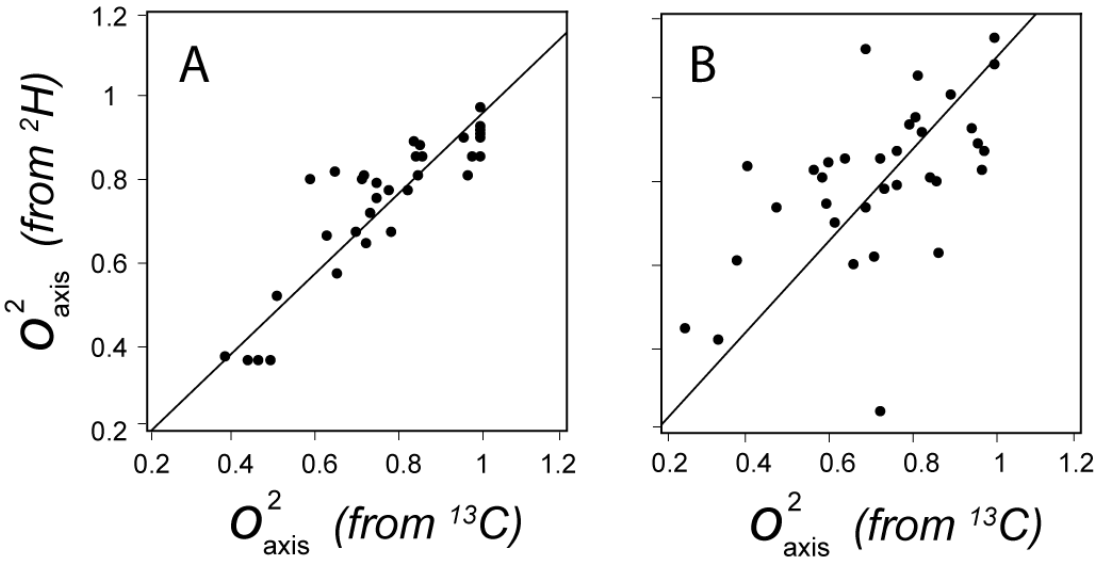


Figure 5.13: (A) The correlation of methyl order parameters for free barnase from the previous ^2H-relaxation measurement and the ^{13}C-relaxation measurement in this study agree reasonably well ($R^2 = 0.8$). (B) Poor correlation ($R^2 = 0.4$) of the methyl order parameters of barnase in the complexed state is observed between the ^2H- and ^{13}C-relaxation studies due to the inaccurate macromolecular tumbling characterization in the previous ^2H-relaxation study.

0.006 ± 0.003; n=5) while those removed from the interface show a comparatively larger increase in motional amplitude ($< O^2_{axis}|_{remote} >= -0.074 \pm 0.010$; n=45). Similarly, the methyl-bearing residues of barstar at the interface also experience small changes upon binding ($< O^2_{axis}|_{interface} >= 0.01 \pm 0.005$; n=3) while those removed from the interface show an even larger increase in motional amplitude than their counterparts in barnase ($< O^2_{axis}|_{remote} >= -0.123 \pm 0.012$; n=37) (Figure 5.14).

The barnase-barstar interface is particularly enriched in aromatic amino acid residues. Using the recently introduced isotopic labeling strategy that eliminates a variety of technical issues (Chapter 3) [26], we were able to characterize the fast motion of 5

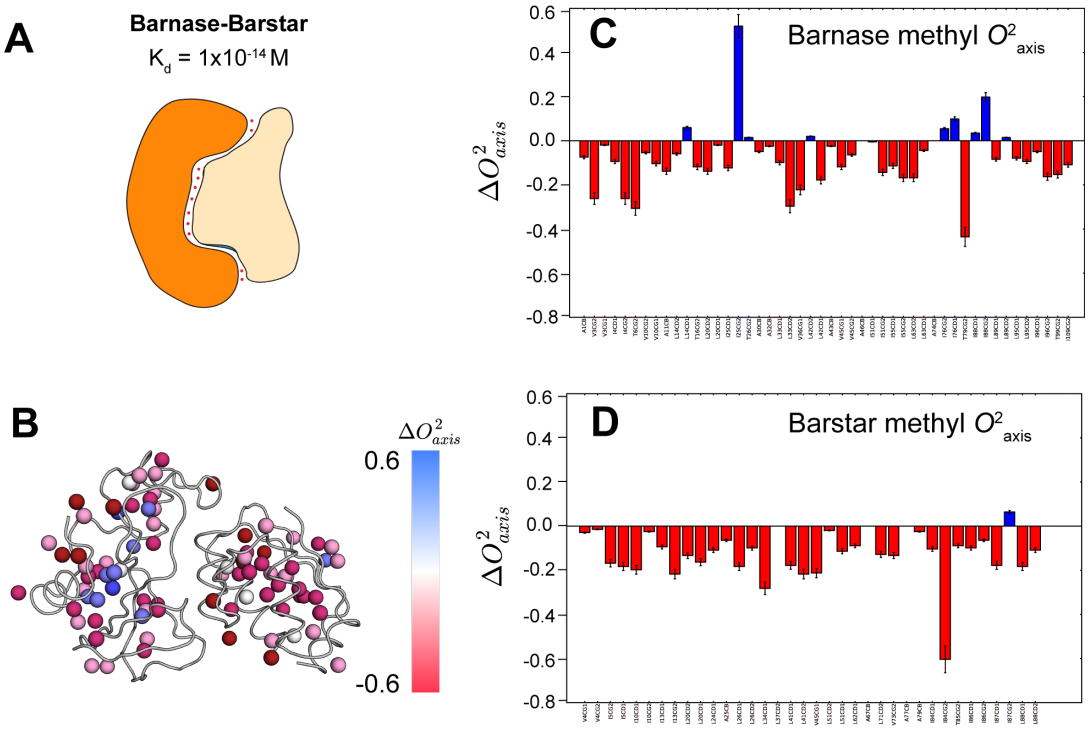


Figure 5.14: (A) Cartoon depiction of the barnase-barstar complex. (B) The changes in motion of the methyl-bearing amino acid side chains (ΔO^2_{axis}) in barnase and barstar upon complex formation iare mapped on to the structure indicating increase flexibility away from the interface and on average reduced flexibility at the interface of binding. The hydrophobic core of barnase which plays an important role in stabilitizing the beta-sheet surface involved in barstar and substrate binding displays reduced motion. Barstar displays an almost complete activation of dynamics indicating its high degree of optimization to bind barnase selectively. (C) & (D) represent the bargraph plot of the changes in order parameter for barnase and barstar, respectively.

Phe, 6 Tyr and 4 Trp aromatic rings. The sensitivity of the fast aromatic ring motions to binding of barstar to barnase mirrors that of the methyl-bearing side chains. We find that the motions of the aromatic side-chains at the interface, which are predominantly involved in base-stacking and hydrogen bond interactions, are quenched ($< O^2_{aro}|_{interface} >= 0.042 \pm 0.007$; n=9), while the motions of those remote from the interface are activated ($< O^2_{aro}|_{remote} >= -0.05 \pm 0.007$; n=6) (Figure 5.15).

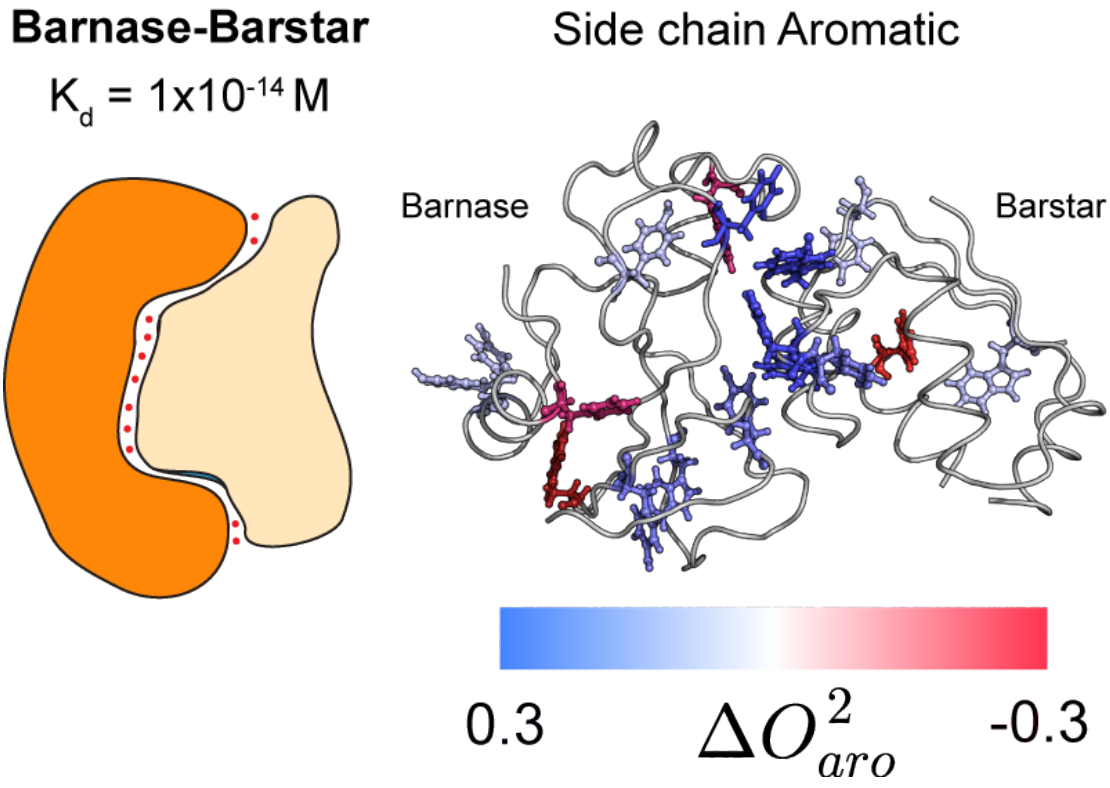


Figure 5.15: (A) Cartoon depiction of the barnase-barstar complex. (B) The motions of the aromatic amino acid side chains (O^2_{aro}) are lowered in the interface of binding while those remote from the binding interface show increased flexibility mirroring the behaviour of the methyl motions. PDB code barnase-barstar complex: 1BRS

5.3.3 Dynamics of the interface waters in barnase-barstar complex

The three-dimensional structure of the barnase-barstar complex indicates that the interface consists of about 40 water molecules involved in extensive water-mediated hydrogen bond interactions between barnase and barstar (Figure 5.16) [122]. Seven of these are completely buried, a further twelve waters have very low solvent accessibility and the rest exhibit partial solvent accessibility.

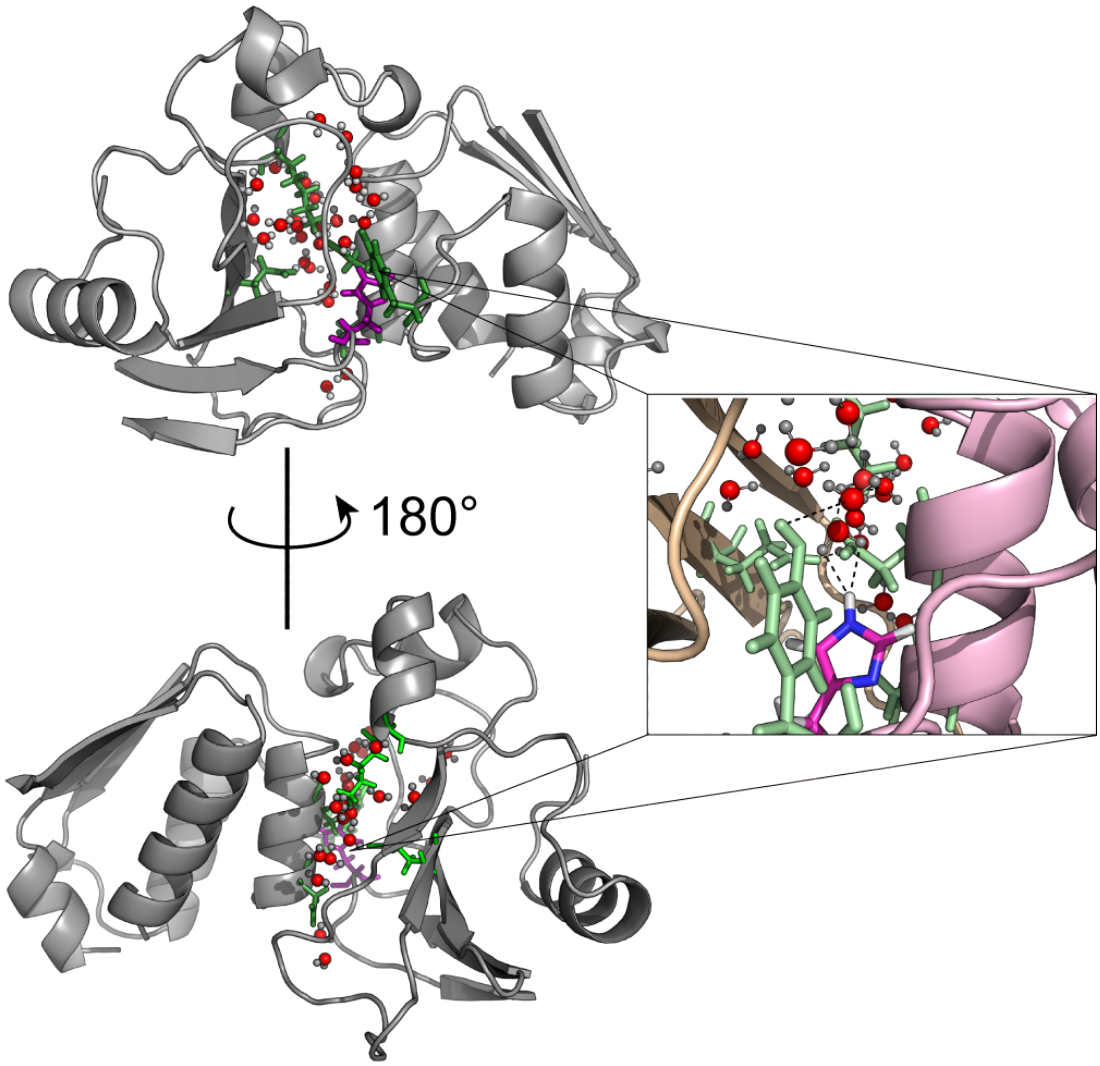


Figure 5.16: The relative location of the crystallographic waters and the important residues at the interface of barnase-barstar complex are indicated. Many of the interactions between barnase and barstar are mediated by a network of water-mediated hydrogen bonds. The catalytic His102 of barnase (indicated as purple) is surrounded by multiple water molecules held in position by a network of residues from barnase and barstar. PDB code barnase-barstar complex: 1BRS

Interface waters has the potential to contribute unfavourably to the binding entropy if its residual entropy is lower than that of the hydration water associated with the corresponding surfaces of the free proteins. The experimental measurements of the dynamics of water molecules across protein surfaces in aqueous solutions have proven difficult due to several technical considerations, in particular the low residence time of water across the protein surface and exchange with bulk solvent [140, 144]. The recently developed reverse micelle methodology involving encapsulation of proteins within the nanoscale water pool of a reverse micelle has been shown to be effective in overcoming these technical issues [141]. Application of this approach to the barnase-barstar complex reveals that the interfacial waters are relatively rigid whereas those at the corresponding surfaces of free proteins are generally quite mobile (Figure 5.17 and Figure 5.18).

Reverse micelle encapsulation of barnase, barstar and barnase-barstar complex. Recent investigations in the Wand lab have outlined several optimal encapsulation conditions i.e. surfactant mixture conditions under which different proteins have been stably incorporated within the nanoscale pool of the reverse micelle [141]. In the case of barnase, barstar and barnase-barstar complex, several of these conditions were first screened to determine the best set of conditions applicable to each protein. Barnase and barnase-barstar are more stable compared to barstar and can be made up to sufficiently high concentrations to enable the use of the direct injection method for reverse micelle preparation. Barnase, barnase-dCGAC and barnase-barstar were determined to be stable and folded with ^{15}N-HSQC spectrum in 75 mM CTAB/450 mM Hexanol with W_o = 20 matching excellently with the aqueous spectrum. The proteins were also suffciently stable under the same conditions making it ideal for hydration measurements using 3D-NOESY-HSQC and 3D-ROESY-HSQC experiments.

The changes in hydration across the free proteins versus the complex indicates an

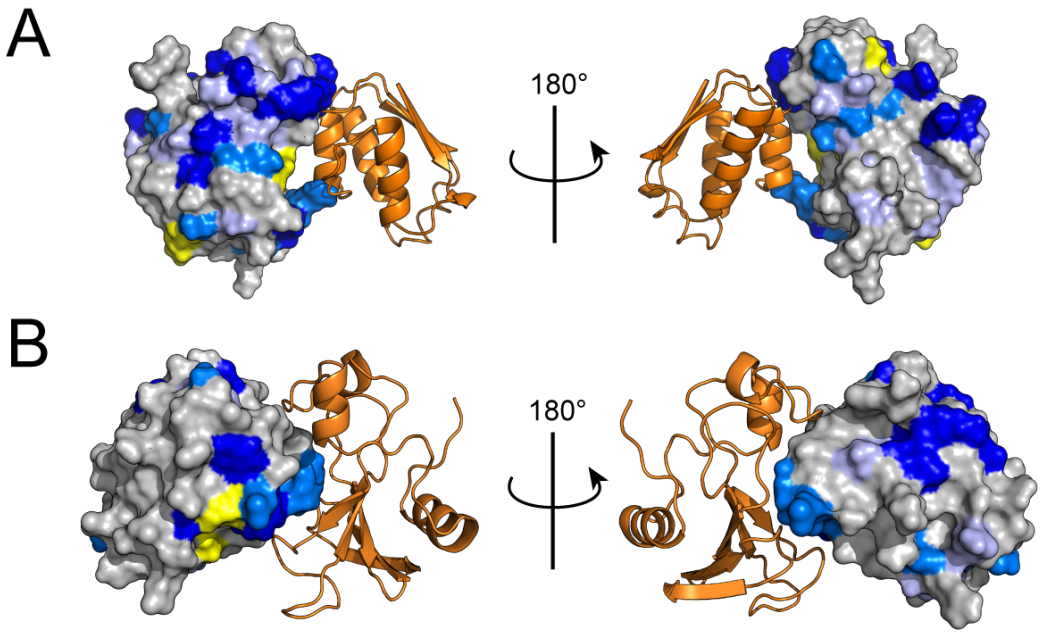


Figure 5.17: Hydration map of barnase-barstar complex. All amide hydrogens within NOE distance of solvent are indicated on the protein surface of (A) barnase in the barnase-barstar complex and (B) barstar in the barnase-barstar complex as follows: dark blue, bound hydration sites ($\frac{NOE}{ROE} \leq -0.35$); blue, slow hydration sites ($-0.35 \leq \frac{NOE}{ROE} \leq -0.2$); light blue, relatively slow hydration sites ($-0.2 \leq \frac{NOE}{ROE} \leq 0.0$). Colored in yellow correspond to sites where the NOE/ROE ratio was positive indicating hydrogen exchange at these sites. The dark blue/light blue hydration class corresponds to sites where water is within NOE distance (4.3Å) of the amide hydrogen and has a residence time on the order of the correlation time of the protein ($\sim 10ns$) or longer.

abundance of bound waters at the interface of binding. Encapsulating free barstar in CTAB/Hexanol mixture resulted in unfolded protein. Hence, the hydration dynamics measurement of free barstar was performed in LDAO/10MAG surfactant mixture where barstar was folded. However, the interpretation of the hydration measurements performed in LDAO/10MAG are complicated by the hydrogen exchange observed due to the hydroxyl head group of 10MAG. Nevertheless, we did not observe any bound waters

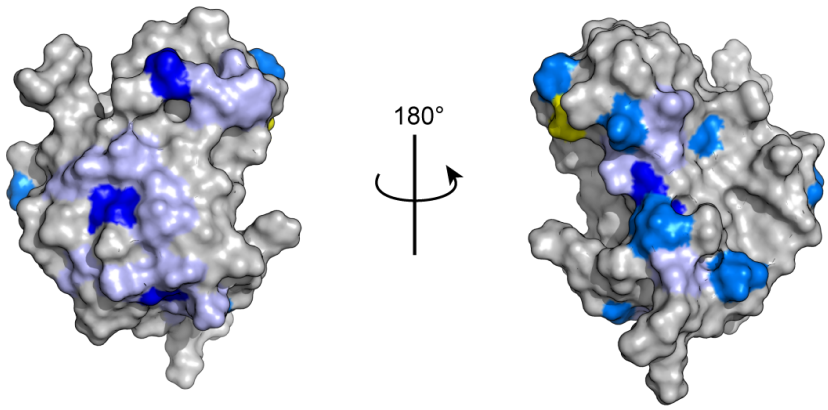


Figure 5.18: Hydration map of free barnase. All amide hydrogens within NOE distance of solvent are indicated on the protein surface of barnase as follows: dark blue, bound hydration sites ($\frac{NOE}{ROE} \leq -0.35$); blue, slow hydration sites ($-0.35 \leq \frac{NOE}{ROE} \leq -0.2$); light blue, relatively slow hydration sites ($-0.2 \leq \frac{NOE}{ROE} \leq 0.0$). Colored in yellow correspond to sites where the NOE/ROE ratio was positive. The dark blue/light blue hydration class corresponds to sites where water is within NOE distance (4.3Å) of the amide hydrogen and has a residence time on the order of the correlation time of the protein ($\sim 10ns$) or longer.

to free barstar.

The hydration dynamics across the rest of the protein surface in the complex (away from the interface) also exhibits slow waters. This suggests that there is a concerted change in the conformational dynamics of the proteins and the solvent dynamics across the protein. Consideration of the change in conformational entropy determined above and the estimated loss in rotational-translational entropy of the proteins ($-T\Delta S_{RT} \sim$ 9 kcal mol^{-1}) [145] (and independent calculations from MD simulations of barnase-barstar, barnase and barstar (not shown)) in light of the measured total binding entropy suggests that the solvent entropy contributes unfavourably by 13 kcal mol^{-1} or about 0.5 kT per interfacial water (agreeing well with the loss of one degree of freedom, translational

in the case hydrogen bound waters). The large favourable change in conformational entropy arising from the association of barnase and barstar raises the possibility that it could serve as a driving force for the association of the substrate analog dCGAC to barnase. This is not only important to understand whether conformational entropy can act as a driving force for binding but also to understand the relative contribution of conformational entropy in binding interactions of vastly differing affinities.

5.3.4 Conformational entropy can act as a driving force for binding interactions

The changes in motion of the polypeptide backbone of barnase upon binding dCGAC are on average small while the motions of the methyl bearing amino acid side chains of barnase increase upon formation of the complex ($< O^2_{axis} >= -0.040 \pm 0.010$; n=50) (Figure 5.19 and 5.20). The aromatic amino acid side chain on the other hand display decreased motions at the interface and slight increase in motions away from the interface, however, on average the changes in motions of the aromatic amino acid side chains are small ($< O^2_{aro} >= -0.002 \pm 0.010$; n=11) (Figure 5.21).

The increased changes in motion of the methyl groups supplemented with the negligible changes on average in both the backbone and aromatic side chains imply that the dominant contribution to conformational entropy changes arise from the changes in methyl dynamics. Using the 'entropy meter', this translates to a favourable increase

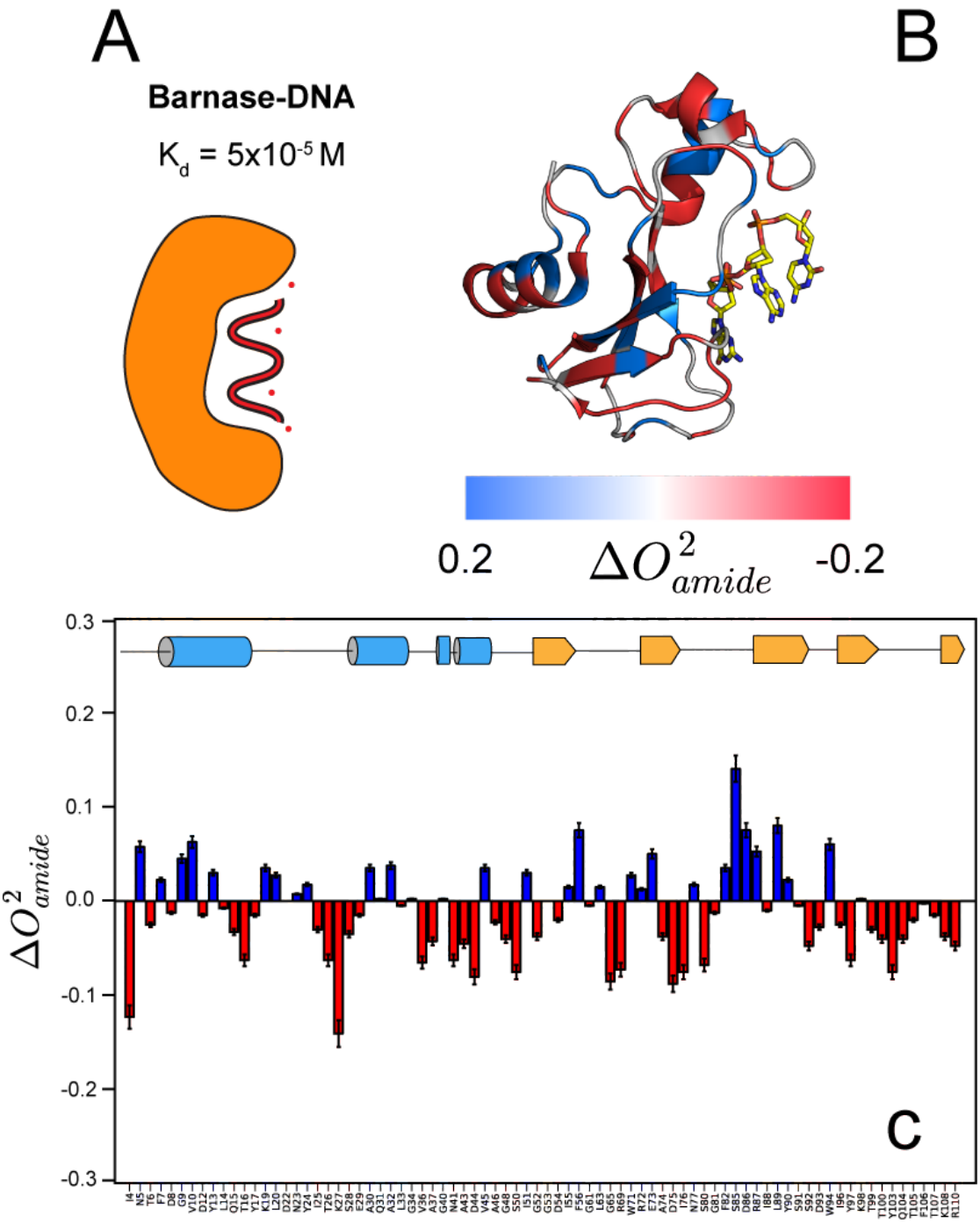


Figure 5.19: (A) Cartoon depiction of the barnase-dCGAC complex. (B)The spatial distribution of the changes in order parameter of the amide backbone (ΔO^2_{NH}) of barnase upon dCGAC binding. (C) Bar graph plot of the per residue change in O^2_{NH}.

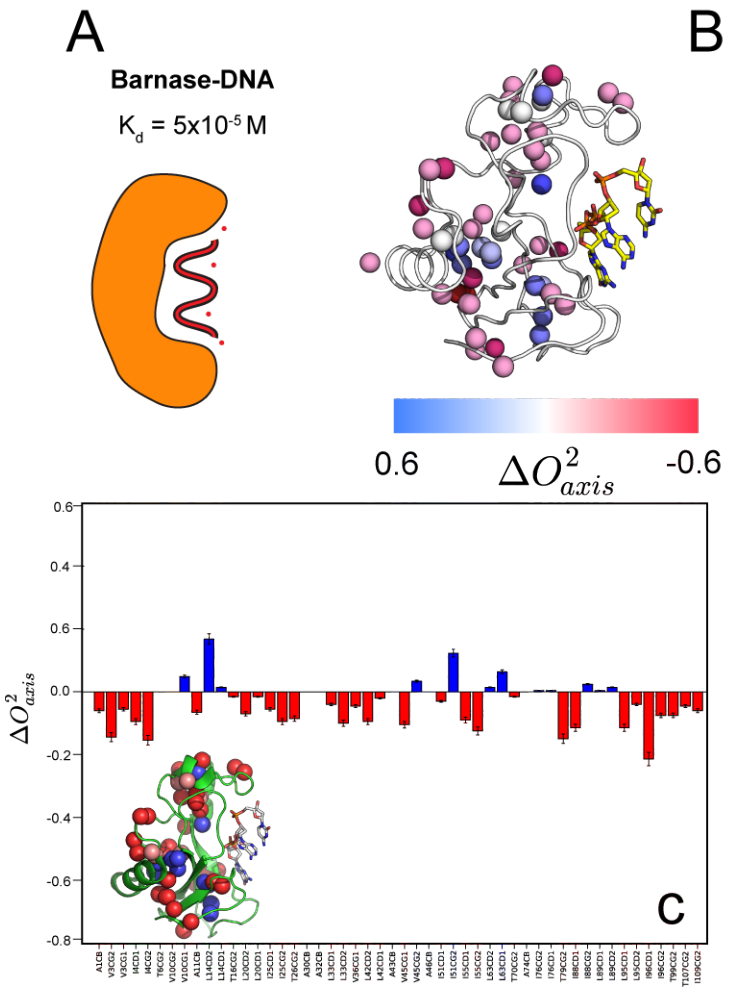


Figure 5.20: (A) Cartoon depiction of the barnase-dCGAC complex. (B) The spatial distribution of the change in O^2_{axis} indicates that on average barnase is more dynamic when bound to dCGAC. The hydrophobic core of barnase however becomes rigid upon binding similar to the barnase-barstar complex. (C) The bargraph plot of the residue by residue change in O^2_{axis} suggests that even though on average barnase becomes more dynamic, the degree of flexbility is comparatively lower than barnase in barnase-barstar complex. Inset shows the spatial distribution of residues which become rigid (blue) and dynamic (red).

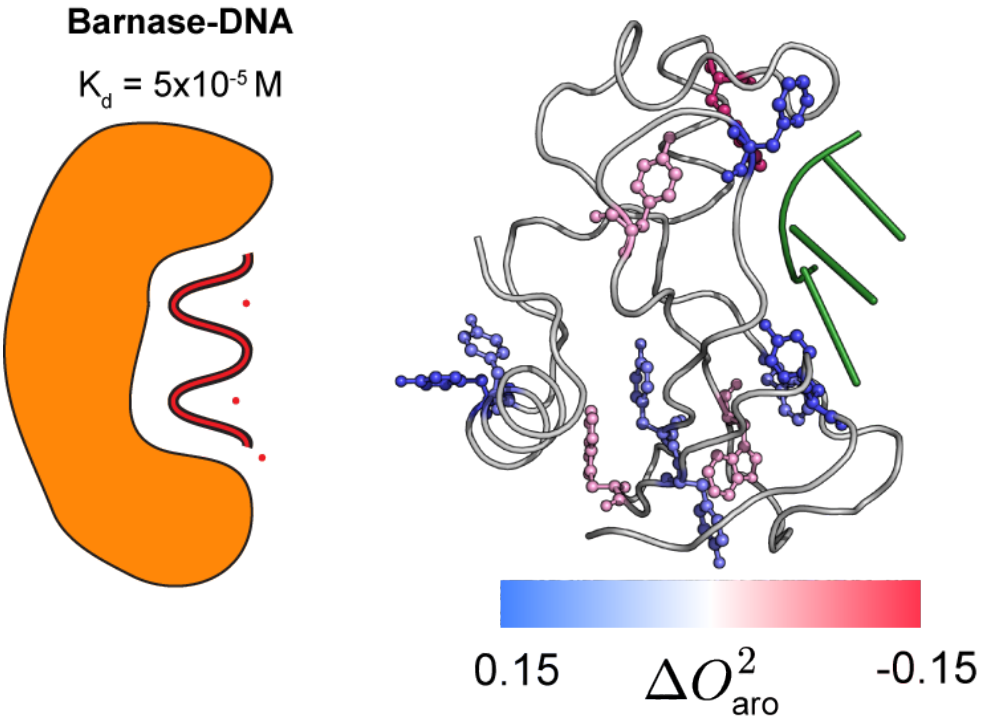


Figure 5.21: The spatial distribution of the change in O^2_{aro} indicates that aromatic amino acid residues close to the interface of dCGAC binding become less dynamic than those away from the interface. On average the motions of the aromatic amino acids do not change significantly ($< \Delta O^2_{aro} >$ = -0.002±0.010; n = 11). The 4mer DNA dCGAC is shown in green.

in the conformational entropy of barnase of 5.1 kcal mol^{-1}. The change in solvent entropy estimated from the change in accessible surface area [10] upon formation of the barnase-dCGAC complex also contributes favourably ($-T\Delta S_{solvent}$ = -9.5 kcal mol^{-1}) and is offset by the predicted loss of translational-rotational entropy [145]. Since the binding enthalpy is negligible (Figure 5.5), the change in conformational entropy essentially defines the Gibbs free energy of binding. To further corroborate the solvent entropy estimate from changes in accessible surface area, the hydration dynamics measurements were made for barnase-dCGAC complex. This analysis reveals that most sites

for which NOE and ROE cross peaks were observed in the free barnase were also seen in the barnase-dCGAC complex. Additionally, these sites displayed higher NOE/ROE ratios in the barnase-dCGAC complex indicating that the waters are less tightly bound in the dCGAC bound complex hinting at the potential favourable solvent entropy changes (Figure 5.22).

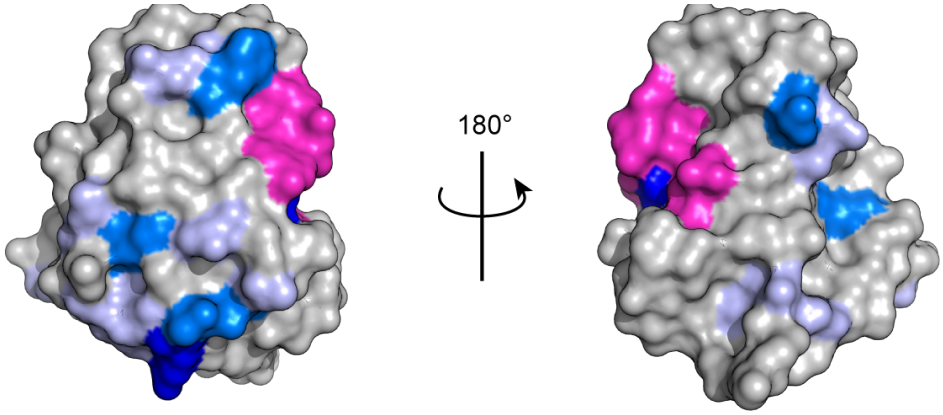


Figure 5.22: Hydration map of barnase bound to 4mer DNA dCGAC . All amide hydrogens within NOE distance of solvent are indicated on the protein surface of barnase, as follows: dark blue, bound hydration sites ($\frac{NOE}{ROE} \leq -0.35$); blue, slow hydration sites ($-0.35 \leq \frac{NOE}{ROE} \leq -0.2$); light blue, relatively slow hydration sites ($-0.2 \leq \frac{NOE}{ROE} \leq 0.0$). The dark blue/light blue hydration class corresponds to sites where water is within NOE distance (4.3Å) of the amide hydrogen and has a residence time on the order of the correlation time of the protein ($\sim 10ns$) or longer.

The results from barnase-barstar and barnase-dCGAC suggest that conformational entropy changes can not only drive binding interactions but also play a vital role in offsetting other entropic penalties incurred upon binding. This compensation phenomenon contributes to the barnase-barstar interaction attaining very high affinity of binding. It is thus important to understand if such a phenomenon is observed in other high affinity interactions especially those involving protein-small molecule ligand binding due to its

relevance to pharmaceutical drug design.

5.3.5 Conformational entropy offsets unfavourable entropic contributions to binding

In the case of barnase-barstar, conformational entropy almost completely offsets the unfavourable entropic contributions to Gibbs free energy of binding arising from solvent, protein rotational and translational entropy. In order to examine the potential role of conformational entropy in the binding of small ligands by proteins, histamine binding protein (HBP(D24R)) harboring a D24R mutation, which abolishes one of two histamine binding sites is studied here. As mentioned before, the binding of HBP(D24R) to histamine is driven by large favourable enthalpy and a net unfavourable total entropy of binding (Figure 5.3). Motions in the sub-nanosecond time scale of the backbone and side-chain methyl were probed to understand the role played by the protein conformational entropy in the binding of high affinity histamine. The overall changes in the motions of the backbone amide N-H bonds are minimal. In contrast, the motions of the methyl bearing side-chains are increased on average in the histamine bound complex ($< \Delta O^2_{axis} >= -0.025 \pm 0.005$; n=56) and corresponds to a favourable contribution from conformational entropy of -3.7 kcal mol^{-1} (Figure 5.23).

The distribution of the dynamical response of side chains to histamine binding indicates that the methyl groups in the beta-barrel core of the protein on average become more dynamic ($< \Delta O^2_{axis} >= -0.033 \pm 0.008$; n=14). Despite only a few methyl bearing side

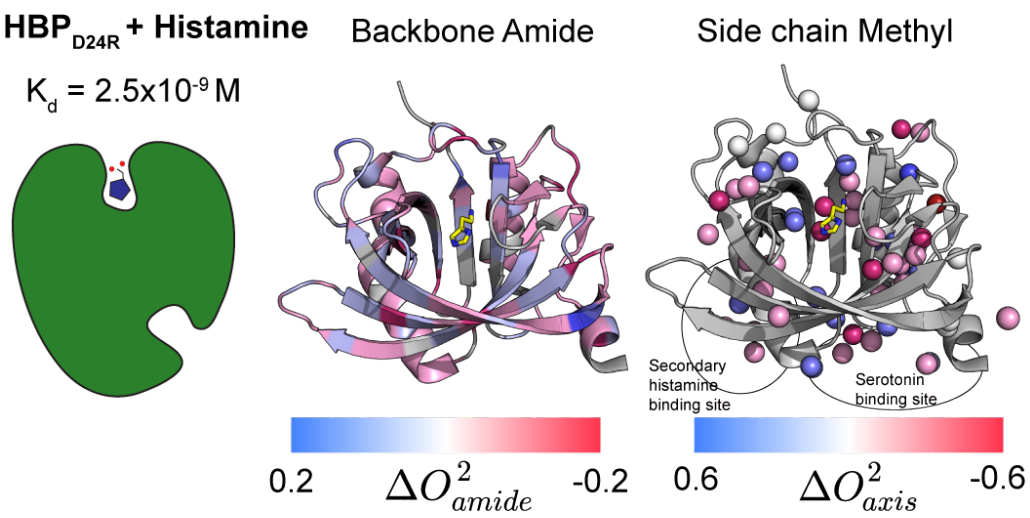


Figure 5.23: The spatial distribution of the changes in motion of backbone amide and methyl side chains of HBP(D24R) are indicated. In particular, the cluster of methyl bearing amino acid side chains within the beta-barrel core, the secondary histamine binding site as well as the remote serotonin binding site become more dynamic upon histamine binding.

chains ($\sim$ 4Å away) residing near the histamine binding site, a distinct propagation of dynamics is observed at the remote (11Å away) secondary histamine binding site ($< \Delta O^2_{axis} >= -0.030 \pm 0.005$; n=7). The favourable conformational entropy contribution offsets in part the unfavourable entropy contribution arising form the bound waters at the histamine binding site and the loss in rotational-translational entropy. Without a favourable conformational entropy contribution to HBP(D24R)-histamine interaction, the binding affinity of this interaction would be in the micro-molar regime compared to the very high affinity (nM) observed. Importantly, these results suggest that small molecule ligands are indeed capable of eliciting global dynamical changes in proteins and can thereby modulate the conformational entropy of proteins.

5.3.6 Conformational entropy modulates binding affinity

The increase in motions upon high affinity histamine binding (i.e. the site of D24R mutation) could potentially contribute a conformational entropy penalty upon second ligand binding resulting in the weaker affinity observed at the secondary binding site. In order to test this hypothesis, the entropy driven binding of a second small molecule ligand, serotonin, to HBP(D24R) is studied. Serotonin binds at a remote site (10Å away) from both the primary and secondary histamine binding site. ITC and NMR chemical shift titration experiments (Figure 5.7) established that the binding of serotonin is very weak and is entirely entropy driven. Serotonin binding to HBP(D24R) ($< \Delta O^2_{axis} >= 0.006 \pm 0.005$; n=56) and HBP(D24R)-histamine ($< \Delta O^2_{axis} >= 0.011 \pm 0.005$; n=56) results in an overall reduction in motion of the methyl bearing side chains. The increased dynamics at the serotonin binding site ($< \Delta O^2_{axis} >= -0.030 \pm 0.005$; n=7) observed upon the high affinity histamine binding is now quenched upon serotonin binding to HBP(D24R)-histamine complex ($< \Delta O^2_{axis} >= 0.014 \pm 0.005$; n=7). The overall changes in motions of the methyl bearing side chains when serotonin binds HBP(D24R) results in a conformational entropy change ($-T\Delta S_{conf}$) of $+1.0$ kcal mol^{-1} versus $+2.1$ kcal mol^{-1} ($-T\Delta S_{conf}$) when serotonin binds HBP(D24R)-histamine complex. The additional unfavourable conformational entropy penalty (HBP(D24R)-histamine-serotonin versus HBP(D24R)-serotonin) results in the much lower affinity observed for serotonin binding to HBP(D24R)-histamine ($K_d \sim 5mM$) versus HBP(D24R) ($K_d \sim 1.8mM$).

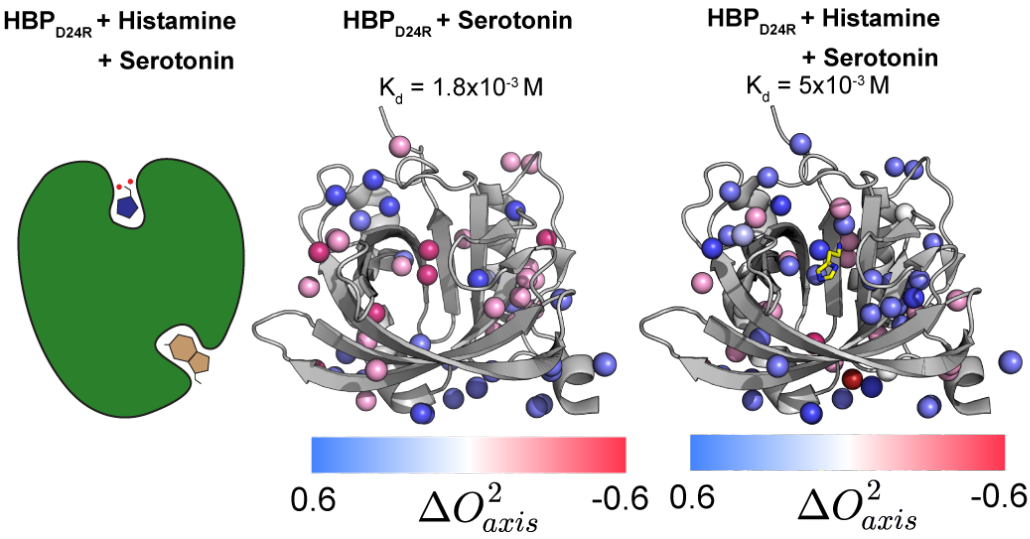


Figure 5.24: The spatial distribution of the changes in motion of the methyl side chains of HBP(D24R) and HBP(D24R)-histamine complex upon serotonin binding are indicated. The increased motions of the methyl probes in the serotonin binding site (Figure 5.23) upon histamine binding is now quenched upon serotonin binding resulting in additional conformational entropy penalty compared to the serotonin binding apo-HBP(D24R).

5.4 Conclusions and Discussion

The view of high affinity interactions has been so far largely restricted to enthalpy driven processes whereby favourable interactions are the major determinants of the free energy of binding [7, 123]. This view has been supplemented with the favourable contributions from solvent entropy upon ligand binding due to the release of waters from the interface [127, 128]. This classical view of protein-ligand binding interactions has been the backbone of pharmaceutical drug design wherein small molecule ligands which result in favourable solvent entropy change are selected for further refinement to gain

favourable enthalpy contributions [123]. This process has largely restricted pharmaceutical drug design to hydrophobic ligands and has completely neglected the role played by protein conformational entropy. Recent measurements of the contributions of protein conformational entropy on moderately high affinity complexes ($K_d = 10^{-6} - 10^{-8}$ M) revealed that conformational entropy can indeed modulate the free energy of ligand binding [14, 27]. In order to understand the role played by protein conformational entropy in high affinity interactions, if any, the very high affinity barnase-barstar (protein-protein) interaction and the high affinity histamine-binding protein (HBP(D24R)) bound to small molecule histamine are examined.

While both these high affinity interactions consist of a large favourable enthalpy, the contributions of the total binding entropy differ considerably [126, 133]. The interaction of barnase-barstar is comprised of almost negligible entropy change whereas HBP(D24R)-histamine binding results in unfavourable total binding entropy [126, 133]. Using NMR relaxation experiments to measure the motions of the side chain methyl groups and their use as a dynamic proxy for conformational entropy, both barnase-barstar and HBP(D24R)-histamine show favourable conformation entropy changes upon binding. In the case of barnase-barstar, the conformational entropy change ($-T\Delta S_{conf}$ = -18.5 kcal mol^{-1}) is more than the enthalpy change accompanying the binding interaction. This large favourable conformational entropy offsets the unfavourable solvent entropy and loss in rotational-translational entropy accompanying this interaction. The

interface of barnase-barstar is "wet" i.e. it consists of trapped waters. The encapsulation of barnase, barstar and barnase-barstar complex in reverse micelles revealed that the solvent dynamics is in general slowed down in the complex and that the interface of the complex consists of many bound waters. There is a complete compensation of the loss in solvent entropy and rotational-translational entropy by the corresponding increase in conformational entropy of barnase and barstar upon binding (Figure 5.25). This remarkable compensation mechanism suggests that protein conformational entropy plays a vital role in achieving the femtomolar affinity barnase-barstar interaction and suggests that barstar has not only evolved structurally to bind specifically to barnase but has also evolved to modulate the conformational dynamics to achieve this very high affinity and high specificity binding.

Can small molecule ligands modulate the conformational entropy of proteins to result in favourable contributions to the free energy of binding? This question is particularly relevant with regards to the design of pharmaceutical drugs. In the case of the high affinity HBP(D24R)-histamine interaction, the binding of histamine results in a favourable change in conformational entropy of HBP(D24R) ($-T\Delta S_{conf}$ = -3.7 kcal mol^{-1}). This favourable change in conformational entropy is offset by the loss in rotational-translational entropy and the unfavourable solvent entropy of binding. However, since the gain in conformational entropy does not completely offset the unfavourable entropic contributions to binding, the net total binding entropy is unfavourable (Figure 5.25). The important feature to note here is that the same compensation mechanism observed in

the very high affinity barnase-barstar interaction is also observed here for the small molecule histamine binding to HBP(D24R) and suggests that small molecule ligands can indeed modulate the global conformational dynamics of proteins. The favourable conformational entropy contributions help maintain the high affinity histamine binding.

In order to further demonstrate that protein conformational entropy changes can prove vital for ligand binding, the binding of 4mer DNA dCGAC to barnase is examined. Both these interactions are completely entropy driven. In the case of barnase-dCGAC interaction the favourable solvent entropy change upon dCGAC binding is further supplemented by the favourable change in conformational entropy ($-T\Delta S_{conf}$ = -5.1 kcal mol^{-1}). The favourable change in solvent entropy is completely offset by the unfavourable change in rotational-translational entropy resulting in protein conformational entropy change as the sole driving force for barnase-dCGAC interaction. This suggests that conformational entropy can function as a driving force for ligand binding. Furthermore in order to provide direct evidence that protein conformational entropy changes can modulate the binding affinity, the binding of serotonin to HBP(D24R) and HBP(D24R)-histamine is examined. Serotonin binds to HBP(D24R) is completely entropy driven with very low affinity ($K_d \sim$ 1.8 mM) whereas serotonin binds to HBP(D24R)-histamine complex at even lower affinity ($K_d \sim$ 5 mM). Histamine binding to HBP(D24R) results in global changes in motion of the side chain methyl groups, in particular, the motions of the methyl bearing side chains near the serotonin binding site are enhanced. Serotonin binding to HBP(D24R) subsequently results in quenching of these enhanced motions

which together with an overall reduction in side chain motions results in unfavourable conformational entropy. This additional unfavourable conformational entropy change incurred when serotonin binds to HBP(D24R)-histamine complex versus HBP(D24R) results in the lower affinity observed. This activation of dynamics at a remote site could serve as a plausible mechanism for dynamics driven allostery as envisioned by Cooper & Dryden [32] and is currently being further investigated in the Wand lab.

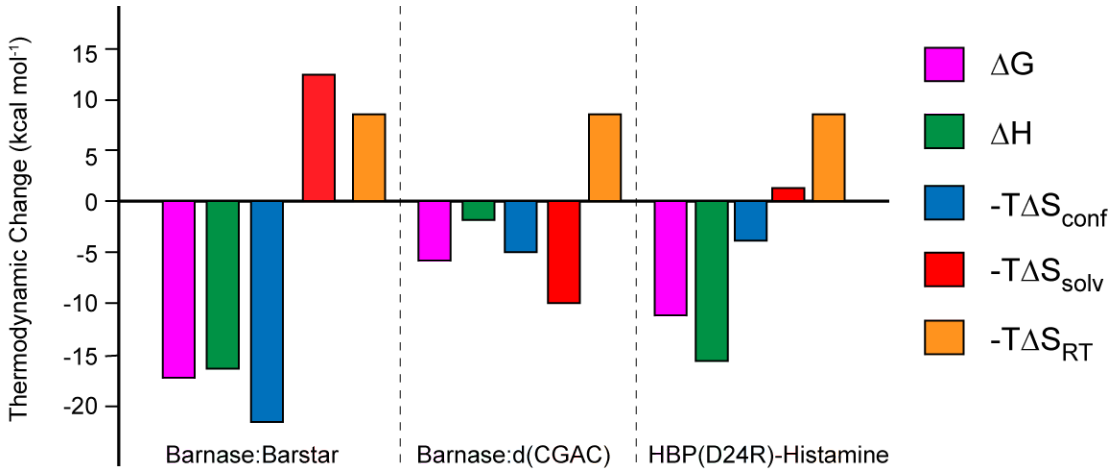


Figure 5.25: Shown above is the breakdown of the different contributions to binding for barnase-barstar, barnase-dCGAC and HBP(D24R)-histamine complexes.

The observation that protein conformational entropy can contribute favourably to the Gibbs free energy of binding and that this can be brought upon by small molecules suggests a reevaluation of the current pharmaceutical drug design strategies. The results presented here strongly suggest that protein conformational entropy is the missing piece of the puzzle in the quest for the rational design of high-affinity and high-specificity pharmaceutical drugs. Conformational entropy changes provide an additional dimension to the drug design process where the interplay between enthalpy, solvent entropy

and conformational entropy can now be explored. In particular, it is clear from the NMR measurements of side chain motions and hydration dynamics that ligand binding results in changes to both the conformational dynamics and the solvent dynamics concurrently. The use of the entropy meter to quantitate the conformational entropy contributions has so far utilized estimates of solvent entropy change obtained via the parametric relationship between the change in accessible surface area upon ligand binding and the solvent entropy change [10]. This parametric relationship however assumes that the change in protein conformational entropy change is negligible upon ligand binding. This discrepancy is addressed in the next chapter where I have also expanded the scope of the entropy meter formulation by including over 35 different protein-ligand complexes with vastly varying binding affinities and ligand types such that conformational entropy changes for any protein-ligand interaction can now be measured accurately.

Even though conformational entropy changes here have been shown to be vital in achieving high affinity interactions and that ligands can modulate the conformational entropy change of proteins, a mechanistic understanding on how ligand binding results in dynamics changes at remote sites, far away from the interface, remains to be studied in more detail.

Chapter 6

A Universal Entropy Meter

Sections of the chapter pertaining to the calmodulin mutants bound to peptide were contributed by Kyle Harpole, graduate student in the Wand Lab.

6.1 Introduction

Specific molecular interactions are a hallmark of complex cellular processes and often involve proteins, small molecules and nucleic acids. Protein-ligand interactions often involve dozens of amino acids over thousands of Å^2 of contact area [146]. The non-uniform contributions of individual amino acid residues to the free energy of binding has hindered the computation of the energetics governing protein-ligand interactions from static structures alone [9]. The characterization of the thermodynamic landscape

governing these interactions is hence vital for understanding their function. However, dissecting this thermodynamic landscape into enthalpy and the different components of entropy, has proven difficult as structure based analysis of such interactions has masked the important role played by entropy. Of particular interest here is the interplay between enthalpy and entropy, in particular the role played by protein conformational entropy in modulating the free energy of the association of protein with a ligand. The entropy of binding obtained by calorimetric measurements is comprised of contributions from protein, ligand and solvent:

$$\Delta G_{total} = \Delta H_{total} - T\left[\Delta S_{solvent} + \Delta S_{protein} + \Delta S_{ligand}\right] \tag{6.1}$$

The contributions of solvent entropy have historically been framed in terms of the hydrophobic effect, wherein a parametric relationship between the change in accessible surface area of the protein upon ligand binding and changes in solvent entropy has been developed [10, 147]. The entropic contributions of the protein and ligand, classified into changes in conformational entropy (ΔS_{conf}) and rotational-translational entropy (ΔS_{RT}) have been less understood [11, 32, 148]. While it has been recognized that the residual conformational entropy of proteins is significant, its measurement has proven particularly challenging [11]. NMR spectroscopy has been used to connect the measures of fast sub-nanosecond time scale motions to conformational entropy through the use of a model dependent interpretation [15–17]. The application of such a model

dependent interpretation to the binding of calcium-saturated calmodulin (CaM) to different target peptides revealed a remarkable linear relationship between the total binding entropy and the conformational entropy derived from the model using NMR measured changes in motions [13]. However, the use of a model dependent approach raised several concerns regarding the effect of correlated motion, the use of a more complex potential energy function and the completeness of the oscillator count [17]. The technical issues concerning the use of specific models to convert measures of fast internal motions of methyl bearing amino acids to conformational entropy were subsequently addressed using an unbiased model-independent approach [14]. The model-independent approach relied on the use of the motions of methyl bearing amino acids as a 'dynamics proxy' for providing quantitative estimates of protein conformational entropy. The use of this dynamical proxy revealed a linear relationship between the total binding entropy and calculated solvent entropy to the conformational entropy measured by NMR, resulting in an 'Entropy Meter' [14, 24]. This empirical calibration approach has since been applied to the binding of catabolite-activator protein (CAP) to DNA resulting in a similar linear relationship between the NMR measured protein conformational entropy and the total binding entropy and entropy of solvent [27]. These studies revealed that protein conformational entropy is critical for regulating the activity of these proteins and that protein conformational entropy tunes the binding affinity of protein-ligand interactions [14, 27]. Even though these initial studies signify a great leap forward in experimental measurement of protein conformational entropy, they exhibit an apparent difference in the

empirical calibration of their respective entropy meters. This raises several questions regarding both the general validity of the dynamical proxy method and the generality of the entropy meter. In chapter 2, I have used molecular dynamics simulations to report that the population distribution between different rotameric states of the amino acid side chains in proteins was the dominant contributor to the protein conformational entropy and that scaling of the NMR measured motions of the methyl bearing side chains by the number of side chain torsion angles helped resolve this apparent discrepancy. This unified entropy meter resulted in an identical empirical calibration for both CaM bound to target peptides and the CAP-DNA complexes [107].

Although the entropy meter formulation for CaM and CAP complexes agreed well (Chapter 2), one of the key questions which remains unaddressed is the generality and the broad applicability of the entropy meter to different classes of protein-ligand interactions. Here, I have expanded the entropy meter from the initial set of 17 protein-ligand complexes (comprised of 6 CaM-peptide and 11 CAP-DNA complexes) to 35 protein-ligand complexes spanning a broad range of binding affinities (K_d = $10^{-6} - 10^{-14}$ M) and containing different classes of ligands (Figure 6.1) [14, 27, 49, 149–154].

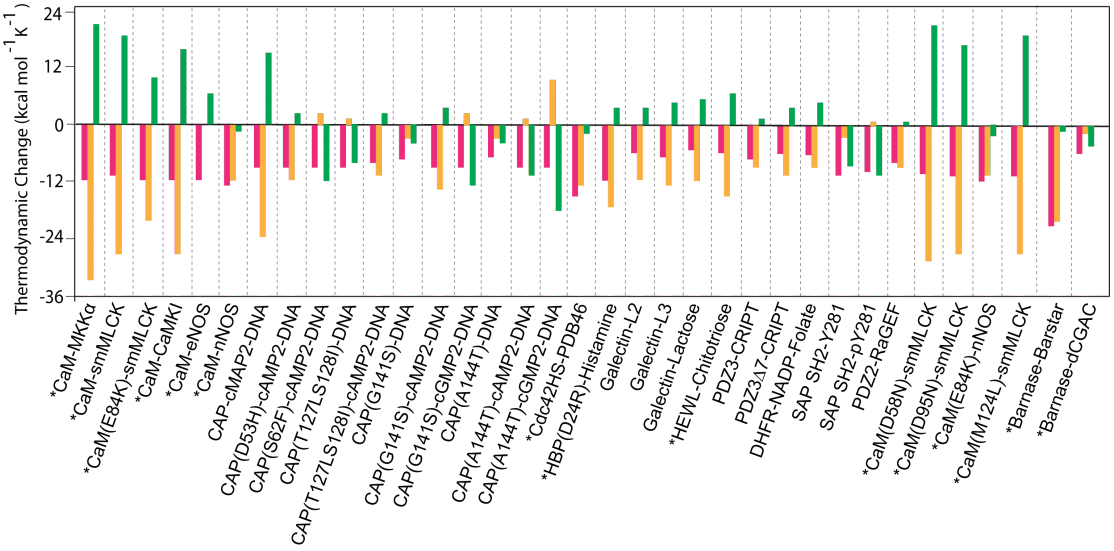


Figure 6.1: The broad spectrum of thermodynamic signatures: (red) change in Gibbs free energy of binding, (yellow) change in binding enthalpy and (green) change in entropy of binding, of the different protein-ligand complexes used here represents an excellent repository for understanding the role played by protein conformational entropy in ligand binding. Asterix ($*$) represents protein-ligand complexes for which both the isothermal calorimetry and NMR-relaxation analysis was performed in the Wand lab. The calorimetric and NMR measured motions of methyl side chains for the rest of the protein-ligand complexes were catalogued from previously published studies (see Appendix).

6.2 Results

The entropy meter constructed from the 35 different protein-ligand complexes (Figure 6.1) using their respective calorimetric measurements (total binding entropy), NMR-measured motions of methyl bearing side chains and the solvent entropy calculated using changes in accessible surface area displays a linear relationship with identical calibration (slope = -0.002±0.0003 kcal mol^{-1} K^{-1}) as reported in Chapter 2. The entropy meter formulation is valid across a spectrum of protein-ligand complexes suggesting

that protein conformational entropy of any protein-ligand interaction can be quantified accurately.

6.2.1 Expanding the 'Entropy Meter'

While calorimetric measurements can aid in the measurement of the total binding entropy, they cannot help resolve the different entropic contributions to binding (i.e. ΔS_{conf}, $\Delta S_{solvent}$ and ΔS_{RT}). The entropic contribution from solvent has been estimated using a parametric relationship proposed by Hilser and Freire based on the changes in polar and apolar accessible surface area upon ligand binding [10]. Taken together, the configurational entropy of a protein-ligand interaction can be estimated using the expression:

$$\Delta S_{conf}^{protein+ligand} + \Delta S_{RT}^{protein+ligand} = \Delta S_{total} - \Delta S_{solvent} \tag{6.2}$$

NMR relaxation measurements of the internal motions in proteins have been able to provide direct access to the conformational entropy of proteins [17, 24]. By combining

- Calorimetric measurements (ΔS_{total})
- Structural information ($\Delta S_{solvent}$)
- NMR measured conformational dynamics

the conformational entropy changes accompanying protein-ligand can be accurately quantified. Motions in sub-nanosecond time scale for the backbone amide and methyl bearing side chains are probed using standing ^{15}N and ^{2}H or ^{13}C NMR relaxation experiments [17], respectively and interpreted in terms of the so-called Lipari-Szabo model free squared generalized order parameter (O^2) [18]. O^2 is a measure of the degree of spatial restriction of a given vector and varies between zero, which represents complete isotropic disorder and one, which corresponds to no internal motion within the molecular frame. Previous studies have shown that the dynamics of protein backbone are generally unperturbed upon protein-ligand binding [17, 24]. Due to the abundance of methyl groups in proteins, the motions of the methyl bearing amino acids are of principal interest. The 'entropy meter' approach utilizes an empirical calibration between changes in O^2 and the change in measurable total binding entropy and calculable changes in solvent entropy [14]. The entropy meter is thus formulated as:

$$\Delta S_{conf} = m.[N_{\chi}^{protein}. < \Delta O_{axis}^{2} >^{protein} + N_{\chi}^{ligand}. < \Delta O_{axis}^{2} >^{ligand}] + \Delta S_{other} \qquad (6.3)$$

Where the slope ('m') allows for the empirical calibration of the conversion of the changes in side chain dynamics (ΔO_{axis}^{2}) to conformational entropy and N_{χ} represents the total number of side chain torsion angles.

The thermodynamics of binding of four previously studied calmodulin mutants in complex with peptide targets were characterized by ITC and their respective motional changes

in methyl bearing side chains measured by NMR relaxation experiments [106]. Additionally, we have recently reported both the thermodynamics of binding and NMR measurements of motions of methyl bearing side chains for twelve different protein-ligand complexes [14, 49, 154] (also refer to Chapter 5). This together with the catalogued calorimetric information and NMR measured methyl side chain motions for nineteen different protein-ligand complexes constitute the entire set of 34 different protein-ligand complexes studied here [14, 27, 49, 149–154]. The changes in polar and apolar accessible surface area for each of the respective protein-ligand interaction has been tabulated to aid in the measurement of the solvent entropy changes (see Appendix). The total binding entropy and the entropy of solvent for the different protein-ligand complexes correlate excellently with their corresponding NMR "dynamical proxy" based conformational entropy measurement (Figure 6.2), expanding the purview of the "Entropy Meter".

The 'Entropy Meter' formulation above is the first exhaustive calibration for protein conformational entropy involving several different protein-ligand complexes each with vastly differing thermodynamic signatures and spanning a wide range of binding affinities. Moreover, the set of protein-ligand complexes reported here spain a wide variety and size of ligands: small molecules, peptides, proteins and nucleic acids. This establishes that the conformational entropy change for any class of protein-ligand complex can be quantified accurately using the 'Entropy Meter'.

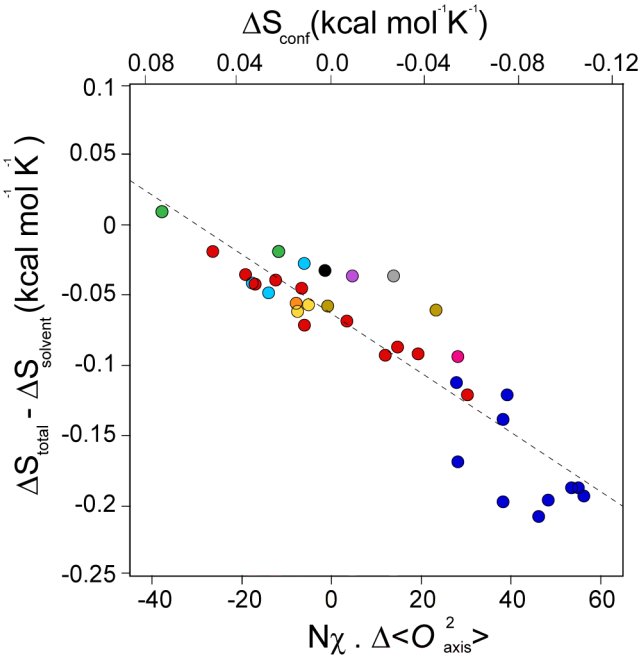


Figure 6.2: A simple linear relationship exists between the total binding entropy and the entropy of solvent to the conformational entropy by NMR relaxation parameters derived from methyl-bearing amino acids (Equation 6.3). The empirical calibration of the entropy meter thus obtained resulted in a slope (m = -0.002±0.0003 kcal mol^{-1} K^{-1}) and a ΔS_{RT} = -0.26±0.12 kcal mol^{-1} K^{-1}. The slope obtained here is identical to that reported previously (refer Chapter 2) and suggests the generality of the dynamical proxy method for quantitative estimates of protein conformational entropy. A linear regression statistic of R = 0.92 (P < 0.01) was obtained suggesting a strong linear correlation. The different protein:ligand complexes are represented by: (dark blue) CaM-peptide, (red) CAP-DNA, (light blue) Galectin-L2/L3/Lactose, (green) Barnase-barstar/Barnase-dCGAC, (yellow) SAPSH2-Y281/pY281, (purple) HEWL-Chitotriose, (pink) Cdc42HS-PBD46, (brown) PDZ3/PDZ3Δ7-CRIPT, (grey) DHFR-NADP-Folate, (black) PDZ2-RaGEF, (orange) HBP(D24R)-histamine.

6.2.2 Recalibration of solvent entropy for a 'Universal Entropy Meter'

The formulation of the entropy meter so far has utilized the parametric relationship between the change in polar and apolar accessible surface area of proteins upon ligand binding and the changes in entropy of the solvent [10]. However, as mentioned before

this approach of estimating the solvent entropy changes does not take into account the changes in residual conformational entropy of proteins which has been shown to be significant. Several studies have demonstrated that ligand binding results in dynamical changes in proteins far removed from the binding site [27, 49, 150]. The measurement of hydration dynamics across the surface of barnase, barstar and barnase-barstar complex (Chapter 5) using reverse micelle encapsulation indicated that the distribution of dynamics of water is highly heterogeneous and that ligand binding results in changes in hydration across the protein surface (refer to Chapter 5). The binding of ligand therefore results in changes to both the hydration dynamics as well as changed in protein conformational dynamics. This warrants that the contribution of solvent entropy changes and protein conformational entropy changes to total binding entropy should be accounted for together and not independently. The entropy meter formulation in Figure 6.2 results in an estimate of ΔS_{RT} of -0.26±0.12 kcal mol^{-1} K^{-1}, this estimated value is twice that of independently estimated values for the loss in rotational-translational entropy [145]. This discrepancy could be due to the parametrization of solvent entropy and conformational entropy independently.

In order to address this issue, the solvent entropy and conformational entropy are calibrated simultaneously using:

$$\Delta S_{total} = m.[N_{\chi}^{protein}. < \Delta O_{axis}^{2} >^{protein} + N_{\chi}^{ligand}. < \Delta O_{axis}^{2} >^{ligand}] + [a1.\Delta ASA_{polar} + a2.\Delta ASA_{apolar}] + \Delta S_{RT} \quad (6.4)$$

where slope ('m') is the calibration of the conformational entropy change using the dynamical proxy method, a1 and a2 are the parametrization constants for the solvent entropy changes and ΔS_{RT} is the loss in rotational-translational entropy. All four parameters (m, a1, a2 and ΔS_{RT}) are fit simultaneously using the calorimetric data for the 34 different protein-ligand complexes used here (Barnase-Barstar complex excluded due to its wet interface of interaction). The new empirical calibration of the entropy meter and the improved calibration of the changes in solvent entropy resulted in a slightly better correlation of the total binding entropy to the NMR measured conformational entropy (Figure 6.3).

The solvent entropy changes estimated using equation 6.4 does not neglect the change in protein conformational entropy and additionally, the estimated change in rotational-translational entropy from the new entropy meter (Figure 6.3) is in excellent agreement with the predicted loss in rotational-translational entropy. This 'Universal Entropy Meter' not only provides quantitative estimates of protein conformational entropy but also allows for an in depth understanding of the interplay between different entropic components of the Gibbs-Helmholtz free energy of binding. It is remarkable that the entropy

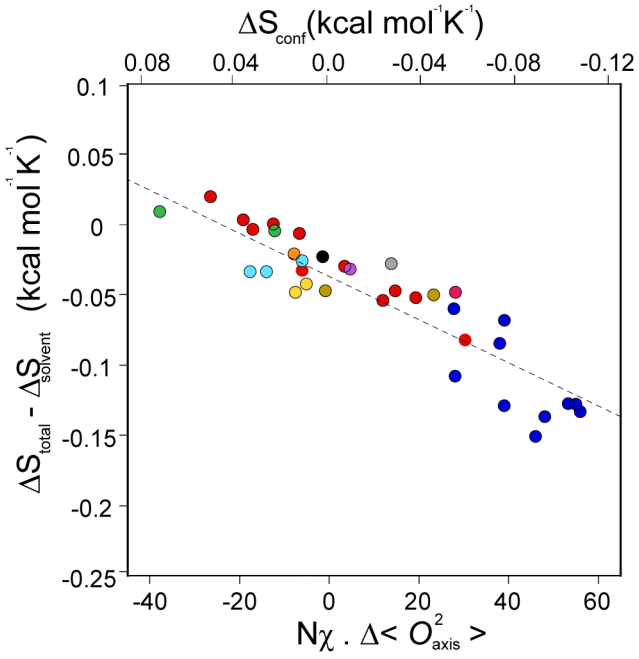


Figure 6.3: The above 'Entropy Meter' simultaneously calibrates both the conformational entropy and the solvent entropy from changes in NMR measured dynamics of methyl-bearing amino acids and changes in accessible surface area of protein and ligand upon binding, respectively (Equation 6.4). The empirical calibration thus obtained results in a slope (m = -0.0016±0.0002 kcal mol^{-1} K^{-1}) and a ΔS_{RT} = -0.14±0.07 kcal mol^{-1} K^{-1}. The ΔS_{RT} value obtained here agrees excellently with the predicted loss in rotational-translational entropy. The linear regression statistic obtained here (R=0.95, P < 0.01) is a slight improvement over the calibration obtained in Figure 6.2. The different protein:ligand complexes are represented by: (dark blue) CaM-peptide, (red) CAP-DNA, (light blue) Galectin-L2/L3/Lactose, (green) Barnase-barstar/Barnase-dCGAC, (yellow) SAPSH2-Y281/pY281, (purple) HEWL-Chitotriose, (pink) Cdc42HS-PBD46, (brown) PDZ3/PDZ3Δ7-CRIPT, (grey) DHFR-NADP-Folate, (black) PDZ2-RaGEF, (orange) HBP(D24R)-histamine.

meter formulation suggests an entropic compensation mechanism between changes in solvent entropy and changes in conformational entropy for high-affinity interactions. In the case of the nano-molar affinity Ca^{2+}-CaM-peptide interaction, favourable changes in solvent entropy compensate for the unfavourable changes in conformational entropy.

Intriguingly, in the case of the very high affinity barnase-barstar (femtomolar affinity) interaction, where the current model of estimating solvent entropy from change in accessible surface area is not applicable, unfavourable changes in solvent entropy (Refer to Chapter 5) are almost completely compensated by the remarkable favourable changes in protein conformational entropy. The entropy meter formulation suggests that protein conformational entropy which has so far been poorly understood and hence largely neglected, is an additional tunable parameter for modulating ligand binding.

6.3 Conclusions

Structure based analysis of molecular interactions have helped understand in great detail the role played by enthalpy [7]. This structural analysis has also contributed to the understanding of the changes in entropy of solvent arising from the changes at the interface of ligand binding [127, 128]. However, the role played by entropy, particularly conformational entropy, has largely been obscured by this view. Protein conformational entropy has been recognized to play a vital role in modulating protein allostery and function [32]. Yet there has been no comprehensive experimental measurement of protein conformational entropy for protein-ligand interactions. Recent use of NMR relaxation methods has paved the way for quantitative estimates of protein conformational entropy through the use of motions of methyl bearing amino acid side chains as a dynamical proxy [14, 24, 27, 107]. While these studies laid the groundwork for the use

of an 'Entropy Meter' to not only understand the role played by protein conformational entropy but also to provide a quantitative estimate for the same, it has remained unclear whether the entropy meter constitutes a general method to measure the conformational entropy of any protein-ligand interaction. Here, I have shown that the 'Entropy Meter' formulation consisting of 35 different protein-ligand complexes each with a unique thermodynamic signature, different class of proteins (size, structural characteristics and flexibility) and different types of ligands (small molecules, peptides, proteins and nucleic acids) is indeed general. Furthermore, the improved empirical calibration for the solvent entropy changes presented here take into account the changes in protein conformational entropy. The resulting 'Universal Entropy Meter' is largely independent of any assumptions (the one major assumption being that the distribution of methyl groups in uniform so as to provide an excellent overall measure) and shows excellent agreement between the loss in rotational-translational entropy obtained here with previous independent estimates [145]. This is the first study where the total binding entropy has been successfully dissected into its different entropic components, providing an accurate quantitative estimate of the changes in protein conformational entropy and solvent entropy for different protein-ligand interactions (6.4). The entropy meter presented here could be utilized to measure the changes in protein conformational entropy of any protein-ligand interaction. The universal entropy meter provides a solid foundation to understand and dissect the interplay between the solvent entropy and protein conformational entropy thereby significantly increasing the energy landscape for the

rational design of pharmaceutical drugs [124, 155].

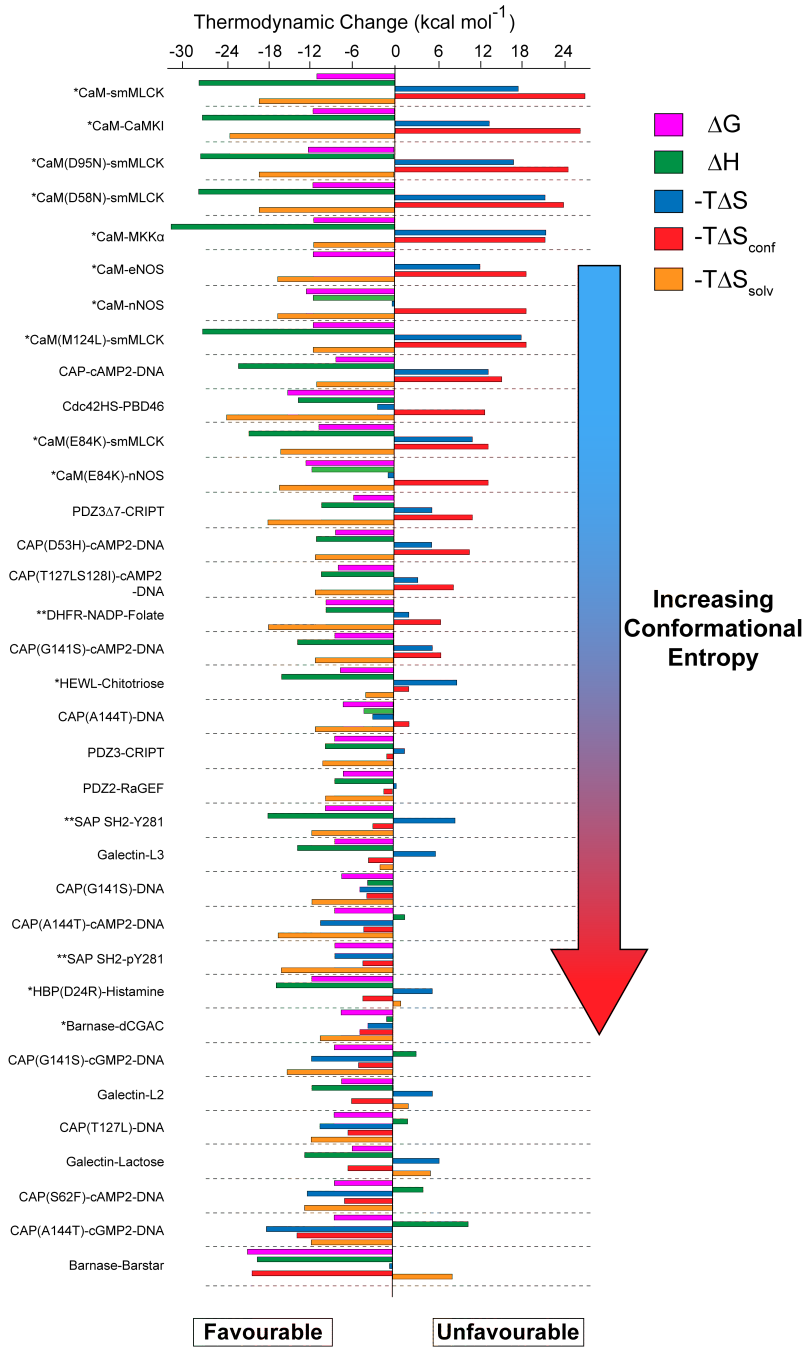


Figure 6.4: The different energetic contributions to the free energy of binding for the protein-ligand complexes studied. Shown above are: (Magenta) ΔG, (Green) ΔH, (Blue) ΔS_{total}, (Red) $\Delta S_{conformational}$, (Orange) $\Delta S_{solvent}$. The change in conformational entropy and solvent entropy were calculated based on the 'Universal Entropy Meter' formulation (Equation 6.4).

Chapter 7

Conclusions

7.1 Summary

The principal goal of the work presented in this thesis is to provide a detailed insight into the role played by conformational entropy in very high-affinity interactions. Prior to addressing this problem, several technical issues concerning the measurement of protein conformational entropy by NMR methods needed to be addressed. In particular, it has been well established that motions in the fast sub-nanosecond time scale within proteins contribute significantly to the conformational entropy [17]. The recent body of work from the Wand lab has laid the foundation to accurately quantify protein conformational entropy by using methyl bearing amino acid side chain motions as a dynamical proxy, resulting in an 'Entropy Meter' [13, 14]. The application of this dynamical proxy to two

different protein-ligand systems revealed a remarkable linear relationship between the total binding entropy and the NMR measured protein conformational entropy (dynamical proxy) [14, 27]. However, the apparent discrepancy between the empirical calibration of the entropy meter between the two studies raised several questions:

- Can methyl side chain motions be used for measuring protein conformational entropy?
- What are the origins of the conformational entropy measured using methyl side chain motions?
- What are the effects of correlated motions within proteins on the measurement of protein conformational entropy?
- Is the 'Entropy Meter' universal and hence applicable to different protein-ligand complexes?

In Chapter 2 [107] and 7, I have addressed all of the above questions and indeed shown that the 'Entropy Meter' is universal and can be utilized to provide accurate quantitative estimates of protein conformational entropy changes. Additionally, it also helps understand the interplay between the different entropic components to ligand binding. Extending the dynamical proxy method to very high-affinity complexes to understand the role played by conformational entropy would be insufficient without the characterization of the dynamics changes occuring at the interface of binding. The interface of high

affinity interactions generally contain limited numbers of methyl containing amino acids but are instead populated by aromatic amino acids. In order to quantify the dynamical changes of aromatic amino acids, I have introduced a new ^{13}C isotopic labelling scheme in Chapter 3 [26]. Using this new isotopic labelling scheme, I have described the fundamentals governing the fast internal motions of aromatic amino acids within proteins and shown that the interior of proteins are more liquid-like than previously understood (Chapter 4). In Chapter 4 I have also described the origins of the conformational entropy contributions from aromatic amino acid side chain motions. Unlike methyl side chains, whose dominant contribution to conformational entropy arises from population distribution between different rotamers [107], aromatic amino acid side chain motions consist predominantly of librational (restricted motion) with occasional ring flip motions in the fast sub-nanosecond time scale. Due to the qualitatively different energetics governing the contributions from methyl- and aromatic amino acids, a cumulative estimate of protein conformational entropy changes from both methyl- and aromatic amino acids has yet to be formulated.

Combining all of the strategies developed, i.e. the new ^{13}C isotopic labelling for aromatic amino acids (Chapter 3 and 4, the new empirical calibration for the entropy meter (Chapter 2), the role played by protein conformational entropy in the very high-affinity barnase-barstar interaction and the HBP(D24R)-histamine interaction is investigated in Chapter 5. Protein conformational entropy was found to play a pivotal role in maintaining the very high affinity observed for both barnase-barstar and HBP(D24R)-histamine

complexes where favourable changes in the conformational entropy offset unfavourable changes in solvent and rotational-translational entropies. This entropic compensation mechanism has not been previously characterized as protein conformational entropy has typically been neglected. Moreover, protein conformational entropy was also found to act as a driving force in the absence of enthalpic contributions in the moderate affinity barnase-dCGAC interaction. The observation that protein conformational entropy can not only contribute favourable to ligand binding but also act as an independent driving force strongly suggests that the rational design of pharmaceutical drugs must utilize protein conformational entropy changes as an additional dimension, apart from enthalpy and solvent entropy.

By combining NMR relaxation methods and hydration measurements using reverse micelle encapsulation, I have shown that ligand binding induces global conformational dynamics change together with change in hydration/water dynamics across the surface of the protein (Chapter 5). This observation proves crucial to the formulation of the 'Univeral Entropy Meter' where 34 different protein-ligand complexes each with widely varying binding affinity and ligands are shown to agree excellently with the entropy meter formulation. The resulting 'Universal Entropy Meter' is the first comprehensive study where the different entropic contributions to the Gibbs free energy of binding have been calculated for several different protein-ligand complexes. Furthermore, it provides a platform from which the entropic contributions for any protein-ligand complex can be accurately estimated.

7.2 Future Directions

The results presented here represent a significant step forward in both understanding the role played by protein conformational entropy in very high affinity interactions as well as the different microscopic contributions to the protein conformational entropy arising from methyl- and aromatic- side chains. However, a cumulative estimate of the protein conformational entropy change by NMR from both methyl- and aromatic amino acids together with the changes in motion of the backbone amides has yet to be formulated. Such a formulation though not expected to result in significant deviations from the 'Universal Entropy Meter' would however represent a true global estimate of the changes in protein conformational entropy. Future work will be directed towards formulating an expression for using the changes in motions of methyl-, aromatic amino acid side chains and backbone amides together in the empirical calibration of the entropy meter.

In addition, In Chapter 4 the temperature dependence of the aromatic amino acid side chain motions revealed an intriguing transition behaviour centered at $\sim$ 37 degree C. The generality and significance of such a transition with respect to protein function is still an open question. This would help provide further insight into the energetics governing aromatic amino acid motions within the hydrophobic protein milieu. Such a description would also help understand the effect of correlation motion between aromatic amino acids and methyl bearing amino acids which has been hinted at in previous studies (refer to Chapter 4 and [116]).

Finally, the observation that ligand binding induces changes in both conformational dynamics across proteins as well as changes to the hydration dynamics across protein surfaces could yet reveal the role played by solvent in protein function. The limited number of studies have so far hampered our understanding of the role played by the dynamics of water, if any. In particular, hydration measurements using reverse micelles has hinted at clusters of water dynamics across protein surfaces [140], especially surfaces involved in binding interactions. Given the importance of pharmaceutical drug design in disrupting binding interfaces, the role played by water could yet prove crucial. Future studies looking at the relationship between ligand binding induced conformational dynamics changes and the change in solvent dynamics would shed valuable light on this issue. In particular a detailed understanding of the physics governing how ligand binding affects the dynamics of side chains far removed from the interface could prove very important.

Appendix A

Appendix Chapter 2

Table A.1: Methyl order parameters and Side chain entropies of ubiquitin from MD simulations

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
1	MET	3	0.485	0.286
2	GLN	3		0.588
3	ILE	2	0.740	0.090
4	PHE	2		0.387
5	VAL	1	0.793	0.143
6	LYS	4		0.543
7	THR	1	0.805	0.000
8	LEU	2	0.232	0.521
9	THR	1	0.360	0.426
11	LYS	4		0.408
12	THR	1	0.857	0.043
13	ILE	2	0.594	0.304
14	THR	1	0.753	0.172
15	LEU	2	0.502	0.322
16	GLU	3		0.452
17	VAL	1	0.818	0.012
18	GLU	3		0.714
19	PRO	2		0.342
20	SER	1		0.746

Continued on next page

Table A.1 Ubiquitin – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
21	ASP	2		0.575
22	THR	1	0.865	0.023
23	ILE	2	0.728	0.089
24	GLU	3		0.503
25	ASN	2		0.792
26	VAL	1	0.860	0.002
27	LYS	4		0.259
29	LYS	4		0.466
30	ILE	2	0.829	0.125
31	GLN	3		0.428
32	ASP	2		0.732
33	LYS	4		0.644
34	GLU	3		0.464
36	ILE	2	0.753	0.204
37	PRO	2		0.312
38	PRO	2		0.285
39	ASP	2		0.731
40	GLN	3		0.551
41	GLN	3		0.197
42	ARG	5		0.467
43	LEU	2	0.473	0.375
44	ILE	2	0.202	0.510
45	PHE	2		0.180
48	LYS	4		0.752
49	GLN	3		0.455
50	LEU	2	0.790	0.054
51	GLU	3		0.631
52	ASP	2		0.530
54	ARG	5		0.403
55	THR	1	0.888	0.006
56	LEU	2	0.522	0.242
57	SER	1		0.719
58	ASP	2		0.368
59	TYR	2		0.301
60	ASN	2		0.647
61	ILE	2	0.340	0.380
62	GLN	3		0.656
63	LYS	4		0.558
64	GLU	3		0.403

Continued on next page

Table A.1 Ubiquitin – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
65	SER	1		0.464
66	THR	1	0.872	0.022
67	LEU	2	0.229	0.432
68	HSD	2		0.710
69	LEU	2	0.653	0.223
70	VAL	1	0.337	0.806
71	LEU	2	0.388	0.373
72	ARG	5		0.436
73	LEU	2	0.247	0.421
74	ARG	5		0.342

Table A.2: Methyl order parameters and Side chain entropies of alpha3D from MD simulations

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
1	MET	3	0.041	0.808
3	SER	1		0.791
4	TRP	2		0.162
6	GLU	3		0.544
7	PHE	2		0.672
8	LYS	4		0.247
9	GLN	3		0.514
10	ARG	5		0.237
11	LEU	2	0.353	0.318
14	ILE	2	0.278	0.601
15	LYS	4		0.466
16	THR	1	0.534	0.461
17	ARG	5		0.200
18	LEU	2	0.740	0.039
19	GLN	3		0.468
21	LEU	2	0.383	0.223
24	SER	1		0.620
25	GLU	3		0.422
27	GLU	3		0.604
28	LEU	2	0.327	0.322
31	PHE	2		0.626
32	GLU	3		0.409

Continued on next page

Table A.2 alpha3D – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
33	LYS	4		0.450
34	GLU	3		0.573
35	ILE	2	0.233	0.622
38	PHE	2		0.497
39	GLU	3		0.626
40	SER	1		0.726
41	GLU	3		0.619
42	LEU	2	0.384	0.323
43	GLN	3		0.524
45	TYR	2		0.383
46	LYS	4		0.335
48	LYS	4		0.462
50	ASN	2		0.506
51	PRO	2		0.342
52	GLU	3		0.556
53	VAL	1	0.381	0.637
54	GLU	3		0.552
56	LEU	2	0.592	0.197
57	ARG	5		0.420
58	LYS	4		0.467
59	GLU	3		0.597
63	ILE	2	0.660	0.154
64	ARG	5		0.444
65	ASP	2		0.732
66	GLU	3		0.570
67	LEU	2	0.103	0.485
68	GLN	3		0.700
70	TYR	2		0.639
71	ARG	5		0.344
72	HSD	2		0.807
73	ASN	2		0.548

Table A.3: Methyl order parameters and Side chain entropies of CaM-smMLCK from MD simulations

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
5	THR	1	0.420	0.555
				Continued on next page

Table A.3 CaM-smMLCK – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
6	GLU	3		0.898
7	GLU	3		0.890
8	GLN	3		0.772
9	ILE	2	0.288	0.618
11	GLU	3		0.656
12	PHE	2		0.479
13	LYS	4		0.674
14	GLU	3		0.735
16	PHE	2		0.298
17	SER	1		0.875
18	LEU	2	0.355	0.362
19	PHE	2		0.543
20	ASP	2		0.515
21	LYS	4		0.737
22	ASP	2		0.522
24	ASP	2		0.103
26	THR	1	0.468	0.483
27	ILE	2	0.631	0.236
28	THR	1	0.747	0.110
29	THR	1	0.516	0.498
30	LYS	4		0.722
31	GLU	3		0.451
32	LEU	2	0.634	0.228
34	THR	1	0.773	0.098
35	VAL	1	0.531	0.343
36	MET	3	0.292	0.407
37	ARG	5		0.725
38	SER	1		0.498
39	LEU	2	0.378	0.303
41	GLN	3		0.702
42	ASN	2		0.505
43	PRO	2		0.276
44	THR	1	0.378	0.791
45	GLU	3		0.680
47	GLU	3		0.770
48	LEU	2	0.556	0.189
49	GLN	3		0.380
50	ASP	2		0.880
51	MET	3	0.235	0.639

Continued on next page

Table A.3 CaM-smMLCK – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
52	ILE	2	0.396	0.428
53	ASN	2		0.374
54	GLU	3		0.739
55	VAL	1	0.667	0.230
56	ASP	2		0.323
58	ASP	2		0.465
60	ASN	2		0.574
62	THR	1	0.425	0.540
63	ILE	2	0.713	0.144
64	ASP	2		0.080
65	PHE	2		0.448
66	PRO	2		0.340
67	GLU	3		0.123
68	PHE	2		0.016
69	LEU	2	0.270	0.417
70	THR	1	0.649	0.091
71	MET	3	0.318	0.389
72	MET	3	0.370	0.341
74	ARG	5		0.358
75	LYS	4		0.681
76	MET	3	0.112	0.764
77	LYS	4		0.703
78	ASP	2		0.808
79	THR	1	0.390	0.484
80	ASP	2		0.631
81	SER	1		0.863
82	GLU	3		0.676
83	GLU	3		0.623
84	GLU	3		0.505
85	ILE	2	0.301	0.469
86	ARG	5		0.562
87	GLU	3		0.463
89	PHE	2		0.004
90	ARG	5		0.318
91	VAL	1	0.761	0.073
92	PHE	2		0.508
93	ASP	2		0.391
94	LYS	4		0.561
95	ASP	2		0.394

Continued on next page

Table A.3 CaM-smMLCK – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
97	ASN	2		0.130
99	TYR	2		0.350
100	ILE	2	0.789	0.024
101	SER	1		0.785
104	GLU	3		0.354
105	LEU	2	0.768	0.045
106	ARG	5		0.268
107	HSD	2		0.498
108	VAL	1	0.510	0.448
109	MET	3	0.276	0.420
110	THR	1	0.593	0.338
111	ASN	2		0.164
112	LEU	2	0.540	0.227
114	GLU	3		0.470
115	LYS	4		0.782
116	LEU	2	0.208	0.543
117	THR	1	0.733	0.181
118	ASP	2		0.540
119	GLU	3		0.887
120	GLU	3		0.672
121	VAL	1	0.283	0.740
122	ASP	2		0.572
123	GLU	3		0.716
124	MET	3	0.516	0.308
125	ILE	2	0.498	0.222
126	ARG	5		0.459
127	GLU	3		0.523
129	ASP	2		0.011
130	ILE	2	0.321	0.599
131	ASP	2		0.706
133	ASP	2		0.271
135	GLN	3		0.558
136	VAL	1	0.812	0.000
137	ASN	2		0.771
138	TYR	2		0.423
139	GLU	3		0.747
140	GLU	3		0.147
141	PHE	2		0.284
142	VAL	1	0.391	0.757

Continued on next page

Table A.3 CaM-smMLCK – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
143	GLN	3		0.661
144	MET	3	0.095	0.602
145	MET	3	0.254	0.595
146	THR	1	0.390	0.457
797	ARG	5		0.441
798	ARG	5		0.409
799	LYS	4		0.490
800	TRP	2		0.494
801	GLN	3		0.517
802	LYS	4		0.508
803	THR	1	0.756	0.502
805	HSD	2		0.410
807	VAL	1	0.686	0.199
808	ARG	5		0.082
810	ILE	2	0.542	0.603
812	ARG	5		0.291
813	LEU	2	0.602	0.345
814	SER	1		0.962
815	SER	1		0.625

Table A.4: Methyl order parameters and Side chain entropies of CaM-nNOS from MD simulations

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
2	ASP	2		0.683
3	GLN	3		0.835
4	LEU	2	0.280	0.321
5	THR	1	0.607	0.171
6	GLU	3		0.791
7	GLU	3		0.833
8	GLN	3		0.638
9	ILE	2	0.307	0.573
11	GLU	3		0.722
12	PHE	2		0.411
13	LYS	4		0.764
14	GLU	3		0.772
16	PHE	2		0.016

Continued on next page

Table A.4 CaM-nNOS – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
17	SER	1		0.730
18	LEU	2	0.477	0.241
19	PHE	2		0.494
20	ASP	2		0.212
21	LYS	4		0.731
22	ASP	2		0.527
24	ASP	2		0.097
26	THR	1	0.496	0.587
27	ILE	2	0.783	0.031
28	THR	1	0.830	0.000
29	THR	1	0.538	0.396
30	LYS	4		0.764
31	GLU	3		0.661
32	LEU	2	0.673	0.104
34	THR	1	0.625	0.361
35	VAL	1	0.500	0.407
36	MET	3	0.411	0.329
37	ARG	5		0.753
38	SER	1		0.084
39	LEU	2	0.798	0.000
41	GLN	3		0.240
42	ASN	2		0.516
43	PRO	2		0.107
44	THR	1	0.350	0.631
45	GLU	3		0.701
47	GLU	3		0.800
48	LEU	2	0.592	0.164
49	GLN	3		0.582
50	ASP	2		0.736
51	MET	3	0.158	0.630
52	ILE	2	0.434	0.341
53	ASN	2		0.534
54	GLU	3		0.599
55	VAL	1	0.289	0.742
56	ASP	2		0.001
58	ASP	2		0.417
60	ASN	2		0.609
62	THR	1	0.828	0.098
63	ILE	2	0.603	0.184

Continued on next page

Table A.4 CaM-nNOS – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
64	ASP	2		0.110
65	PHE	2		0.335
66	PRO	2		0.361
67	GLU	3		0.118
68	PHE	2		0.002
69	LEU	2	0.486	0.253
70	THR	1	0.524	0.435
71	MET	3	0.212	0.487
72	MET	3	0.142	0.571
74	ARG	5		0.632
75	LYS	4		0.494
76	MET	3	0.106	0.739
77	LYS	4		0.664
78	ASP	2		0.772
79	THR	1	0.250	0.705
80	ASP	2		0.684
81	SER	1		0.927
82	GLU	3		0.816
83	GLU	3		0.726
84	GLU	3		0.697
85	ILE	2	0.505	0.368
86	ARG	5		0.707
87	GLU	3		0.572
89	PHE	2		0.003
90	ARG	5		0.684
91	VAL	1	0.817	0.050
92	PHE	2		0.455
93	ASP	2		0.225
94	LYS	4		0.705
95	ASP	2		0.306
97	ASN	2		0.137
99	TYR	2		0.238
100	ILE	2	0.846	0.001
101	SER	1		0.760
104	GLU	3		0.130
105	LEU	2	0.731	0.081
106	ARG	5		0.645
107	HSD	2		0.573
108	VAL	1	0.700	0.147

Continued on next page

Table A.4 CaM-nNOS – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
109	MET	3	0.183	0.467
110	THR	1	0.685	0.215
111	ASN	2		0.307
112	LEU	2	0.275	0.405
114	GLU	3		0.392
115	LYS	4		0.769
116	LEU	2	0.232	0.576
117	THR	1	0.787	0.006
118	ASP	2		0.560
119	GLU	3		0.859
120	GLU	3		0.687
121	VAL	1	0.157	0.987
122	ASP	2		0.590
123	GLU	3		0.721
124	MET	3	0.237	0.473
125	ILE	2	0.505	0.311
126	ARG	5		0.730
127	GLU	3		0.691
129	ASP	2		0.115
130	ILE	2	0.312	0.507
131	ASP	2		0.395
133	ASP	2		0.286
135	GLN	3		0.652
136	VAL	1	0.774	0.000
137	ASN	2		0.570
138	TYR	2		0.316
139	GLU	3		0.773
140	GLU	3		0.002
141	PHE	2		0.121
142	VAL	1	0.762	0.095
143	GLN	3		0.689
144	MET	3	0.409	0.297
145	MET	3	0.219	0.566
146	THR	1	0.499	0.456
148	LYS	4		0.746
5	ILE	2	0.450	0.415
7	PHE	2		0.470
8	LYS	4		0.566
9	LYS	4		0.503

Continued on next page

Table A.4 CaM-nNOS – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
10	LEU	2	0.797	0.021
12	GLU	3		0.649
14	VAL	1	0.834	0.013
15	LYS	4		0.342
16	PHE	2		0.493
17	SER	1		0.507
19	LYS	4		0.687
20	LEU	2	0.431	0.256
21	MET	3	0.259	0.652
23	GLN	3		0.748

Table A.5: Methyl order parameters and Side chain entropies of Adipocytle lipid binding protein (ADBP) from MD simulations

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
1	CYS	1		0.708
2	ASP	2		0.549
4	PHE	2		0.512
5	VAL	1	0.307	0.737
7	THR	1	0.759	0.194
8	TRP	2		0.004
9	LYS	4		0.719
10	LEU	2	0.642	0.188
11	VAL	1	0.396	0.590
12	SER	1		0.724
13	SER	1		0.081
14	GLU	3		0.604
15	ASN	2		0.393
16	PHE	2		0.000
17	ASP	2		0.423
18	ASP	2		0.526
19	TYR	2		0.000
20	MET	3	0.428	0.323
21	LYS	4		0.722
22	GLU	3		0.349
23	VAL	1	0.890	0.000
25	VAL	1	0.572	0.327

Continued on next page

Table A.5 ADBP – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
27	PHE	2		0.777
29	THR	1	0.805	0.075
30	ARG	5		0.265
31	LYS	4		0.723
32	VAL	1	0.137	0.953
35	MET	3	0.071	0.811
37	LYS	4		0.691
38	PRO	2		0.302
39	ASN	2		0.320
40	MET	3	0.376	0.419
41	ILE	2	0.279	0.353
42	ILE	2	0.821	0.035
43	SER	1		0.857
44	VAL	1	0.630	0.197
45	ASN	2		0.640
47	ASP	2		0.526
48	LEU	2	0.156	0.441
49	VAL	1	0.861	0.003
50	THR	1	0.727	0.226
51	ILE	2	0.819	0.026
52	ARG	5		0.215
53	SER	1		0.689
54	GLU	3		0.618
55	SER	1		0.490
56	THR	1	0.655	0.256
57	PHE	2		0.477
58	LYS	4		0.690
59	ASN	2		0.127
60	THR	1	0.773	0.149
61	GLU	3		0.401
62	ILE	2	0.740	0.135
63	SER	1		1.029
64	PHE	2		0.014
65	LYS	4		0.715
66	LEU	2	0.640	0.163
68	VAL	1	0.604	0.412
69	GLU	3		0.460
70	PHE	2		0.000
71	ASP	2		0.455

Continued on next page

Table A.5 ADBP – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
72	GLU	3		0.397
73	ILE	2	0.185	0.660
74	THR	1	0.886	0.000
76	ASP	2		0.458
77	ASP	2		0.578
78	ARG	5		0.193
79	LYS	4		0.706
80	VAL	1	0.722	0.174
81	LYS	4		0.646
82	SER	1		0.525
83	ILE	2	0.642	0.276
84	ILE	2	0.614	0.193
85	THR	1	0.566	0.402
86	LEU	2	0.194	0.463
87	ASP	2		0.573
91	LEU	2	0.256	0.375
92	VAL	1	0.816	0.088
93	GLN	3		0.227
94	VAL	1	0.730	0.192
95	GLN	3		0.205
96	LYS	4		0.713
97	TRP	2		0.022
98	ASP	2		0.685
100	LYS	4		0.692
101	SER	1		0.995
102	THR	1	0.891	0.000
103	THR	1	0.712	0.244
104	ILE	2	0.334	0.351
105	LYS	4		0.712
106	ARG	5		0.263
107	LYS	4		0.690
108	ARG	5		0.472
109	ASP	2		0.755
111	ASP	2		0.730
112	LYS	4		0.696
113	LEU	2	0.252	0.323
114	VAL	1	0.855	0.024
115	VAL	1	0.893	0.000
116	GLU	3		0.495

Continued on next page

Table A.5 ADBP – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
117	CYS	1		0.071
118	VAL	1	0.433	0.571
119	MET	3	0.573	0.382
120	LYS	4		0.706
122	VAL	1	0.427	0.566
123	THR	1	0.825	0.078
124	SER	1		0.000
125	THR	1	0.863	0.029
126	ARG	5		0.175
127	VAL	1	0.872	0.026
128	TYR	2		0.001
129	GLU	3		0.535
130	ARG	5		0.501

Table A.6: Methyl order parameters and Side chain entropies of Hen egg white lysozyme (HEWL) from MD simulations

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
1	LYS	4		0.496
2	VAL	1	0.529	0.481
3	PHE	2		0.324
5	ARG	5		0.331
7	GLU	3		0.640
8	LEU	2	0.814	0.020
12	MET	3	0.686	0.252
13	LYS	4		0.425
14	ARG	5		0.405
15	HSD	2		0.129
17	LEU	2	0.612	0.173
18	ASP	2		0.496
19	ASN	2		0.506
20	TYR	2		0.224
21	ARG	5		0.463
23	TYR	2		0.204
24	SER	1		0.707
25	LEU	2	0.580	0.109
27	ASN	2		0.306

Continued on next page

Table A.6 HEWL – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
28	TRP	2		0.097
29	VAL	1	0.857	0.020
33	LYS	4		0.492
34	PHE	2		0.235
35	GLU	3		0.463
36	SER	1		0.344
37	ASN	2		0.731
38	PHE	2		0.136
39	ASN	2		0.387
40	THR	1	0.893	0.000
41	GLN	3		0.526
43	THR	1	0.343	0.763
44	ASN	2		0.268
45	ARG	5		0.438
46	ASN	2		0.033
47	THR	1	0.300	0.595
48	ASP	2		0.486
50	SER	1		0.002
51	THR	1	0.876	0.000
52	ASP	2		0.514
53	TYR	2		0.069
55	ILE	2	0.494	0.233
56	LEU	2	0.743	0.045
57	GLN	3		0.005
58	ILE	2	0.494	0.288
59	ASN	2		0.132
60	SER	1		0.000
61	ARG	5		0.314
62	TRP	2		0.656
63	TRP	2		0.271
65	ASN	2		0.431
66	ASP	2		0.532
68	ARG	5		0.324
69	THR	1	0.853	0.125
70	PRO	2		0.329
72	SER	1		0.755
73	ARG	5		0.472
74	ASN	2		0.304
75	LEU	2	0.334	0.429

Continued on next page

Table A.6 HEWL – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
77	ASN	2		0.630
78	ILE	2	0.489	0.270
79	PRO	2		0.293
81	SER	1		0.781
83	LEU	2	0.828	0.023
84	LEU	2	0.812	0.074
85	SER	1		0.774
86	SER	1		0.851
87	ASP	2		0.799
88	ILE	2	0.788	0.045
89	THR	1	0.698	0.302
91	SER	1		0.186
92	VAL	1	0.857	0.023
93	ASN	2		0.373
96	LYS	4		0.309
97	LYS	4		0.372
98	ILE	2	0.808	0.057
99	VAL	1	0.576	0.114
100	SER	1		0.862
101	ASP	2		0.700
103	ASN	2		0.562
105	MET	3	0.531	0.426
106	ASN	2		0.513
108	TRP	2		0.424
109	VAL	1	0.396	0.572
111	TRP	2		0.000
112	ARG	5		0.531
113	ASN	2		0.412
114	ARG	5		0.559
116	LYS	4		0.589
118	THR	1	0.145	0.920
119	ASP	2		0.856
120	VAL	1	0.283	0.752
121	GLN	3		0.479
123	TRP	2		0.226
124	ILE	2	0.647	0.136
125	ARG	5		0.523
128	ARG	5		0.433
129	LEU	2	0.474	0.218

Table A.7: Methyl order parameters and Side chain entropies of Cytochrome c2 (Cyt c2) from MD simulations

Residue No.	Residue Type	Nchi	MD O2axis	PInP/Nchi
2	ASP	2		0.838
5	LYS	4		0.167
7	GLU	3		0.445
8	LYS	4		0.627
9	GLU	3		0.524
10	PHE	2		0.000
11	ASN	2		0.372
12	LYS	4		0.249
13	CYS	1		0.000
14	LYS	4		0.480
15	THR	1	0.309	0.674
16	CYS	1		0.265
17	HSD	2		0.002
18	SER	1		0.885
19	ILE	2	0.606	0.225
20	ILE	2	0.519	0.233
22	PRO	2		0.337
23	ASP	2		0.669
25	THR	1	0.735	0.140
26	GLU	3		0.555
27	ILE	2	0.174	0.457
28	VAL	1	0.634	0.322
29	LYS	4		0.505
32	LYS	4		0.617
33	THR	1	0.682	0.242
35	PRO	2		0.143
36	ASN	2		0.308
37	LEU	2	0.849	0.000
38	TYR	2		0.039
40	VAL	1	0.899	0.000
41	VAL	1	0.706	0.226
43	ARG	5		0.235
44	THR	1	0.728	0.232
47	THR	1	0.626	0.291
48	TYR	2		0.000
49	PRO	2		0.319

Continued on next page

Table A.7 Cyt c2 – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
50	GLU	3		0.658
51	PHE	2		0.012
52	LYS	4		0.754
53	TYR	2		0.015
54	LYS	4		0.572
55	ASP	2		0.645
56	SER	1		0.023
57	ILE	2	0.539	0.266
58	VAL	1	0.807	0.264
60	LEU	2	0.753	0.038
63	SER	1		0.766
65	PHE	2		0.270
67	TRP	2		0.250
68	THR	1	0.872	0.017
69	GLU	3		0.685
70	GLU	3		0.825
71	ASP	2		0.187
72	ILE	2	0.532	0.273
74	THR	1	0.889	0.000
75	TYR	2		0.000
76	VAL	1	0.227	0.953
77	LYS	4		0.612
78	ASP	2		0.554
79	PRO	2		0.351
82	PHE	2		0.003
83	LEU	2	0.340	0.327
84	LYS	4		0.501
85	GLU	3		0.497
86	LYS	4		0.424
87	LEU	2	0.331	0.374
88	ASP	2		0.837
89	ASP	2		0.551
90	LYS	4		0.734
91	LYS	4		0.602
93	LYS	4		0.719
94	THR	1	0.597	0.000
96	MET	3	0.882	0.131
98	PHE	2		0.002
99	LYS	4		0.539

Continued on next page

Table A.7 Cyt c2 – Continued from previous page

Residue No.	Residue Type	Nchi	MD O2axis	PlnP/Nchi
100	LEU	2	0.279	0.338
102	LYS	4		0.754
105	GLU	3		0.631
106	ASP	2		0.316
107	VAL	1	0.313	0.673
110	TYR	2		0.000
111	LEU	2	0.824	0.000
113	SER	1		0.840
114	VAL	1	0.518	0.524
115	VAL	1	0.280	0.681
116	LYS	4		0.671

Appendix B

Appendix Chapter 4

Table B.1: Temperature dependence of ubiquitin aromatic order parameter and internal correlation times

Temperature	F45 O2	F45 te (e-10)	Y59 O2	F45 te (e-10)
283	0.832 ± 0.015	1.4 ± 0.021	0.973 ± 0.010	1.00E-03
293	0.845 ± 0.010	1.24 ± 0.015	0.981 ± 0.010	1.00E-03
303	0.850 ± 0.017	1.28 ± 0.017	0.989 ± 0.013	1.00E-03
308	0.764 ± 0.013	2.16 ± 0.020	0.912 ± 0.012	5.88 ± 0.040
313	0.646 ± 0.019	2.84 ± 0.030	0.773 ± 0.015	6.36 ± 0.034
318	0.533 ± 0.019	3.56 ± 0.070	0.619 ± 0.016	5.56 ± 0.013
323	0.511 ± 0.018	4.56 ± 0.040	0.580 ± 0.014	5.72 ± 0.021
328	0.484 ± 0.020	4.84 ± 0.030	0.565 ± 0.019	5.80 ± 0.030

Table B.2: Temperature dependence of ubiquitin aromatic order parameter and internal correlation times II

Temperature	F4 O2	F4 te (e-10)
283	0.721 ± 0.009	1.60 ± 0.012
293	0.673 ± 0.010	1.44 ± 0.012
303	0.696 ± 0.011	1.84 ± 0.024
308	0.592 ± 0.013	2.40 ± 0.030
313	0.506 ± 0.013	2.56 ± 0.021
318	0.421 ± 0.015	3.48 ± 0.013

Continued on next page

Table B.2 – Continued from previous page

Temperature	F4 O2	F4 te (e-10)
323	0.375 ± 0.015	3.76 ± 0.040
328	0.366 ± 0.016	3.92 ± 0.017

Table B.3: Pressure dependence of ubiquitin F45 aromatic order parameter.

Temperature	P = 1 bar	P = 1200 bar	P = 2500 bar
283	0.832 ± 0.015	0.787 ± 0.013	0.837 ± 0.011
303	0.850 ± 0.017	0.907 ± 0.010	0.912 ± 0.012
323	0.511 ± 0.511	0.547 ± 0.016	0.601 ± 0.015

Table B.4: Pressure dependence of ubiquitin Y59 aromatic order parameter.

Temperature	P = 1 bar	P = 1200 bar	P = 2500 bar
283	0.973 ± 0.010	0.973 ± 0.012	0.963 ± 0.011
303	0.989 ± 0.013	0.998 ± 0.012	0.998 ± 0.011
323	0.580 ± 0.014	0.621 ± 0.014	0.671 ± 0.013

Table B.5: Pressure dependence of ubiquitin F4 aromatic order parameter.

Temperature	P = 1 bar	P = 1200 bar	P = 2500 bar
283	0.721 ± 0.009	0.666 ± 0.014	0.691 ± 0.011
303	0.696 ± 0.011	0.726 ± 0.014	0.746 ± 0.013
323	0.375 ± 0.015	0.421 ± 0.015	0.479 ± 0.013

Table B.6: Pressure dependence of ubiquitin F45 aromatic internal correlation times.

Temperature	P = 1 bar (e-10)	P = 1200 bar (e-10)	P = 2500 bar (e-10)
283	1.4 ± 0.021	2.28 ± 0.040	1.92 ± 0.014
303	1.28 ± 0.017	0.161 ± 0.024	1.36 ± 0.034
323	4.56 ± 0.040	5.12 ± 0.030	4.08 ± 0.021

Table B.7: Pressure dependence of ubiquitin Y59 aromatic internal correlation times.

Temperature	P = 1 bar (e-10)	P = 1200 bar (e-10)	P = 2500 bar (e-10)
283	1.00E-03	9.10 ± 0.034	9.40 ± 0.030
303	1.00E-03	9.80 ± 0.034	9.70 ± 0.030
323	5.72 ± 0.021	6.76 ± 0.030	9.90 ± 0.03

Table B.8: Pressure dependence of ubiquitin F4 aromatic internal correlation times.

Temperature	P = 1 bar (e-10)	P = 1200 bar (e-10)	P = 2500 bar (e-10)
283	1.60 ± 0.012	1.68 ± 0.013	1.32 ± 0.011
303	1.84 ± 0.024	1.28 ± 0.020	1.48 ± 0.024
323	3.76 ± 0.040	3.96 ± 0.024	3.48 ± 0.030

Table B.9: Temperature dependence of ubiquitin aromatic order parameter from 120ns MD simulations

Temperature	F45 O2	Y59 O2	F4 O2
283	0.811	0.871	0.815
293	0.655	0.876	0.784
303	0.442	0.873	0.356
308	0.379	0.854	0.328
313	0.364	0.863	0.330
323	0.506	0.868	0.315

Table B.10: Pressure dependence of ubiquitin F45 aromatic order parameter from MD simulations.

Temperature	**P = 1 bar**	**P = 1200 bar**	**P = 2500 bar**
293	0.710	0.690	0.695
323	0.390	0.420	0.490

Table B.11: Pressure dependence of ubiquitin Y59 aromatic order parameter from MD simulations.

Temperature	**P = 1 bar**	**P = 1200 bar**	**P = 2500 bar**
293	0.980	0.985	0.982
323	0.590	0.615	0.680

Table B.12: Pressure dependence of ubiquitin F4 aromatic order parameter from MD simulations.

Temperature	**P = 1 bar**	**P = 1200 bar**	**P = 2500 bar**
293	0.820	0.795	0.830
323	0.505	0.535	0.605

Appendix C

Appendix Chapter 5

Table C.1: Backbone order parameters of free barnase and barnase in complex with barstar

Residue No.	Residue Name	O2 free	O2 complex
4	ILE	0.865	0.809
5	ASN	0.897	
6	THR	0.953	0.905
7	PHE	0.910	0.925
8	ASP	0.900	0.938
9	GLY	0.915	1.000
10	VAL	0.915	0.923
12	ASP	0.960	0.845
13	TYR	0.940	0.950
14	LEU	0.945	
15	GLN	0.970	
16	THR	0.902	0.907
17	TYR	0.985	
19	LYS	0.928	
20	LEU	0.862	
22	ASP	0.875	
23	ASN	0.933	0.867
24	TYR	0.872	0.817
25	ILE	0.935	
26	THR	0.940	0.910
27	LYS	0.900	0.943
28	SER	0.965	0.907
29	GLU	0.980	0.925

Continued on next page

Table C.1 Barnase amide O2 – Continued from previous page

Residue No.	Residue Name	O2 free	O2 complex
30	ALA	0.938	
31	GLN	0.960	0.910
32	ALA	0.948	0.832
33	LEU	0.912	
34	GLY	0.985	
36	VAL	0.912	1.000
37	ALA	0.912	0.860
40	GLY	0.905	0.973
41	ASN	0.980	
43	ALA	0.940	0.995
44	ASP	0.910	
45	VAL	0.897	1.000
46	ALA	0.930	0.822
48	GLY	0.867	0.865
50	SER	0.907	
51	ILE	0.930	
52	GLY	0.917	
53	GLY	0.882	
54	ASP	0.948	0.867
55	ILE	0.897	
56	PHE	0.925	
61	GLY	0.930	0.877
62	LYS		0.978
63	LEU	0.905	
65	GLY	0.862	0.862
69	ARG	0.860	
70	THR		0.789
71	TRP	0.935	0.915
72	ARG	0.955	
73	GLU	0.933	
74	ALA	0.960	
75	ASP	0.980	
76	ILE	0.915	
77	ASN	0.943	0.985
80	SER	0.847	0.887
81	GLY	0.885	0.915
82	PHE	0.915	
85	SER	0.860	0.960
86	ASP	0.925	0.779

Continued on next page

Table C.1 Barnase amide O2 – Continued from previous page

Residue No.	Residue Name	O2 free	O2 complex
87	ARG	0.948	
88	ILE	0.938	
89	LEU	0.890	
90	TYR	0.960	1.000
91	SER	0.930	
92	SER	0.958	
93	ASP	0.958	0.988
94	TRP	0.910	0.993
96	ILE	0.907	
97	TYR	0.897	
98	LYS	0.928	
99	THR	0.975	
100	THR	0.933	
103	TYR	1.000	
104	GLN	0.970	
105	THR	0.923	0.925
106	PHE	0.917	
107	THR	0.890	0.968
108	LYS	0.865	0.890
110	ARG	0.940	0.928

Table C.2: Backbone order parameters of free barstar and barstar in complex with barnase

Residue No.	Residue Name	O2 free	O2 complex
2	LYS	0.923	0.887
3	ALA	0.923	0.905
4	VAL	0.905	
6	ASN	0.940	
7	GLY	0.925	
8	GLU	0.980	0.897
9	GLN	0.938	0.980
10	ILE	0.900	0.892
11	ARG	0.885	
12	SER	0.840	0.912
13	ILE	0.875	0.935
14	SER	0.885	1.000

Continued on next page

Table C.2 Barstar amide O2 – Continued from previous page

Residue No.	Residue Name	O2 free	O2 complex
15	ASP	0.744	
16	LEU	0.923	0.900
17	HIS	0.890	0.935
19	THR	0.905	1.000
21	LYS	0.920	
22	LYS	0.928	0.968
23	GLU	0.923	0.968
24	LEU	0.945	0.867
25	ALA	0.935	0.973
26	LEU	0.917	1.000
28	GLU	0.865	
29	TYR	0.923	0.845
30	TYR	0.935	0.960
31	GLY	0.907	
33	ASN	0.875	
34	LEU		0.598
35	ASP		0.500
36	ALA	0.907	
38	TRP	0.973	0.978
38	TRP	0.892	
40	CYS	0.923	0.935
41	LEU	0.958	
42	THR	0.938	
43	GLY	0.860	
45	VAL	0.950	
46	GLU	0.963	0.968
47	TYR	0.880	0.905
49	LEU	0.978	0.503
50	VAL	0.935	
51	LEU	0.985	1.000
52	GLU	0.970	
57	GLU	0.975	
58	GLN	0.907	0.887
60	LYS	0.958	0.945
61	GLN	0.935	
62	LEU	0.912	
63	THR	0.930	0.938
64	GLU	0.960	
65	ASN	0.983	

Continued on next page

Table C.2 Barstar amide O2 – Continued from previous page

Residue No.	Residue Name	O2 free	O2 complex
66	GLY	0.963	1.000
68	GLU	0.887	0.948
69	SER	0.915	0.920
70	VAL	0.915	0.960
71	LEU	0.923	
72	GLN	0.950	0.970
73	VAL	0.985	0.960
74	PHE	0.928	0.980
75	ARG	0.935	0.990
76	GLU	0.935	
77	ALA	0.990	0.917
79	ALA	0.807	0.958
80	GLU	0.950	
81	GLY	0.897	0.867
82	CYS	0.963	0.938
83	ASP	0.923	0.923
84	ILE	0.885	0.902
85	THR	0.965	

Table C.3: Methyl side chain order parameters of free barnase and barnase in complex with barstar

Residue No.	Residue Name	Residue Type	O2 free	O2 complex
1	ALA	CB	0.35055	0.2763
3	VAL	CG2	0.6327	0.37035
3	VAL	CG1	0.50895	0.48915
4	ILE	CD1	0.57825	0.4842
4	ILE	CG2	0.97425	0.7119
6	THR	CG2	0.99405	0.68715
10	VAL	CG2	0.99405	0.9396
10	VAL	CG1	0.8505	0.74655
11	ALA	CB	0.99405	0.85545
14	LEU	CD2	0.58815	0.52875
14	LEU	CD1	0.74655	0.80595
16	THR	CG2	0.95445	0.83565
20	LEU	CD2	0.8208	0.6822
20	LEU	CD1	0.83565	0.81585

Continued on next page

Table C.3 Barnase methyl O2axis – Continued from previous page

Residue No.	Residue Name	Residue Type	O2axis free	O2axis complex
25	ILE	CD1	0.77625	0.6525
25	ILE	CG2	0.24165	0.7614
26	THR	CG2	0.9792	0.99405
30	ALA	CB	0.99405	0.94455
32	ALA	CB	0.99405	0.9693
33	LEU	CD1	0.7119	0.6129
33	LEU	CD2	0.7317	0.4347
36	VAL	CG1	0.90495	0.6822
37	ALA	CB		0.91485
42	LEU	CD2	0.71685	0.73665
42	LEU	CD1	0.85545	0.67725
43	ALA	CB	0.99405	0.9693
45	VAL	CG1	0.96435	0.84555
45	VAL	CG2	0.85545	0.7911
46	ALA	CB	0.99405	0.99405
51	ILE	CD1	0.4644	0.45945
51	ILE	CG2	0.8406	0.69705
55	ILE	CD1	0.3852	0.27135
55	ILE	CG2	0.8406	0.6723
63	LEU	CD2	0.69705	0.52875
63	LEU	CD1	0.7317	0.68715
70	THR	CG2	0.4644	
74	ALA	CB	0.99405	0.99405
76	ILE	CG2	0.83565	0.8901
76	ILE	CD1	0.64755	0.74655
79	THR	CG2	0.84555	0.40995
88	ILE	CD1	0.7812	0.81585
88	ILE	CG2	0.5337	0.7317
89	LEU	CD1	0.7218	0.63765
89	LEU	CD2	0.62775	0.6426
95	LEU	CD1	0.43965	0.36045
95	LEU	CD2	0.4941	0.40005
96	ILE	CD1	0.6525	0.603
96	ILE	CG2	0.99405	0.8307
99	THR	CG2	0.99405	0.6921
107	THR	CG2	0.97425	
109	ILE	CG2	0.74655	0.63765

Table C.4: Methyl side chain order parameters of free barstar and barstar in complex with barnase

Residue No.	Residue Name	Residue Type	O2 free	O2 complex
3	ALA	CB	0.994	
4	VAL	CG1	0.786	0.756
4	VAL	CG2	0.821	0.806
5	ILE	CG2	0.920	0.752
5	ILE	CD1	0.464	0.281
10	ILE	CD1	0.657	0.459
10	ILE	CG2	0.900	0.875
13	ILE	CD1	0.514	0.420
13	ILE	CG2	0.549	0.331
16	LEU	CD2		0.474
19	THR	CG2	0.831	
20	LEU	CD2	0.727	0.593
20	LEU	CD1	0.826	0.662
24	LEU	CD1	0.549	0.440
24	LEU	CD2	0.539	
25	ALA	CB	0.994	0.930
26	LEU	CD1	0.994	0.811
26	LEU	CD2	0.915	0.816
34	LEU	CD1	0.707	0.425
34	LEU	CD2	0.727	
36	ALA	CB	0.994	0.276
37	LEU	CD1		0.687
37	LEU	CD2	0.989	0.989
41	LEU	CD1	0.895	0.717
41	LEU	CD2	0.761	0.544
42	THR	CG2	0.994	
45	VAL	CG1	0.925	0.712
45	VAL	CG2	0.994	
49	LEU	CD2		0.430
50	VAL	CG1		0.831
50	VAL	CG2	0.563	
51	LEU	CD2	0.539	0.519
51	LEU	CD1	0.534	0.420
62	LEU	CD1	0.450	0.360
63	THR	CG2	0.940	
67	ALA	CB	0.994	0.994
70	VAL	CG1	0.969	
71	LEU	CD2	0.890	0.761

Continued on next page

Table C.4 Barstar methyl O2axis – Continued from previous page

Residue No.	Residue Name	Residue Type	O2axis free	O2axis complex
73	VAL	CG1	0.969	
73	VAL	CG2	0.910	0.776
77	ALA	CB	0.994	0.994
79	ALA	CB	0.994	0.969
84	ILE	CD1	0.811	0.707
84	ILE	CG2	0.717	0.113
85	THR	CG2	0.994	0.905
86	ILE	CD1	0.662	0.563
86	ILE	CG2	0.875	0.811
87	ILE	CD1	0.608	0.430
87	ILE	CG2	0.841	0.905
88	LEU	CD1	0.588	0.405
88	LEU	CD2	0.534	0.425

Table C.5: Backbone order parameters of free barnase and barnase in complex with 4mer dCGAC

Residue No.	Residue Name	O2 free	O2 complex
4	ILE	0.741	0.865
5	ASN	0.955	0.897
6	THR	0.928	0.953
7	PHE	0.933	0.910
8	ASP	0.887	0.900
9	GLY	0.960	0.915
10	VAL	0.978	0.915
12	ASP	0.945	0.960
13	TYR	0.970	0.940
14	LEU	0.938	0.945
15	GLN	0.938	0.970
16	THR	0.840	0.902
17	TYR	0.970	0.985
19	LYS	0.963	0.928
20	LEU	0.890	0.862
22	ASP	0.875	0.875
23	ASN	0.940	0.933
24	TYR	0.890	0.872
25	ILE	0.905	0.935

Continued on next page

Table C.5 Barnase-DNA amide O2 – Continued from previous page

Residue No.	Residue Name	O2 free	O2 complex
26	THR	0.877	0.940
27	LYS	0.759	0.900
28	SER	0.930	0.965
29	GLU	0.965	0.980
30	ALA	0.973	0.938
31	GLN	0.963	0.960
32	ALA	0.985	0.948
33	LEU	0.907	0.912
34	GLY	0.988	0.985
36	VAL	0.847	0.912
37	ALA	0.870	0.912
40	GLY	0.907	0.905
41	ASN	0.917	0.980
43	ALA	0.895	0.940
44	ASP	0.829	0.910
45	VAL	0.933	0.897
46	ALA	0.907	0.930
48	GLY	0.827	0.867
50	SER	0.832	0.907
51	ILE	0.960	0.930
52	GLY	0.880	0.917
53	GLY	0.882	0.882
54	ASP	0.928	0.948
55	ILE	0.912	0.897
56	PHE	1.000	0.925
61	GLY	0.925	0.930
63	LEU	0.920	0.905
65	GLY	0.777	0.862
69	ARG	0.787	0.860
71	TRP	0.963	0.935
72	ARG	0.968	0.955
73	GLU	0.983	0.933
74	ALA	0.923	0.960
75	ASP	0.892	0.980
76	ILE	0.840	0.915
77	ASN	0.960	0.943
80	SER	0.779	0.847
81	GLY	0.872	0.885
82	PHE	0.950	0.915

Continued on next page

Table C.5 Barnase-DNA amide O2 – Continued from previous page

Residue No.	Residue Name	O2 free	O2 complex
85	SER	1.000	0.860
86	ASP	1.000	0.925
87	ARG	1.000	0.948
88	ILE	0.928	0.938
89	LEU	0.970	0.890
90	TYR	0.983	0.960
91	SER	0.925	0.930
92	SER	0.910	0.958
93	ASP	0.930	0.958
94	TRP	0.970	0.910
96	ILE	0.882	0.907
97	TYR	0.834	0.897
98	LYS	0.930	0.928
99	THR	0.945	0.975
100	THR	0.892	0.933
103	TYR	0.925	1.000
104	GLN	0.930	0.970
105	THR	0.902	0.923
106	PHE	0.915	0.917
107	THR	0.875	0.890
108	LYS	0.827	0.865
110	ARG	0.892	0.940

Table C.6: Methyl side chain order parameters of free barnase and barnase in complex with 4mer dCGAC

Residue No.	Residue Name	Residue Type	O2 free	O2 complex
1	ALA	CB	0.291	0.351
3	VAL	CG2	0.489	0.633
3	VAL	CG1	0.455	0.509
4	ILE	CD1	0.484	0.578
4	ILE	CG2	0.821	0.974
6	THR	CG2	0.994	0.994
10	VAL	CG2	0.994	0.994
10	VAL	CG1	0.900	0.851
11	ALA	CB	0.930	0.994
14	LEU	CD2	0.756	0.588

Continued on next page

Table C.6 Barnase-DNA methyl O2axis – Continued from previous page

Residue No.	Residue Name	Residue Type	O2axis free	O2axis complex
14	LEU	CD1	0.761	0.747
16	THR	CG2	0.940	0.954
20	LEU	CD2	0.752	0.821
20	LEU	CD1	0.821	0.836
25	ILE	CD1	0.722	0.776
25	ILE	CG2	0.148	0.242
26	THR	CG2	0.895	0.979
30	ALA	CB	0.994	0.994
32	ALA	CB	0.994	0.994
33	LEU	CD1	0.672	0.712
33	LEU	CD2	0.633	0.732
36	VAL	CG1	0.860	0.905
42	LEU	CD2	0.623	0.717
42	LEU	CD1	0.836	0.855
43	ALA	CB	0.994	0.994
45	VAL	CG1	0.860	0.964
45	VAL	CG2	0.890	0.855
46	ALA	CB	0.994	0.994
51	ILE	CD1	0.435	0.464
51	ILE	CG2	0.964	0.841
55	ILE	CD1	0.296	0.385
55	ILE	CG2	0.717	0.841
63	LEU	CD2	0.712	0.697
63	LEU	CD1	0.796	0.732
70	THR	CG2	0.450	0.464
74	ALA	CB	0.994	0.994
76	ILE	CG2	0.841	0.836
76	ILE	CD1	0.653	0.648
79	THR	CG2	0.697	0.846
88	ILE	CD1	0.667	0.781
88	ILE	CG2	0.558	0.534
89	LEU	CD1	0.727	0.722
89	LEU	CD2	0.643	0.628
95	LEU	CD1	0.326	0.440
95	LEU	CD2	0.455	0.494
96	ILE	CD1	0.440	0.653
96	ILE	CG2	0.920	0.994
99	THR	CG2	0.514	0.994
105	THR	CG2	0.628	

Continued on next page

Table C.6 Barnase-DNA methyl O2axis – Continued from previous page

Residue No.	Residue Name	Residue Type	O2axis free	O2axis complex
107	THR	CG2	0.930	0.974
109	ILE	CG2	0.687	0.747

Table C.7: Aromatic side chain order parameters of free barnase and barnase in complex with barstar

Residue No.	Residue Name	Residue Type	O2 free	O2 complex
7	PHE	CD1	0.9977	0.9477
13	TYR	CD1	0.9577	0.9977
17	TYR	CD1	0.8543	0.9010
24	TYR	CD1	-	-
35	TRP	CE3	0.8210	0.7177
56	PHE	CD1	0.9643	-
71	TRP	CE3	0.8510	0.9710
78	TYR	CD1	0.9843	0.9977
82	PHE	CD1	0.5077	0.8110
90	TYR	CD1	0.9510	-
94	TRP	CE3	0.9910	0.7577
97	TYR	CD1	0.9110	-
103	TYR	CD1	0.8777	0.9777
107	PHE	CD1	-	0.9977

Table C.8: Aromatic side chain order parameters of free barstar and barstar in complex with barnase

Residue No.	Residue Name	Residue Type	O2 free	O2 complex
29	TYR	CD1	0.7814	0.997667
30	TYR	CD1	0.8751	0.984333
38	TRP	CE3	0.8954	0.957667
44	TRP	CE3	0.7421	0.9613
47	TYR	CD1	-	0.937667
53	TRP	CE3	0.9752	0.9894
56	PHE	CD1	-	0.854333
74	PHE	CD1	0.615	0.431

Table C.9: Aromatic side chain order parameters of free barnase and barnase in complex with dCGAC

Residue No.	Residue Name	Residue Type	O2 free	O2 complex
7	PHE	CD1	0.9977	0.1577
13	TYR	CD1	0.9577	0.9877
17	TYR	CD1	0.8543	0.9077
24	TYR	CD1	-	0.9943
35	TRP	CE3	0.8210	0.6810
56	PHE	CD1	0.9643	0.9810
71	TRP	CE3	0.8510	0.7877
78	TYR	CD1	0.9843	0.8843
82	PHE	CD1	0.5077	0.6310
90	TYR	CD1	0.9510	0.9843
94	TRP	CE3	0.9910	0.9043
97	TYR	CD1	0.9110	0.9210
103	TYR	CD1	0.8777	0.9843
106	PHE	CD1	-	0.9910

Table C.10: Barnase hydration measurement (NOE/ROE) in the free state and complex states

Residue	NOE/ROE with barstar	NOE/ROE with DNA	NOE/ROE free
i4	-0.31383		
t6	-0.10345	-0.07387	-0.08169
f7	-0.47402	-0.25567	-0.38696
g9	-0.06525	-0.08617	-0.10501
l14	-0.13517		
t16	-0.08312	-0.08328	-0.23599
y17	-0.49935		
k19	-0.39349		
l20		-0.32465	
d22	-0.49648		
y24	-0.33556		
t26	-0.27772	-0.1488	-0.14807
k27	-0.23279		-0.15923

Continued on next page

Table C.10 Barnase hydration measurement – Continued from previous page

Residue	NOE/ROE with barstar	NOE/ROE with DNA	NOE/ROE free
s28	-0.18553	-0.24764	-0.26124
e29	-0.12183	-0.07486	-0.09739
a32	-0.28779		
v36	-0.40917		-0.30863
a37	-0.50979		0.432066
s38	-0.43713		
g40	-0.13553	-0.10721	-0.14698
a43	-0.17571		
a46	-0.14341		
g48	-0.46469		
k49	0.61671		
i51		-0.09801	-0.13606
g52		-0.13967	
d54	-0.14205		-0.4889
i55			-0.30815
f56	-0.50978		
r59	-0.2313		
e60	-0.37694		
g61	-0.2326		-0.26064
k62			-0.10144
l63	-0.2588		-0.18161
t70	-0.15308	-0.13538	-0.16441
w71	-0.50909	-0.13968	-0.21597
e73	0.339393		
d75		-0.16303	-0.19034
i76		-0.13831	
n77	-0.43194		-0.18003
y78	-0.05796		
g79	-0.11372		
s80	-0.4138	-0.06004	-0.50303
g81	-0.17613	-0.12627	-0.15769
f82	-0.92315	-0.3257	-0.06354
r83	-0.13494		
n84	-0.01543		
s85	-0.27545		
d86	-0.25101		-0.14958
l89	-0.46293		
s92		-0.19384	-0.20693
d93	-0.07211		

Continued on next page

Table C.10 Barnase hydration measurement – Continued from previous page

Residue	NOE/ROE with barstar	NOE/ROE with DNA	NOE/ROE free
w94	0.460922		
t99	0.829938		
t100		-0.06178	-0.08667
h102	0.167052	-0.47705	-0.31539
t105	-0.04828	-0.07617	-0.06124
f106		-0.25614	-0.38166
t107	-0.30529	-0.22862	-0.17293
k108	0.986259	-0.41877	
i109		-0.09269	-0.10588
r110		-0.1713	
w35ne1	-0.22122		-0.12526
w94ne	-0.09177		-0.71506

Table C.11: Barstar hydration measurement (NOE/ROE) complex state bound to barnase

Residue	NOE/ROE with barnase
q9	-0.02295
s12	-0.19827
i13	0.900329
d15	-0.24767
n23	-0.50716
l24	-0.17259
a25	-0.08076
y29	-0.23839
y30	-0.30865
g31	-0.40998
a36	-0.4317
l37	-0.47051
w38	-0.46375
w44	-0.47758
y47	-0.40967
l49	-0.4124
e52	-0.48031
q58	-0.2804
l62	-0.10756
s69	-0.50004

Continued on next page

Table C.11 Barstar hydration measurement – Continued from previous page

Residue	NOE/ROE with barnase
v70	-0.15717
a77	0.159592
g81	-0.32998
c82	-0.07514
i84	-0.29326
t85	-0.7571

Appendix D

Appendix Chapter 6

Table D.1: List of all protein-ligand complexes used in Chapter 7

Index No.	Protein-Ligand	K_d (nM)	PDB code (free, bound)
1	CaM:MKKa(p)	3.5	1XO2, 1CKK
2	CaM:smMLCK(p)	24	1XO2, 1CDL
3	CaM(E84K):smMLCk(p)	4.5	1XO2, 1CDL
4	CaM:CaMKI(p)	3.4	1XO2, 1MXE
5	CaM:eNOS(p)	3.4	1XO2, 1NIW
6	CaM:nNOS(p)	1.5	1XO2, 2O60
7	CAP:cAMP2:DNA	2200	1G6N, 1CGP
8	CAP(D53H):cAMP2:DNA	2200	1G6N, 1CGP
9	CAP(S62F):cAMP2:DNA	2200	1G6N, 1CGP
10	CAP(T127LS128I):DNA	2200	2WC2, 1CGP
11	CAP(T127LS128I):cAMP2:DNA	5000	1G6N, 1CGP
12	CAP(G141S):DNA	11000	2WC2, 1CGP
13	CAP(G141S):cAMP2:DNA	2200	1G6N, 1CGP
14	CAP(G141S):cGMP2:DNA	2200	1G6N, 1CGP
15	CAP(A144T):DNA	25000	2WC2, 1CGP
16	CAP(A144T):cAMP2:DNA	2200	1G6N, 1CGP
17	CAP(A144T):cGMP2:DNA	2200	1G6N, 1CGP
18	Galectin:L2	18200	1A3K, 2XG3
19	Galectin:L3	3300	1A3K, 1KJR
20	Galectin:Lactose	230000	1A3K, 2NN8
21	HEWL:Chitotriose	15000	1LZA, 1LZB
22	PDZ3:CRIPT	2000	1BFE, 1BE9
23	PDZ3?7:CRIPT	36000	1BFE, 1BE9
24	DHFR:NADP:Folate	8000	1RX7, 1RX2

Continued on next page

Table D.1 – Continued from previous page

Index No.	Protein-Ligand	K_d (nM)	PDB code (free, bound)
25	SAP SH2:Y281	200	1D1Z, 1D4T
26	SAP SH2:pY281	900	1D1Z, 1D4W
27	PDZ2:RaGEF	8000	1GM1, 1D5G
28	CaM(D58N):smMLCK(p)	4	1XO2, 1CDL
29	CaM(D95N):smMLCK(p)	5	1XO2, 1CDL
30	CaM(M124L):smMLCK(p)	30	1XO2, 1CDL
31	CaM(E84K):nNOS(p)	3.4	1XO2, 1CDL
32	Cdc42HS:PBD46	0.012	2QRZ, 1EES
33	Barnase:Barstar	0.00001	1BNR, 1BTA, 1BRS
34	Barnase:dCGAC	50000	1BNR, 1BRN
35	HBP(D24R):Histamine	2.5	3GAQ, 3G7X

Table D.2: Thermodynamics and NMR relaxation parameters for the calibration of the 'Entropy Meter'.

Index No.	$N_\chi . < \Delta O^2_{axis} >$	ΔS_{total} kcal/mol/K	ΔASA_{polar} Å^2	ΔASA_{apolar} Å^2
1	46	-0.072	-1778	-1965
2	56	-0.062	-1824	-2433
3	28	-0.033	-1867	-2386
4	55	-0.056	-1836	-2421
5	38	-0.021	-1661	-2248
6	39	-0.002	-1618	-2000
7	30.179	-0.039	-1085	-1451
8	19.063	-0.01	-1085	-1451
9	-19.338	0.046	-1085	-1451
10	-17.113	0.039	-1085	-1451
11	14.655	-0.005	-1085	-1451
12	-6.078	0.01	-1085	-1451
13	11.787	-0.011	-1085	-1451
14	-12.708	0.043	-1085	-1451
15	3.223	0.013	-1085	-1451
16	-6.676	0.036	-1085	-1451
17	-26.567	0.062	-1085	-1451
18	-14.076	-0.017	-432.5	-621.8
19	-6.072	-0.021	-65.6	-19.7
20	-17.723	-0.024	-219.4	-256.8
21	4.38	-0.027	-127.3	-201.4

Continued on next page

Table D.2 – Continued from previous page

Index No.	$N_\chi . < \Delta O^2_{axis} >$	ΔS_{total} **kcal/mol/K**	ΔASA_{polar} Å^2	ΔASA_{apolar} Å^2
22	-0.864	-0.033	-246.5	-155.3
23	23.209	-0.034	-242.1	-89.9
24	13.779	-0.019	-238.4	-354.1
25	-5.154	-0.025	-425.4	-538.7
26	-7.72	-0.03	-387.6	-420.4
27	-1.57	-0.008	-226.8	-59.2
28	48.076	-0.066	-1824	-2433
29	53.337	-0.057	-1824	-2433
30	36.561	-0.06	-1824	-2433
31	27.827	0.007	-1618	-2000
32	28.16	0.008	-1104.3	-950.9
33	-38	0.001	-707	-894
34	-12.44	0.004	-160	-150
35	-8.029	-0.019	-435.9	-469.9

Table D.3: Thermodynamics and NMR relaxation parameters for the calibration of the 'Entropy Meter'.

Index No.	$\Delta S^a_{solvent}$ **kcal/mol/K**	$\Delta S^b_{solvent}$ **kcal/mol/K**
1	0.136	0.079
2	0.13	0.071
3	0.135	0.075
4	0.131	0.072
5	0.118	0.064
6	0.119	0.067
7	0.081	0.042
8	0.081	0.042
9	0.081	0.042
10	0.081	0.042
11	0.081	0.042
12	0.081	0.042
13	0.081	0.042
14	0.081	0.042
15	0.081	0.042
16	0.081	0.042
17	0.081	0.042
18	0.031	0.016

Continued on next page

Table D.3 – Continued from previous page

Index No.	$\Delta S^{a}_{solvent}$ kcal/mol/K	$\Delta S^{b}_{solvent}$ kcal/mol/K
19	0.007	0.004
20	0.018	0.009
21	0.009	0.004
22	0.024	0.014
23	0.026	0.015
24	0.017	0.008
25	0.032	0.017
26	0.031	0.017
27	0.025	0.015
28	0.13	0.071
29	0.13	0.071
30	0.13	0.071
31	0.119	0.067
32	0.101	0.056
33	-0.01	-0.01
34	0.007	0.007
35	0.036	0.001

[a]The change in solvent entropy was calculated according to Hilser and Freire ([10]). Both the free and bound 3D structures had both identical atom type and number of atoms prior to the calculation of change in accessible surface area.
[b]The change in solvent entropy was obtained by simultaneous calibration of the conformational entropy and the solvent entropy using Equation 6.4 in Chapter 7.

Bibliography

[1] Malcolm H Levitt. *Spin dynamics: basics of nuclear magnetic resonance*. John Wiley & Sons, 2008.

[2] JM Berg, JL Tymoczko, and L Stryer. Biochemistry. sixth., 2007.

[3] Tony Pawson and Piers Nash. Protein–protein interactions define specificity in signal transduction. *Genes & Development*, 14(9):1027–1047, 2000.

[4] Clinton D Stoner. Inquiries into the nature of free energy and entropy in respect to biochemical thermodynamics. *Entropy*, 2(3):106–141, 2000.

[5] Michael M Pierce, CS Raman, and Barry T Nall. Isothermal titration calorimetry of protein–protein interactions. *Methods*, 19(2):213–221, 1999.

[6] Joël Janin, Ranjit P Bahadur, and Pinak Chakrabarti. Protein–protein interaction and quaternary structure. *Quarterly reviews of biophysics*, 41(02):133–180, 2008.

[7] Ozlem Keskin, Attila Gursoy, Buyong Ma, and Ruth Nussinov. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chemical reviews*, 108(4):1225–1244, 2008.

[8] Loredana Lo Conte, Cyrus Chothia, and Joël Janin. The atomic structure of protein-protein recognition sites. *Journal of molecular biology*, 285(5):2177–2198, 1999.

[9] Tim Clackson and James A Wells. A hot spot of binding energy in a hormone-receptor interface. *Science*, 267(5196):383–386, 1995.

[10] Vincent J Hilser, Bertrand García-Moreno E, Terrence G Oas, Greg Kapp, and Steven T Whitten. A statistical thermodynamic model of the protein ensemble. *Chemical reviews*, 106(5):1545–1558, 2006.

[11] M Karplus, T Ichiye, and BM Pettitt. Configurational entropy of native proteins. *Biophysical journal*, 52(6):1083–1085, 1987.

[12] Andrew L Lee and A Joshua Wand. Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature*, 411(6836):501–504, 2001.

[13] Kendra King Frederick, Michael S Marlow, Kathleen G Valentine, and A Joshua Wand. Conformational entropy in molecular recognition by proteins. *Nature*, 448 (7151):325–329, 2007.

[14] Michael S Marlow, Jakob Dogan, Kendra K Frederick, Kathleen G Valentine, and A Joshua Wand. The role of conformational entropy in molecular recognition by calmodulin. *Nature chemical biology*, 6(5):352–358, 2010.

[15] Daiwen Yang and Lewis E Kay. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *Journal of molecular biology*, 263(2):369–382, 1996.

[16] Zhigang Li, Soumya Raychaudhuri, and A Joshua Wand. Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein science*, 5(12): 2647–2650, 1996.

[17] Tatyana I Igumenova, Kendra King Frederick, and A Joshua Wand. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chemical reviews*, 106(5):1672–1699, 2006.

[18] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *Journal of the American Chemical Society*, 104(17):4546–4559, 1982.

[19] M Goldman. Interference effects in the relaxation of a pair of unlike spin-12 nuclei. *Journal of Magnetic Resonance (1969)*, 60(3):437–452, 1984.

[20] Gaetano Barbato, Mitsuhiko Ikura, Lewis E Kay, Richard W Pastor, and Ad Bax. Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry*, 31(23):5269–5278, 1992.

[21] Nico Tjandra, Scott E Feller, Richard W Pastor, and Ad Bax. Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation. *Journal of the American Chemical Society*, 117(50):12562–12566, 1995.

[22] Michael J Osborne, Jason Schnell, Stephen J Benkovic, H Jane Dyson, and Peter E Wright. Backbone dynamics in dihydrofolate reductase complexes: role of loop flexibility in the catalytic mechanism. *Biochemistry*, 40(33):9846–9859, 2001.

[23] Elan Zohar Eisenmesser, Daryl A Bosco, Mikael Akke, and Dorothee Kern. Enzyme dynamics during catalysis. *Science*, 295(5559):1520–1523, 2002.

[24] A Joshua Wand. The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Current opinion in structural biology*, 23(1):75–81, 2013.

[25] Kim A Sharp, Vignesh Kasinath, and A Joshua Wand. Banding 2of nmr-derived methyl order parameters: Implications for protein dynamics. *Proteins: Structure, Function, and Bioinformatics*, 2014.

[26] Vignesh Kasinath, Kathleen G Valentine, and A Joshua Wand. A 13C labeling strategy reveals a range of aromatic side chain motion in calmodulin. *Journal of the American Chemical Society*, 135(26):9560–9563, 2013.

[27] Shiou-Ru Tzeng and Charalampos G Kalodimos. Protein activity regulation by conformational entropy. *Nature*, 488(7410):236–240, 2012.

[28] M Michael Gromiha, N Saranya, S Selvaraj, B Jayaram, and Kazuhiko Fukui. Sequence and structural features of binding site residues in protein-protein complexes: comparison with protein-nucleic acid complexes. *Proteome Sci*, 9(Suppl 1):S13, 2011.

[29] Ken A Dill. Dominant forces in protein folding. *Biochemistry*, 29(31):7133–7155, 1990.

[30] Vincent J Hilser, James O Wrabl, and Hesam N Motlagh. Structural and energetic basis of allostery. *Annual review of biophysics*, 41:585, 2012.

[31] Alan Cooper. Thermodynamic fluctuations in protein molecules. *Proceedings of the National Academy of Sciences*, 73(8):2740–2741, 1976.

[32] A Cooper and DTF Dryden. Allostery without conformational change. *European Biophysics Journal*, 11(2):103–109, 1984.

[33] Neil A Farrow, Ranjith Muhandiram, Alex U Singer, Steven M Pascal, Cyril M Kay, Gerry Gish, Steven E Shoelson, Tony Pawson, Julie D Forman-Kay, and Lewis E Kay. Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, 33(19): 5984–6003, 1994.

[34] DR Muhandiram, Toshio Yamazaki, Brian D Sykes, and Lewis E Kay. Measurement of 2H T1 and T1. rho. relaxation times in uniformly 13C-labeled and fractionally 2H-labeled proteins in solution. *Journal of the American Chemical Society*, 117(46):11536–11544, 1995.

[35] Rieko Ishima, Aneta P Petkova, John M Louis, and Dennis A Torchia. Comparison of methyl rotation axis order parameters derived from model-free analyses of 2H and 13C longitudinal and transverse relaxation rates measured in the same protein sample. *Journal of the American Chemical Society*, 123(25):6164–6171, 2001.

[36] Oscar Millet, DR Muhandiram, Nikolai R Skrynnikov, and Lewis E Kay. Deuterium spin probes of side-chain dynamics in proteins. 1. measurement of five relaxation rates per deuteron in 13C-labeled and fractionally 2H-enriched proteins in solution. *Journal of the American Chemical Society*, 124(22):6439–6448, 2002.

[37] Vitali Tugarinov and Lewis E Kay. Quantitative 13C and 2H NMR relaxation studies of the 723-residue enzyme malate synthase G reveal a dynamic binding interface. *Biochemistry*, 44(49):15970–15977, 2005.

[38] David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.

[39] Ron O Dror, Robert M Dirks, JP Grossman, Huafeng Xu, and David E Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, 41:429–452, 2012.

[40] Kyle A Beauchamp, Yu-Shan Lin, Rhiju Das, and Vijay S Pande. Are protein force fields getting better? a systematic benchmark on 524 diverse NMR measurements. *Journal of chemical theory and computation*, 8(4):1409–1414, 2012.

[41] Alexander D Scouras and Valerie Daggett. The dynameomics rotamer library: Amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. *Protein Science*, 20(2):341–352, 2011.

[42] Mark T Nelson, William Humphrey, Attila Gursoy, Andrew Dalke, Laxmikant V Kalé, Robert D Skeel, and Klaus Schulten. NAMD: a parallel, object-oriented molecular dynamics program. *International Journal of High Performance Computing Applications*, 10(4):251–268, 1996.

[43] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, 26(16):1781–1802, 2005.

[44] Bernard R Brooks, Charles L Brooks, Alexander D MacKerell, Lennart Nilsson, Robert J Petrella, Benoît Roux, Youngdo Won, Georgios Archontis, Christian Bartels, Stefan Boresch, et al. CHARMM: the biomolecular simulation program. *Journal of computational chemistry*, 30(10):1545–1614, 2009.

[45] Bernard R Brooks, Robert E Bruccoleri, Barry D Olafson, David J States, S Swaminathan, and Martin Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2):187–217, 1983.

[46] Scott TR Walsh, Andrew L Lee, William F DeGrado, and A Joshua Wand. Dynamics of a de novo designed three-helix bundle protein studied by 15N, 13C, and 2H NMR relaxation methods. *Biochemistry*, 40(32):9560–9569, 2001.

[47] Keith L Constantine, Mark S Friedrichs, Michael Wittekind, Haris Jamil, Ching-Hsuen Chu, Rex A Parker, Valentina Goldfarb, Luciano Mueller, and Bennett T Farmer. Backbone and side chain dynamics of uncomplexed human adipocyte and muscle fatty acid-binding proteins. *Biochemistry*, 37(22):7965–7980, 1998.

[48] Peter F Flynn, Ramona J Bieber Urbauer, Hui Zhang, Andrew L Lee, and A Joshua Wand. Main chain and side chain dynamics of a heme protein: 15N and 2H NMR relaxation studies of r. capsulatus ferrocytochrome c2. *Biochemistry*, 40 (22):6559–6569, 2001.

[49] Veronica R Moorman, Kathleen G Valentine, and A Joshua Wand. The dynamical response of hen egg white lysozyme to the binding of a carbohydrate ligand. *Protein Science*, 21(7):1066–1073, 2012.

[50] Andrew L Lee, Peter F Flynn, and A Joshua Wand. Comparison of 2H and 13C NMR relaxation techniques for the study of protein methyl group dynamics in solution. *Journal of the American Chemical Society*, 121(12):2891–2902, 1999.

[51] Senadhi Vijay-Kumar, Charles E Bugg, and William J Cook. Structure of ubiquitin refined at 1.8 Å resolution. *Journal of molecular biology*, 194(3):531–544, 1987.

[52] Scott TR Walsh, Hong Cheng, James W Bryson, Heinrich Roder, and William F DeGrado. Solution structure and dynamics of a de novo designed three-helix bundle protein. *Proceedings of the National Academy of Sciences*, 96(10):5486–5491, 1999.

[53] William E Meador, Anthony R Means, and Florante A Quiocho. Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science*, 257(5074):1251–1255, 1992.

[54] Katsumi Maenaka, Masaaki Matsushima, Haiwei Song, Futoshi Sunada, Kimitsuna Watanabe, and Izumi Kumagai. Dissection of protein–carbohydrate interactions in mutant hen egg-white lysozyme complexes and their hydrolytic activity. *Journal of molecular biology*, 247(2):281–293, 1995.

[55] Zhaohui Xu, DA Bernlohr, and L Jo Banaszak. The adipocyte lipid-binding protein at 1.6 Å resolution. crystal structures of the apoprotein and with bound saturated and unsaturated fatty acids. *Journal of Biological Chemistry*, 268(11):7874–7884, 1993.

[56] Matthew M Benning, Gary Wesenberg, Michael S Caffrey, Robert G Bartsch, Terrance E Meyer, Michael A Cusanovich, Ivan Rayment, and Hazel M Holden.

Molecular structure of cytochrome c2 isolated from *rhodobacter capsulatus* determined at 2· 5 Å resolution. *Journal of molecular biology*, 220(3):673–685, 1991.

[57] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[58] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[59] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.

[60] David C Chatfield, Attila Szabo, and Bernard R Brooks. Molecular dynamics of staphylococcal nuclease: Comparison of simulation with 15N and 13C nmr relaxation data. *Journal of the American Chemical Society*, 120(21):5301–5311, 1998.

[61] Linda K Nicholson, Lewis E Kay, Donna M Baldisseri, Julian Arango, Paul E Young, Ad Bax, and Dennis A Torchia. Dynamics of methyl groups in proteins as studied by proton-detected carbon-13 NMR spectroscopy. application to the leucine residues of staphylococcal nuclease. *Biochemistry*, 31(23):5253–5263, 1992.

[62] Wei Chen, Michael K Gilson, Simon P Webb, and Michael J Potter. Modeling protein- ligand binding by Mining Minima. *Journal of chemical theory and computation*, 6(11):3540–3557, 2010.

[63] Martha S Head, James A Given, and Michael K Gilson. Mining minima: direct computation of conformational free energy. *The Journal of Physical Chemistry A*, 101(8):1609–1618, 1997.

[64] Bracken M King and Bruce Tidor. MIST: Maximum information spanning trees for dimension reduction of biological data sets. *Bioinformatics*, 25(9):1165–1172, 2009.

[65] Bracken M King, Nathaniel W Silver, and Bruce Tidor. Efficient calculation of molecular configurational entropies using an information theoretic approximation. *The Journal of Physical Chemistry B*, 116(9):2891–2904, 2012.

[66] Andrew L Lee, Kim A Sharp, James K Kranz, Xiang-Jin Song, and A Joshua Wand. Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry*, 41(46):13814–13825, 2002.

[67] J Andrew McCammon. Dynamics of folded proteins. *Nature*, 267:16, 1977.

[68] Dong Long, Da-Wei Li, Korvin FA Walter, Christian Griesinger, and Rafael Brüschweiler. Toward a predictive understanding of slow methyl group dynamics in proteins. *Biophysical journal*, 101(4):910–915, 2011.

[69] Scott A Showalter, Eric Johnson, Mark Rance, and Rafael Brüschweiler. Toward quantitative interpretation of methyl side-chain dynamics from NMR by molecular dynamics simulations. *Journal of the American Chemical Society*, 129(46): 14146–14147, 2007.

[70] RJ Wittebort and Attila Szabo. Theory of NMR relaxation in macromolecules: Restricted diffusion and jump models for multiple internal rotations in amino acid side chains. *The Journal of Chemical Physics*, 69(4):1722–1736, 1978.

[71] Dennis C Glass, Marimuthu Krishnan, Jeremy C Smith, and Jerome Baudry. Three entropic classes of side chain in a globular protein. *The Journal of Physical Chemistry B*, 117(11):3127–3134, 2013.

[72] James J Chou, David A Case, and Ad Bax. Insights into the mobility of methyl-bearing side chains in proteins from 3 J CC and 3 J CN couplings. *Journal of the American Chemical Society*, 125(29):8959–8966, 2003.

[73] Da-Wei Li, Scott A Showalter, and Rafael BruÌĹschweiler. Entropy localization in proteins. *The Journal of Physical Chemistry B*, 114(48):16036–16044, 2010.

[74] Kateri H DuBay and Phillip L Geissler. Calculation of proteins' total side-chain torsional entropy and its influence on protein–ligand interactions. *Journal of molecular biology*, 391(2):484–497, 2009.

[75] Amit Das, J Chakrabarti, and Mahua Ghosh. Conformational Contribution to Thermodynamics of Binding in Protein-Peptide Complexes through Microscopic Simulation. *Biophysical journal*, 104(6):1274–1284, 2013.

[76] Garland R Marshall. Limiting assumptions in structure-based design: binding entropy. *Journal of computer-aided molecular design*, 26(1):3–8, 2012.

[77] Anthony Mittermaier and Lewis E Kay. New tools provide new insights in NMR studies of protein dynamics. *Science*, 312(5771):224–228, 2006.

[78] Virginia A Jarymowycz and Martin J Stone. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chemical reviews*, 106(5):1624–1671, 2006.

[79] Andrew L Lee, Jeffrey L Urbauer, and A Joshua Wand. Improved labeling strategy for 13C relaxation measurements of methyl groups in proteins. *Journal of biomolecular NMR*, 9(4):437–440, 1997.

[80] David M LeMaster. Isotope labeling in solution protein assignment and structural analysis. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 26:371–419, 1994.

[81] David M LeMaster and Diana M Kushlan. Dynamical mapping of *e. coli* thioredoxin via 13C NMR relaxation analysis. *Journal of the American Chemical Society*, 118(39):9255–9264, 1996.

[82] A Joshua Wand, Ramona J Bieber, Jeffrey L Urbauer, Robert P McEvoy, and ZH Gan. Carbon relaxation in randomly fractionally 13C-enriched proteins. *Journal of Magnetic Resonance, Series B*, 108(2):173–175, 1995.

[83] Masatsune Kainosho and Peter Güntert. SAIL–stereo-array isotope labeling. *Quarterly reviews of biophysics*, 42(04):247–300, 2009.

[84] Kaare Teilum, Ulrika Brath, Patrik Lundström, and Mikael Akke. Biosynthetic 13C labeling of aromatic side chains in proteins for NMR relaxation measurements. *Journal of the American Chemical Society*, 128(8):2506–2507, 2006.

[85] Joshua A Boyer and Andrew L Lee. Monitoring Aromatic Picosecond to Nanosecond Dynamics in Proteins via 13C relaxation: Expanding perturbation mapping of the rigidifying core mutation, v54a, in eglin c. *Biochemistry*, 47(17):4876–4886, 2008.

[86] Patrik Lundström, Kaare Teilum, Tommy Carstensen, Irina Bezsonova, Silke Wiesner, D Flemming Hansen, Tomasz L Religa, Mikael Akke, and Lewis E Kay. Fractional 13C enrichment of isolated carbons using [1-13C]-or [2-13C]-glucose facilitates the accurate measurement of dynamics at backbone Cα and side-chain methyl positions in proteins. *Journal of biomolecular NMR*, 38(3):199–212, 2007.

[87] Kenneth A Thomas, Graham M Smith, Theresa B Thomas, and Richard J Feldmann. Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. *Proceedings of the National Academy of Sciences*, 79(16):4843–4847, 1982.

[88] SK Burley and GA Petsko. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science*, 229(4708):23–28, 1985.

[89] SK Burley and GA Petsko. Amino-aromatic interactions in proteins. *FEBS Letters*, 203(2):139–143, 1986.

[90] Sara Birtalan, Robert D Fisher, and Sachdev S Sidhu. The functional capacity of the natural amino acids for molecular recognition. *Molecular bioSystems*, 6(7): 1186–1194, 2010.

[91] Andrew A Bogan and Kurt S Thorn. Anatomy of hot spots in protein interfaces. *Journal of molecular biology*, 280(1):1–9, 1998.

[92] Gail J Bartlett, Craig T Porter, Neera Borkakoti, and Janet M Thornton. Analysis of catalytic residues in enzyme active sites. *Journal of molecular biology*, 324(1): 105–121, 2002.

[93] K Wüthrich and G Wagner. NMR investigations of the dynamics of the aromatic amino acid residues in the basic pancreatic trypsin inhibitor. *FEBS letters*, 50(2): 265–268, 1975.

[94] Edwin B. Kalan, Bernard D. Davis, P. R. Srinivasan, and David B. Sprinson. The conversion of various carbohydrates to 5-Dehydroshikimic acid by bacterial extracts. *Journal of Biological Chemistry*, 223(2):907–912, 1956.

[95] PR Srinivasan, Harol d T Shigeura, Milon Sprecher, David B Sprinson, and Bernard D Davis. The biosynthesis of shikimic acid from D-glucose. *Journal of Biological Chemistry*, 220(1):477–497, 1956.

[96] James K Kranz, Peter F Flynn, Ernesto J Fuentes, and A Joshua Wand. Dissection of the pathway of molecular recognition by calmodulin. *Biochemistry*, 41(8): 2599–2608, 2002.

[97] Weixia Liu, Peter F Flynn, Ernesto J Fuentes, James K Kranz, Margaret McCormick, and A Joshua Wand. Main chain and side chain dynamics of oxidized flavodoxin from *cyanobacterium anabaena*. *Biochemistry*, 40(49):14744–14753, 2001.

[98] Ninad V Prabhu, Andrew L Lee, A Joshua Wand, and Kim A Sharp. Dynamics and entropy of a calmodulin-peptide complex studied by NMR and molecular dynamics. *Biochemistry*, 42(2):562–570, 2003.

[99] Michael J Plevin, David L Bryce, and Jérôme Boisbouvier. Direct detection of CH/π interactions in proteins. *Nature chemistry*, 2(6):466–471, 2010.

[100] Hitoshi Kuboniwa, Nico Tjandra, Stephan Grzesiek, Hao Ren, Claude B Klee, and Ad Bax. Solution structure of calcium-free calmodulin. *Nature structural biology*, 2(9):768–776, 1995.

[101] Marcel Ottiger and Ad Bax. Determination of relative N-HN, N-C', Cα-C', and Cα-Hα Effective Bond Lengths in a Protein by NMR in a Dilute Liquid Crystalline Phase. *Journal of the American Chemical Society*, 120(47):12334–12341, 1998.

[102] Lishan Yao, Alexander Grishaev, Gabriel Cornilescu, and Ad Bax. Site-specific backbone amide 15N chemical shift anisotropy tensors in a small protein from liquid crystal and cross-correlated relaxation measurements. *Journal of the American Chemical Society*, 132(12):4295–4309, 2010.

[103] Martin J Dellwo and A Joshua Wand. Model-independent and model-dependent analysis of the global and internal dynamics of cyclosporin a. *Journal of the American Chemical Society*, 111(13):4571–4578, 1989.

[104] Chaohui Ye, Riqiang Fu, Jianzhi Hu, Lei Hou, and Shangwu Ding. Carbon-13 chemical shift anisotropies of solid amino acids. *Magnetic resonance in chemistry*, 31(8):699–704, 1993.

[105] Goeran Jonsaell and Per Ahlberg. One-bond carbon-13-carbon-13 coupling constants as a probe for carbocation structure. Doubly carbon-13 labeled 1, 4-bishomotropylium ion. *Journal of the American Chemical Society*, 108(13):3819–3824, 1986.

[106] Tatyana I Igumenova, Andrew L Lee, and A Joshua Wand. Backbone and side chain dynamics of mutant calmodulin-peptide complexes. *Biochemistry*, 44(38): 12627–12639, 2005.

[107] Vignesh Kasinath, Kim A Sharp, and A Joshua Wand. Microscopic insights into the NMR relaxation-based protein conformational entropy meter. *Journal of the American Chemical Society*, 135(40):15092–15100, 2013.

[108] Jack J Skalicky, Jeffrey L Mills, Surabhi Sharma, and Thomas Szyperski. Aromatic ring-flipping in supercooled water: implications for NMR-based structural biology of proteins. *Journal of the American Chemical Society*, 123(3):388–397, 2001.

[109] Gerhard Wagner. Activation volumes for the rotational motion of interior aromatic rings in globular proteins determined by high resolution 1H NMR at variable pressure. *FEBS letters*, 112(2):280–284, 1980.

[110] Mineyuki Hattori, Hua Li, Hiroaki Yamada, Kazuyuki Akasaka, Wolfgang Hengstenberg, Wolfram Gronwald, and Hans Robert Kalbitzer. Infrequent cavity-forming fluctuations in HPr from Staphylococcus carnosus revealed by pressure- and temperature-dependent tyrosine ring flips. *Protein science*, 13(12):3104–3114, 2004.

[111] Ulrich Weininger, Michal Respondek, Christian LoÌĹw, and Mikael Akke. Slow aromatic ring flips detected despite near-degenerate NMR frequencies of the exchanging nuclei. *The Journal of Physical Chemistry B*, 117(31):9241–9247, 2013.

[112] Roman J Lichtenecker, Katharina Weinhäupl, Walther Schmid, and Robert Konrat. α-Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression. *Journal of biomolecular NMR*, 57(4):327–331, 2013.

[113] Arthur G Palmer III, Remo A Hochstrasser, David P Millar, Mark Rance, and Peter E Wright. Characterization of amino acid side chain dynamics in a zinc-finger peptide using carbon-13 NMR spectroscopy and time-resolved fluorescence spectroscopy. *Journal of the American Chemical Society*, 115(14):6333–6345, 1993.

[114] T Michael Sabo, Davood Bakhtiari, Korvin FA Walter, Robert L McFeeters, Karin Giller, Stefan Becker, Christian Griesinger, and Donghan Lee. Thermal coefficients of the methyl groups within ubiquitin. *Protein Science*, 21(4):562–570, 2012.

[115] Michael H Glickman and Aaron Ciechanover. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiological reviews*, 82 (2):373–428, 2002.

[116] Yinan Fu, Vignesh Kasinath, Veronica R Moorman, Nathaniel V Nucci, Vincent J Hilser, and A Joshua Wand. Coupled motion in proteins revealed by pressure perturbation. *Journal of the American Chemical Society*, 134(20):8543–8550, 2012.

[117] Beatriz Ibarra-Molero, Vakhtang V Loladze, George I Makhatadze, and Jose M Sanchez-Ruiz. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry*, 38(25):8138–8149, 1999.

[118] Ronald M Levy and Robert P Sheridan. Combined effect of restricted rotational diffusion plus jumps on nuclear magnetic resonance and fluorescence probes of aromatic ring motions in proteins. *Biophysical journal*, 41(2):217–221, 1983.

[119] Richard J Wittebort, T Michael Rothgeb, Attila Szabo, and FR Gurd. Aliphatic groups of sperm whale myoglobin: 13C NMR study. *Proceedings of the National Academy of Sciences*, 76(3):1059–1063, 1979.

[120] Mike P Allen, Dominic J Tildesley, et al. Computer simulation of liquids, 1987.

[121] Kohki Ishikawa, Mika Okumura, Katsuo Katayanagi, Shigenobu Kimura, Shigenori Kanaya, Haruki Nakamura, and Kosuke Morikawa. Crystal structure of Ribonuclease H from *thermus thermophilus* HB8 Refined at 2· 8 Å resolution. *Journal of molecular biology*, 230(2):529–542, 1993.

[122] Ashley M Buckle, Gideon Schreiber, and Alan R Fersht. Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0 å. resolution. *Biochemistry*, 33(30):8878–8889, 1994.

[123] Ernesto Freire. A thermodynamic approach to the affinity optimization of drug candidates. *Chemical biology & drug design*, 74(5):468–472, 2009.

[124] Christof Winter, Andreas Henschel, Anne Tuukkanen, and Michael Schroeder. Protein interactions in 3D: From interface evolution to drug discovery. *Journal of structural biology*, 179(3):347–358, 2012.

[125] Tracy L Nero, Craig J Morton, Jessica K Holien, Jerome Wielens, and Michael W Parker. Oncogenic protein interfaces: small molecules, big challenges. *Nature Reviews Cancer*, 14(4):248–262, 2014.

[126] Christian Frisch, Gideon Schreiber, Christopher M Johnson, and Alan R Fersht. Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation. *Journal of molecular biology*, 267 (3):696–706, 1997.

[127] Joël Janin. Wet and dry interfaces: the role of solvent in protein–protein and protein-DNA recognition. *Structure*, 7(12):R277–R279, 1999.

[128] Yaakov Levy and José N Onuchic. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.*, 35:389–415, 2006.

[129] Motomu Shimaoka and Timothy A Springer. Therapeutic antagonists and conformational regulation of integrin function. *Nature Reviews Drug Discovery*, 2(9): 703–716, 2003.

[130] Terence A Kelly, Deborah D Jeanfavre, Daniel W McNeil, Joseph R Woska, Patricia L Reilly, Elizabeth A Mainolfi, Karen M Kishimoto, Gerald H Nabozny, Rosemarie Zinter, Barbara-Jean Bormann, et al. Cutting edge: a small molecule antagonist of LFA-1-mediated cell adhesion. *The Journal of Immunology*, 163(10): 5173–5177, 1999.

[131] Joseph R Woska Jr, Kathleen Last-Barney, Robert Rothlein, Rachel R Kroe, Patricia L Reilly, Deborah D Jeanfavre, Elizabeth A Mainolfi, Terence A Kelly, Gary O Caviness, Steven E Fogal, et al. Small molecule LFA-1 antagonists compete with an anti-LFA-1 monoclonal antibody for binding to the CD11a i domain: development of a flow-cytometry-based receptor occupancy assay. *Journal of immunological methods*, 277(1):101–115, 2003.

[132] GC Paesen, PL Adams, K Harlos, PA Nuttall, and DI Stuart. Tick histamine-binding proteins: isolation, cloning, and three-dimensional structure. *Molecular cell*, 3(5):661–671, 1999.

[133] Neil R Syme, Caitriona Dennis, Agnieszka Bronowska, Guido C Paesen, and Steve W Homans. Comparison of entropic contributions to binding in a hydrophilic versus hydrophobic ligand- protein interaction. *Journal of the American Chemical Society*, 132(25):8682–8689, 2010.

[134] Mark Bycroft, Svend Ludvigsen, Alan R Fersht, and Flemming M Poulsen. Determination of the three-dimensional solution structure of barnase using nuclear magnetic resonance spectroscopy. *Biochemistry*, 30(35):8697–8701, 1991.

[135] Michael J Lubienski, Mark Bycroft, Stefan MV Freund, and Alan R Fersht. Three-dimensional solution structure and 13c assignments of barstar using nuclear magnetic resonance spectroscopy. *Biochemistry*, 33(30):8866–8877, 1994.

[136] Gideon Schreiber and Alan R Fersht. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry*, 32(19):5145–5150, 1993.

[137] Mazen Ahmad, Wei Gu, Tihamér Geyer, and Volkhard Helms. Adhesive water networks facilitate binding of protein interfaces. *Nature communications*, 2:261, 2011.

[138] S Sangamnatdej, GC Paesen, M Slovak, and PA Nuttall. A high affinity serotonin- and histamine-binding lipocalin from tick saliva. *Insect molecular biology*, 11(1): 79–86, 2002.

[139] Ashley M Buckle and Alan R Fersht. Subsite Binding in an RNase: Structure of a Barnase-Tetranucleotide complex at 1.76 Å. resolution. *Biochemistry*, 33(7): 1644–1653, 1994.

[140] Nathaniel V Nucci, Maxim S Pometun, and A Joshua Wand. Site-resolved measurement of water-protein interactions by solution NMR. *Nature structural & molecular biology*, 18(2):245–249, 2011.

[141] Igor Dodevski, Nathaniel V Nucci, Kathleen G Valentine, Gurnimrat K Sidhu, Evan S O'Brien, Arthur Pardi, and A Joshua Wand. Optimized reverse micelle surfactant system for high-resolution NMR spectroscopy of encapsulated proteins and nucleic acids dissolved in low viscosity fluids. *Journal of the American Chemical Society*, 136(9):3465–3474, 2014.

[142] Anastasia Zhuravleva, Dmitry M Korzhnev, Svetlana B Nolde, Lewis E Kay, Alexander S Arseniev, Martin Billeter, and Vladislav Yu Orekhov. Propagation of dynamic changes in barnase upon binding of barstar: an nmr and computational study. *Journal of molecular biology*, 367(4):1079–1092, 2007.

[143] David NM Jones, Mark Bycroft, Michael J Lubienski, and Alan R Fersht. Identification of the barstar binding site of barnase by nmr spectroscopy and hydrogen-deuterium exchange. *FEBS letters*, 331(1):165–172, 1993.

[144] Bertil Halle. Protein hydration dynamics in solution: a critical survey. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1448):1207–1224, 2004.

[145] Kyle W Harpole and Kim A Sharp. Calculation of configurational entropy with a Boltzmann-Quasiharmonic model: the origin of high-affinity protein-ligand binding. *The Journal of Physical Chemistry B*, 115(30):9461–9472, 2011.

[146] Shoshana J Wodak and Joël Janin. Structural basis of macromolecular recognition. *Advances in protein chemistry*, 61:9–73, 2002.

[147] Julian M Sturtevant. Heat capacity and entropy changes in processes involving proteins. *Proceedings of the National Academy of Sciences*, 74(6):2236–2240, 1977.

[148] Izchak Z Steinberg and Harold A Scheraga. Entropy changes accompanying association reactions of proteins. *Journal of Biological Chemistry*, 238(1):172–181, 1963.

[149] Carl Diehl, Olof EngstroÌĹm, Tamara Delaine, Maria Håkansson, Samuel Genheden, Kristofer Modig, Hakon Leffler, Ulf Ryde, Ulf J Nilsson, and Mikael Akke. Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *Journal of the American Chemical Society*, 132(41):14577–14589, 2010.

[150] Chad M Petit, Jun Zhang, Paul J Sapienza, Ernesto J Fuentes, and Andrew L Lee. Hidden dynamic allostery in a PDZ domain. *Proceedings of the National Academy of Sciences*, 106(43):18249–18254, 2009.

[151] Jason R Schnell, H Jane Dyson, and Peter E Wright. Effect of cofactor binding and loop conformation on side chain methyl dynamics in dihydrofolate reductase. *Biochemistry*, 43(2):374–383, 2004.

[152] Patrick J Finerty Jr, Ranjith Muhandiram, and Julie D Forman-Kay. Side-chain dynamics of the SAP SH2 domain correlate with a binding hot spot and a region with conformational plasticity. *Journal of molecular biology*, 322(3):605–620, 2002.

[153] Stoyan Milev, Saša Bjelic, Oleg Georgiev, and Ilian Jelesarov. Energetics of peptide recognition by the second PDZ domain of human protein tyrosine phosphatase 1e. *Biochemistry*, 46(4):1064–1078, 2007.

[154] Veronica R Moorman, Kathleen G Valentine, Sabrina Bédard, Vignesh Kasinath, Jakob Dogan, Fiona M Love, and A Joshua Wand. Dynamic and thermodynamic response of the Ras protein Cdc42Hs upon association with the effector domain of PAK3. *Journal of Molecular Biology*, 2014.

[155] Simon J Teague. Implications of protein flexibility for drug discovery. *Nature reviews Drug discovery*, 2(7):527–541, 2003.