

Protein Stabilisation In Salt Solutions : Specific Ion Effects

A thesis submitted towards the partial fulfillment of
BS-MS Dual Degree Programme

By,
Atul C. Thakur
20131139

Under the guidance of
Dr. Arnab Mukherjee
Department of Chemistry
IISER Pune, India




Department of Chemistry
Indian Institute of Science Education and Research,
Pune-411 008
March, 2018

Certificate

This is to certify that this dissertation entitled “**Protein Stabilisation In Salt Solutions : Specific Ion Effects**” towards the partial fulfillment of the BS-MS dual degree program at the Indian Institute of Science Education and Research, Pune represents work carried out by **Atul C. Thakur** at Indian Institute of Science Education and Research, Pune under the supervision of **Dr. Arnab Mukherjee**, Associate Professor, Department of Chemistry during the academic year **2017-18**.

Date : 21-03-2018

Pune (MH), India




Arnab Mukherjee

(Thesis Supervisor)

Declaration

I hereby declare that the matter embodied in the report entitled "**Protein Stabilisation In Salt Solutions : Specific Ion Effects**" are the results of the work carried out by me at the **Department of Chemistry, Indian Institute of Science Education and Research, Pune** under the supervision of **Dr. Arnab Mukherjee** and the same has not been submitted elsewhere for any other degree or diploma from any other University or Institution. I declare that I have adequately cited and referenced the original sources and I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

Date : 21-03-2018
Pune (MH), India


Atul C Thakur
(20131139)

Acknowledgments

First and foremost, I would like to take this opportunity to express my sincere gratitude to my thesis advisor **Dr. Arnab Mukherjee** for giving me the opportunity to work under his supervision. He has always been kind and generous enough to overlook many mistakes that I made. Discussion with him really taught me the scientist's way of perceiving things. I take this opportunity to thank him for his constant encouragement, support and patience throughout the period 18 months which developed my personality more than the rest of the 18 years of my life. I am also grateful to my TAC member **Dr. Amrita Hazra** for valuable suggestion and discussion throughout the year.

I would like to express my deepest gratitude towards my mother **Mrs. Hira Thakur** and to my father **Mr. Chudaman Thakur** for their unbreakable faith, unconditional love the much needed criticism throughout the good, bad and worst patches of this journey. I am deeply indebted to my sister **Ms. Ashawini Thakur** and **Mrs. Harsha Thakur** for raising me into what I am today. I would like to express my special thanks to my special friend **Ms. Neha U. Saste** for always providing me with the shoulder to lean on to. Had it not been for support from all of you, I probably would not have achieved this milestone so easily.

I owe my deepest gratitude to my Lab members **Kanika Kohli** and **Abhijeet Gupta** for many memorable moments, support and healthy criticism. I would like to give my special thanks to **Kanika Kohli** for giving me the CPU cores whenever required on the no questions asked policy. I specially thank **Debasis Saha** for valuable tea time discussions which covered everything from science to food. I am thankful to the most full of thanks person and my senior **Hridya V. M.** for always guiding and answering my stupid question for a period of one complete year and for assisting in the revision of this thesis. Last but not the least, I express my deepest heartfelt appreciation to **Reman K singh**, who taught me the nitty-gritties of the computations. I express my deepest appreciation to him for working, discussing and solving a great deal of my problems. Without your hand on my back, I probably would not have completed the work for half of this thesis.

I would like to extend my thanks to **Dr. Kausik Chakraborty** and **CSIR Fourth Paradigm Institute** for providing me with the excellent computational facility to carry out my research work. I am grateful to **IISER Pune** for allowing me to pursue my master's in such vibrant environment and providing every platform to help me grow as a person and at the end, I thank **Almighty God** for giving me required perseverance and strength and surrounding me with all the beautiful people in my life.



Dedicated...

To My Beloved Mother & Father and To Their Efforts That
Made Me Into What I am Today.



Contents

1	Introduction	2
1.1	Outline of the thesis :	4
2	Methodology	8
2.1	Metadynamics :	8
2.2	Well-Tempered Metadynamics :	9
2.3	Definition of Collective Variables :	10
2.3.1	Native Contacts (N_c) :	10
2.3.2	Radius of Gyration (R_g) :	10
2.4	Calculation of Minimum Free Energy Path :	11
3	Ion Aggregation In Sulphate Solutions	13
3.1	Introduction	13
3.2	Design and Methodology :	14
3.3	Results & Discussions :	15
3.3.1	Ion association in sulphate solutions :	15
3.3.2	Inter-molecular interactions :	17
3.4	Conclusion :	18
4	Ubiquitin Unfolding Under Native Conditions	21
4.1	Introduction :	21
4.2	Design and Methodology :	22
4.2.1	Preparation of Initial Configuration :	22
4.3	Results and Discussions :	23
4.4	Minimum free energy path and barriers of unfolding :	26
4.5	Hydrogen Bonding and Residue Interaction energy :	28
4.5.1	Mechanism of Unfolding of Ubiquitin under native conditions :	30
4.6	Conclusion :	32
5	Ubiquitin Unfolding In Presence of Hofmeister Anions	36
5.1	Introduction	36
5.2	Design and Methodology	37
5.3	Results and Discussion	38
5.3.1	Direct Ion-Protein Interactions :	38
5.3.2	Free energy surface of Ubiquitin unfolding in sulphate solutions :	40
5.3.3	Hydrophilic/Hydrophobic-water Interactions :	41
5.3.4	Mechanism of Unfolding of Ubiquitin :	42
5.4	Outlook	44
5.5	Conclusion	45

List of Figures

1.1	Schematic showing the Hofmeister’s series for protein solubility and stability. Reproduced with permission from Tahara <i>et al.</i> ^[9] . Copyright [Pleaseinsertintopreamble] 2014 American Chemical Society	3
1.2	Figure showing the changes in the Hofmeister series with the pH-pI relations, Reproduced from Hattori <i>et al.</i> ^[30] with permissions. Copyright © 2016 RSC	4
2.1	Figure representing the model free energy surface obtained using hills method. Reproduced with permission from Ensing <i>et al.</i> ^[5] . Copyright © 2014 American Chemical Society	11
3.1	Schematic representing the observed aggregation of sulphate anions across all force fields.	15
3.2	Cumulative number radial distribution function for sulphate-sulphate ion pairs for all the force-fields used in the study	16
3.3	Cumulative number radial distribution function for sulphate-sulphate ion pairs for SIM7, SIM8	17
3.4	Figure representing the interaction energies between sulphate and sodium ions	18
4.1	FES for SIM1	24
4.2	FES for SIM2	24
4.3	FES for SIM3	25
4.4	FES for SIM4	26
4.5	FES for SIM5	26
4.6	FE along MFEP for SIM1, SIM2, SIM3	27
4.7	FE along MFEP for SIM4, SIM5	27
4.8	Hydrogen Bonding	28
4.9	Non-polar solvation energies	28
4.10	Mechanism of unfolding of secondary structural segments of Ubiquitin	31
5.1	Interaction Energies	38
5.2	Hydrogen Bonding	38
5.3	Minimum distance between "N" of protein residues and "O" of sulphate ions as a function of time	39
5.4	FES for SO_4^{2-}	40
5.5	MFEP for SO_4^{2-}	40
5.6	Hydrophilic SASA	41
5.7	Hydrophobic SASA	41
5.8	Figure representing unfolding pathways	43
5.9	FES for ClO_4^- ions	44
5.10	FES for SCN^- ions	44

List of Tables

3.1	Table Summarizing various force fields and the aggregation behaviour of sulphate ions	15
4.1	Table summarizing the force-fields used across this study	22
4.2	Table summarizing the simulation settings for metadynamics	23
4.3	Table summarizing the relative free energies, barriers and free energy differences for SIM1-SIM5	27
4.4	Table summarizing the labels for the secondary structural segments used in this study	30

Abstract

Significant progress has been made in investigating protein folding/unfolding reactions under various conditions in the literature. There is a prevailing belief that folded state of the protein is thermodynamically the most stable state and sits at the global minimum of the free energy surface. Protein unfolding reactions are chaotic in nature with no uniquely defined pathways leading to the unfolded state which itself is not uniquely defined either. Effect of salts on thermodynamics and dynamics of protein folding/unfolding has many times been poorly or insufficiently addressed in the literature compared to their counterparts such as Urea and GdmCl. There is a clear need to expand the horizons of our investigation beyond the usual approach of studying ion-macromolecule interaction in solution, in order to further progress and gain a deeper insight into the mechanism of Hofmeister phenomena. Therefore, in order to map the changes in free energy landscapes with respect to change in solution conditions we've performed MD-MTD simulations of Ubiquitin in different Hofmeister salts.

Our results provide strong evidence of the fact that sulphate ions shift the folded-unfolded equilibrium towards the folded side not by the stabilization of the native state but by destabilization of the unfolded state. Moreover, we also discuss the dependence of unfolding thermodynamics of ions on the water model and show that TIP4P-Ew and TIP4P-2005 increasingly destabilize the native state while TIP3P and TIP4P-D reproduce the folding-unfolding equilibrium quite realistically. Moreover, while studying the unfolding of Ubiquitin in sulphate solutions, we show that ion aggregation in non-polarizable models of sulphates arises because of the lack of optimized interactions between sulphate and sodium ions and thus might be the case across all polyvalent and densely charged ions.

Chapter 1

Introduction

The structure and function of the protein are perhaps the two most intimately related and interdependent quantities. The tertiary structure of the protein governs its interactions with its environment subsequently dictating its function. A diverse range of human proteopathies such as Alzheimer's, Parkinson's, Creutzfeldt–Jakob's and amyloidosis arise due to misfolding and aggregation of proteins^[1;2]. Various studies accounting the protein stability, often highlighted the decisiveness of osmolytes and co-solvents in determining protein stability, protein aggregation suppression and protein folding kinetics both *in vivo* and *in vitro*^[3;4]. A class of osmolytes, favors the native state of the protein by inducing a population shift on the folded side thus promoting native state stability whilst another favors the unfolded conformations thereby helping in protein denaturation, both of which are of fundamental interest and occur ubiquitously in nature. TMAO, a naturally occurring osmolyte is often present in deep sea inhabitants to counteract the effect of Urea, a well known protein denaturant and the external hydrostatic pressure^[3;5].

Salt ions, perhaps similar to osmolytes, are one of the most abundant species in various biological assemblies and has been proven to play a significant role in regulating protein solubility/stability centuries ago^[6;7]. Despite their ubiquity and a substantial contribution in steering protein solubility/stability, many earlier studies accounting protein denaturation and aggregation behavior repeatedly failed to weight the effect of salts on protein folding/unfolding dynamics, its solubility and stability until recently. Many emerging studies underline the effect of ions on the thermal stability of protein^[8] whilst other emphasize on the role of charge-charge interaction between protein and ions in rendering its solubility and stability^[9;10;11]. Nonetheless, the effect of ions in driving protein folding/unfolding equilibrium, their associated thermodynamics and kinetics and the underlying molecular mechanism of ion-protein interactions still remains elusive and is the topic of prime interest for this study.

Franz Hofmeister, in a seminal paper in 1880s, proposed the ion specificity series based on the sequential capabilities of ions to salt out proteins from solutions (Figure 1.1)^[6;7]. An analogues ranking of ions has been found in number of different physical properties viz. protein stability^[12;13], solubility^[10], surface tension at protein water interface^[14], colloidal assembly and chromatographic selectivity. Moreover, the effect of anions is found to be generally more prominent than cations^[8]. Within this framework, high charge density ions induces local water ordering and are therefore called '*Kosmotropes*' (water order-makers) whilst low charge density anions breaks the local water ordering and are therefore called '*Chaotropes*' (water order-breakers)^[15;16]. Kosmotropic anions and chaotropic cations 'salts-out' the protein from solution consequently enhancing their native state stability whereas, chaotropic cations and kosmotropic anions 'salts-in' protein thereby promoting its denaturation. Sodium and chloride ions are

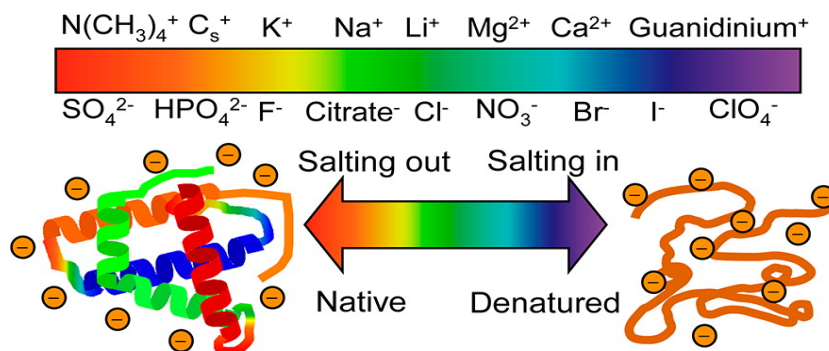


Figure 1.1: Schematic showing the Hofmeister's series for protein solubility and stability. Reproduced with permission from Tahara *et al.*^[9]. Copyright © 2014 American Chemical Society

usually acknowledged as the partitioning line, sectioning the kosmotropes from chaotropes.

Early literature, following the explanations of Franz Hofmeister, often rationalized the specific ion effects in terms of traditional water withdrawing effects of ions^[17]. According to this view, strongly hydrated ions i.e. *kosmotropes* increases the protein-protein interactions by stealing water molecules from its surface and thus being efficient in salting-out whereas weakly hydrated ions *chaotropes* lends water molecules to the protein thereby being efficient in salting-in. However, there are few caveats built into this justification. First, this rationalization of the series fails abruptly for cation since kosmotropic cations lies on the salting-in side of the series and tend to destabilize the native structure. Moreover this notion, does makes no reference, whatsoever, to the chemical structure of protein which is of vital importance as shown by Lysozyme following a reverse hofmeister series at low salt concentrations^[12]. Apart from this, there is mounting experimental and computational evidences, that ions do not affect the water ordering prominently, much beyond their solvation shells and water structure even in high salt solutions largely resembles the bulk water^[18;19]. Therefore, this concept of water withdrawing effects of salts on protein hydration is largely abandoned.

In recent years, numerous spectroscopic and computational studies has manifested the significance of direct ion-protein interactions in explaining the Hofmeister phenomena, which primarily emphasizes on the interactions of ions with protein backbone, the structural motif that is common to all proteins^[20;21]. Within this approach, based on the "Law of matching water affinities" chaotropic anions interact favorably with the chaotropic amide group on the peptide subsequently dictating it's stability whilst kosmotropic anions, do not interact much with their carboxyl counter part on the peptide backbone owing to the presence of substantial waters within their hydration shell^[22]. In this regard, the reverse trend for cations and anions and the greater impact of anions than cations is addressed satisfactorily. On the other hand, few groups reported that destabilizing effect of chaotropes comes largely due their interaction with the hydrophobic part of the protein than their interaction with the amide group on the protein backbone^[23;24].

Besides this direct interaction mechanisms, several other explanations for Hofmeister series has been recently proposed. A phenomenological theory developed by Der and Ramsden *et al.*, exemplifies the role of interfacial surface tension in governing protein conformational ensemble and it's stability whilst providing a keystone in addressing the molecular origin of Hofmeister phenomena^[25]. Pegram and Refcord *et al.*, based on the surface bulk partitioning model shows that the ions exposing protein surface to water (chaotropes) has higher tendency to accumulate at the air-water interface whilst ions removing water from the protein water interface are largely

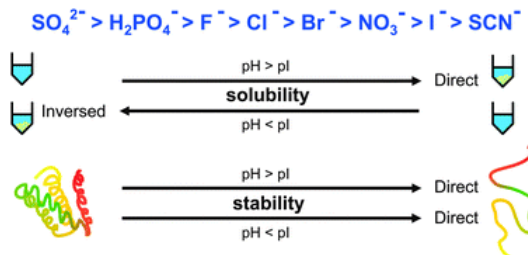


Figure 1.2: Figure showing the changes in the Hofmeister series with the pH-pI relations, Reproduced from Hattori *et al.*^[30] with permissions. Copyright © 2016 RSC

excluded from the air-water interface^[26]. Furthermore, many theories addressing Hofmeister effects in terms of ionic dispersion potentials^[27], excluded volume^[28], electrostatic preferential interaction^[29], interfacial water orientation and structure^[22] has been put forth. However, despite its simplicity and wide history, Hofmeister series still remains an unsolved enigma of the chemical world.

Despite their robustness, Hofmeister ordering of ions is shown to be reversed based on factors including surface charge and polarity, nature of ion-protein interaction, pH-pI relations and hydrophobic-hydrophilic surface area^[31]. Paterova *et al.*^[32], using the capped and uncapped tryglycine as model peptide explicitly showed the direct Hofmeister ordering for capped and reversed for the uncapped tryglycine. Hen egg white Lysozyme, has been famously proven to follow reversed Hofmeister series even under ($\text{pH} < \text{pI}$) acidic conditions and mild concentration ranges for ions, by Cremer and coworkers^[33]. Note that, as the figure 1.2 shows, with respect to changes in pH-pI relations, only the Hofmeister series for solubility is reversed whereas, the Hofmeister ranking for stability remains unchanged. Sedlak *et al.*^[8], under various pH-pI relations showed that the thermal stability of Apoflavodoxin (under $\text{pH} > \text{pI}$) and Cytochrome-C (under $\text{pH} < \text{pI}$) both follow the Hofmeister series.

The above discussion clearly illustrates that the folding/unfolding thermodynamics of proteins in salt solutions is poorly investigated in literature. In the present study, we therefore aim to elucidate the salt induced changes in the protein unfolding dynamics and thermodynamics. We used the all-atom molecular dynamics simulations coupled with metadynamics to calculate the free energy of unfolding of protein in salt solutions. We here explicitly calculate the free energy barrier involved in the unfolding of Ubiquitin for Hofmeister ions. Moreover, we also discuss the effect of water model on the native state stability of the Ubiquitin and also discuss the phenomena of sulphate ion aggregation often found in MD simulations. We choose, Ubiquitin, a 76 residue protein for studying the unfolding thermodynamics in salt solutions because unfolding of Ubiquitin molecule has been widely investigated widely in literature and moreover, it serves as excellent template for studying the protein folding/unfolding reactions within simulation due to its small size.

1.1 Outline of the thesis :

The thesis is sectioned into 4 different chapters which are as follows.

Methodology :

This chapter addresses the various computational methods and algorithms used in the study. We mention the Metadynamics protocol along with the caveats built into it. We also mention

the Well-Tempered Metadynamics method, developed to address the shortcomings of standard metadynamics. We present the recipe for computation of lowest free energy path, along the free energy surface obtained using WTMD method.

Ion aggregation in sulphate solution :

We here, discuss about various pre-existing sulphate ion potential while simultaneously highlighting their inadequacy in producing the correct experimental activity for aqueous sodium sulphate solutions. To quantify its origin, we also test whether this inadequacy arises out of the water models or sodium potentials by rigorously testing the sulphate ion potentials with various models of sodium and water. In the end, we show that the ion association in sulphate solutions arises out of the absence of specific calibration of sulphate-sodium interaction parameters and might be the case across all polyvalent and densely charged anions.

Ubiquitin unfolding under native conditions :

In this chapter, we mainly deal with the effect water models and protein force fields on the unfolding free energy landscape of Ubiquitin using tools discussed in Methods chapter. We show that the TIP4P-Ew and TIP4P-2005 water model has the tendency to cause the population shift towards unfolded states, using explicit all-atom molecular dynamics combined with hills method. We also highlight that TIP3P, despite being poor as a water model, along with dispersion corrected TIP4P-D water model, describes the protein folding/unfolding thermodynamics quite realistically whilst simultaneously highlighting the role of calibration of force fields against specific water models.

Ubiquitin unfolding in the presence of Hofmeister anions :

This chapter is dedicated to verify the effect of Hofmeister ions on the free energy surface of Ubiquitin unfolding. We here, discuss the free energy surface of Ubiquitin unfolding in sodium sulphate solutions. We show that, sulphate ions, though kosmotropic in nature is involved in the direct hydrogen bonding and persistent interactions with the positively charged residues. Moreover, we illustrate that the sulphate induced stabilization does not arise out of the stabilization of folded ensemble but arises out of the destabilization of unfolded ensembles of proteins.

Bibliography

- [1] Christopher M. Dobson. Protein folding and misfolding. *Nature*, 426 (6968):884–890, 2003.
- [2] Reman K. Singh, Neharika G. Chamachi, Suman Chakrabarty, and Arnab Mukherjee. Mechanism of unfolding of human prion protein. *J. Phys. Chem. B*, 121 (3):550–564, 2017.
- [3] Brian J. Bennion Qin Zou, Valerie Daggett, and Kenneth P. Murphy. The molecular mechanism of stabilization of proteins by tmao and its ability to counteract the effects of urea. *J. Am. Chem. Soc.*, 124 (7):1192–1202, 2002.
- [4] Timothy O. Street, D. Wayne Bolen, and George D Rose. A molecular mechanism for osmolyte-induced protein stability. *PNAS*, 103 (38):13997–14002, 2006.
- [5] Samuel S. Cho, Govardhan Reddy, John E. Straub, and D. Thirumalai. Entropic stabilization of proteins by TMAO. *J. Phys. Chem. B*, 115 (45):13401–13407, 2011.
- [6] F Hofmeister. Zur lehre von der wirkung der salze. *Naunyn-Schmiedeberg’s Arch. Pharmacol*, 24:247–260, 1888.
- [7] W Kunz, J Henle, and B. Ninham. Zur lehre von der wirkung der salze (about the science of the effect of salts): Franz hofmeister’s historical papers. *Curr. Opin. Colloid Interface Sci*, 9:19–37, 2004.
- [8] E Sedláková, Stagg, L, and P Wittung-Stafshede. Effect of hofmeister ions on protein thermal stability: Roles of ion hydration and peptide groups? *Archives of Biochemistry and Biophysics*, 479 (1):69–73, 2008.
- [9] Nihonyanagi Satoshi, Yamaguchi Shoichi, and Tahara Tahei. Counterion effect on interfacial water at charged interfaces and its relevance to the hofmeister series. *J. Am. Chem. Soc.*, 136:6155–6158, 2014.
- [10] Eva Y. Chi, KrishnanTheodore Sampathkumar, RandolphJohn W., and Carpenter F. Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharmaceutical Research*, 20 (9):1325–1336, 2003.
- [11] 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.
- [12] and Jana Hladilkova Halil I. Okur, Kelvin B. Rembert, Younhee Cho, Jan Heyda, Joachim Dzubiella, Paul S. Cremer, and Pavel Jungwirth. Beyond the hofmeister series: Ion-specific effects on proteins and their biological functions. *J. Phys. Chem. B*, 121:1997–2014, 2017.
- [13] Baldwin RL. How hofmeister ion interactions affect protein stability. *Biophys J*, 71:2056–2063, 1996.
- [14] Soohaeng Yoo Willow and Sotiris S. Xantheas. Molecular-level insight of the effect of hofmeister anions on the interfacial surface tension of a model protein. *J. Phys. Chem. Lett*, 8:1574–1577, 2017.
- [15] W. M. Cox and J. H. Wolfenden. Viscosity of strong electrolytes measured by a differential method. *Proc. R. Soc. London*, 145:475–488, 1934.

- [16] R. W Gurney. *Ionic Processes in Solution*. McGraw-Hill, New York, 1953.
- [17] M Boström, DRM Williams, and BW Ninham. Why the properties of proteins in salt solutions follow a hofmeister series. *Curr Opin Colloid Interface Sci*, 9:48–52, 2004.
- [18] A. W. Omta, M. F. Kropman, S. Woutersen, and H. J. Bakker. Negligible effect of ions on the hydrogen-bond structure in liquid water. *Science*, 301:347–349, 2003.
- [19] G. Stirnemann, E. Wernersson, P. Jungwirth, and D. Laage. Mechanisms of acceleration and retardation of water dynamics by ions. *J. Am. Chem. Soc.*, 135:11824–11831, 2013.
- [20] Jana Paterova, Kelvin B. Rembert, Jan Heyda, Yadagiri Kurra, Halil I. Okur, Wenshe R. Liu, Christian Hilty, Paul S. Cremer, , and Pavel Jungwirth. Reversal of the hofmeister series: Specific ion effects on peptides. *J. Phys. Chem. B*, 117:8150–8158, 2013.
- [21] Yanjie Zhang and Paul S Cremer. Interactions between macromolecules and ions: the hofmeister series. *Current Opinion in Chemical Biology*, 10:658–663, 2006.
- [22] Collins KD. Ions from the hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods*, 34:300–311, 2004.
- [23] J Heyda, JC Vincent, DJ Tobias, J Dzubiella, and Jungwirth P. Ion specificity at the peptide bond: molecular dynamics simulations of n-methylacetamide in aqueous salt solutions. *J Phys Chem B*, 114:1213–1220, 2010.
- [24] Gibb CLD and Gibb BC. Anion binding to hydrophobic concavity is central to the salting-in effects of hofmeister chaotropes. *J Am Chem Soc*, 133:7344–7347, 2011.
- [25] Der A, Kelemen L, Fabian L, Taneva SG, Fodor E, and Pali T et. al. Interfacial water structure controls protein conformation. *J Phys Chem B*, 111:5344–5350, 2007.
- [26] Pegram LM and Record MT Jr. Hofmeister salt effects on surface tension arise from partitioning of anions and cations between bulk water and the air–water interface. *J Phys Chem B*, 111:5411–5417, 2007.
- [27] Kunz W, Belloni L, Bernard O, and Ninham BW. Osmotic coefficients and surface tensions of aqueous electrolyte solutions: role of dispersion forces. *J Phys Chem B*, 108:2398–2404, 2004.
- [28] Schellman JA. Protein stability in mixed solvents: a balance of contact interaction and excluded volume. *Biophys J*, 85:108–125, 2003.
- [29] Honig B and Nicholls A. Classical electrostatics in biology and chemistry. *Science*, 268:1144–1149, 1995.
- [30] Katsuyoshi Aoki, Kentaro Shirakib, and Toshiaki Hattori. Salt effects on the picosecond dynamics of lysozyme hydration water investigated by terahertz time-domain spectroscopy and an insight into the hofmeister series for protein stability and solubility. *Phys. Chem. Chem. Phys.*, 18:15060–15069, 2016.
- [31] Schwierz N, Horinek D, and Netz RR. Reversed anionic hofmeister series: the interplay of surface charge and surface polarity. *Langmuir*, 26:7370–7379, 2010.
- [32] Paterová J, Rembert KB, Heyda J, Kurra Y, Okur HI, and Liu WR et al. Reversal of the hofmeister series: specific ion effects on peptides. *J Phys Chem B*, 117:8150–8158, 2013.
- [33] Yanjie Zhang and Paul S Cremer. The inverse and direct hofmeister series for lysozyme. *Proceedings of the National Academy of Sciences*, 106:15249–15253, 2009.

Chapter 2

Methodology

Protein folding and unfolding events are very stochastic processes occurring in a regime of several hundreds of nanoseconds (ns) to microsecond (μ s), in a convoluted multidimensional free energy landscape. Computing the underlying free energy surface associated with such processes, using the brute-force technique of all-atom molecular dynamics simulations, is extremely computationally demanding and often not possible owing to the high energy barriers involved and other sampling bottlenecks^[1]. Many adaptive biasing techniques has been introduced in the literature, in order to efficiently cross the free energy barriers and to overcome the sampling bottlenecks on a computationally accessible timescale^[2;3]. We here use the recently introduced Well-Tempered Metadynamics method to construct the free energy surface associated with protein unfolding/folding events as a function of it's slowest relaxing reaction coordinates^[2].

In this chapter, we mainly discuss about Metadynamics and its bottlenecks^[4], Well-Tempered Metadynamics^[2], reaction coordinates used and recipe for the computation of minimum free energy path along the calculated free energy surface^[5].

2.1 Metadynamics :

Metadynamics^[4], an extended version of Wang-Landau approach at finite temperature^[6], is a robust and non-equilibrium accelerated sampling technique, used particularly with an interest to construct the underlying equilibrium free energy surface as a function of few collective variables (CV) of any biological/chemical processes. Metadynamics^[4] is designed in order to force the system to explore the configuration that are higher in free energy by biasing the dynamics by adding history dependent Gaussian potentials along the reaction coordinates until the free energy surface becomes flat. Free energy surface, which will be a function of reaction coordinates (CV), can then be constructed as a negative of the bias potential.

Despite the success of metadynamics^[4], there are few shortcomings of the technique. In normal metadynamics^[4], free energy does not converge but keeps fluctuating around the exact result leading to an average error which is directly proportional to the square root of rate of Gaussian deposition, reducing which guarantees the increase in the time required to explore the free energy surface. Furthermore, in the long time limit, there is always a risk of irreversibly pushing the system into the physically irrelevant regions of configurational space. Various solution had been proposed to address this issues with metadynamics^[4], Well-Tempered metadynamics^[2] being one of them.

2.2 Well-Tempered Metadynamics :

Well-Tempered Metadynamics^[2], inspired from the self-healing umbrella sampling method^[7], is comparatively new but robust method built in order to address the convergence problem of metadynamics. In Well-Tempered metadynamics, the dynamics of the system are biased by adding the history-dependent repulsive Gaussian potentials in the CV space according to the equation,

$$V(s, t) = \Delta T \ln \left(1 + \frac{wN(s, t)}{\Delta T} \right) \quad (2.1)$$

where $s(q)$ represents the set of reaction coordinates (CV) along which we want to bias the dynamics, as a functions of system’s microscopic coordinates q and t represents the time. w is the energy rate , ΔT is a temperature parameter and $N(s, t)$ can be obtained from the biased dynamics.

Well-Tempered Metadynamics^[2], unlike the standard metadynamics^[4], addresses the convergence problem in metadynamics by constantly re-scaling the Gaussian height and thus the deposition rate according to the equation,

$$V(s, t) = w \exp - \left(\frac{V(s, t)}{\Delta T} \right) \delta_{s, s(t)} \quad (2.2)$$

where w is the initial bias deposition rate. Free energy surface, in practice, is constructed as

$$\tilde{F}(s, t) = -\frac{T + \Delta T}{\Delta T} V(s, t) \quad (2.3)$$

where $\frac{T + \Delta T}{\Delta T}$ is known as the bias factor and $\tilde{F}(s, t)$ represents the free energy surface as a function of collective variables (CV).

In practice, we start from the reactant state without adding any Gaussian potentials along the CV space. The fluctuations of the $s(q)$ in the CV space indicates the width of the Gaussian (σ) to be used, while representing the width of the local reactant state minimum in which the system is stuck. The chosen width of the Gaussian (σ), for a chosen $s(q)$, represents the efficiency in terms of exploring the free energy surface. The history dependent potential is then added which forces the system to disfavor the previously visited configurations, thereby sampling new configurations until the system escapes from the reactant state minimum to product state minimum by traversing through the lowest energy transition state whilst compensating for the underlying free energy. When all the local minima are filled and system can move freely from reactant to product state i.e when the metadynamics run has converged, the free energy surface can then be constructed as negative of the added bias.

Well-Tempered Metadynamics^[2] has much to offer computationally over a standard metadynamics. By optimizing the ΔT , one could increase the barrier crossing events or limit the exploration of FES to physically relevant regions in the energy range ($T + \Delta T$) of $s(q)$. Better statistical accuracy can be obtained in the long time limit whilst simultaneously avoiding the overfilling. By checking the convergence, it is rather easy to determine when to stop the run.

2.3 Definition of Collective Variables :

Owing to the stochastic nature of protein folding/unfolding reactions, it is imperative to make a good choice of a finite number of collective variables i.e reaction coordinates, on the grounds of heuristic understanding of the system. In our case, a good reaction coordinate should clearly differentiate between unfolded and folded states of the protein while simultaneously being the slowest relaxing mode in the reaction. After careful investigation of several CV's, we've used the following CV's to monitor the unfolding of Ubiquitin.

2.3.1 Native Contacts (N_c) :

Most protein folding/unfolding studies commonly use native contacts as the reaction coordinate to monitor the folding/unfolding reactions because of it's ability to clearly distinguish between folded and unfolded regions^[8]. Native contacts, as the name suggests, is defined here as the spatial proximity between secondary structural units in the native state of the protein. If one heavy atom from a secondary structure (g_1) falls within the 4.5 Å of another secondary structure unit (g_2) in the native state, then the contact is said to be formed. To make s_{ij} , continuous and differentiable at all points, it is defined as,

$$N_c = \sum_{i \in g_1} \sum_{j \in g_2} s_{ij}$$

where s_{ij} is given as,

$$s_{ij} = \begin{cases} 1 & r_{ij} \leq 0 \\ 1 - \left(\frac{r_{ij}}{r_0}\right)^n & 0 < r_{ij} < r_0 \\ 1 - \left(\frac{r_{ij}}{r_0}\right)^m & r_{ij} \geq r_0 \end{cases}$$

and r_{ij} is $|r_j - r_i| - d_0$, and other parameters were chosen as $n = 6, m = 12, r_0 = 4.5 \text{ \AA}, d_0 = 0 \text{ \AA}$.

On the account that native contacts are defined over the native state of the protein, unfolded proteins can usually be characterized by the lesser number of native contacts.

2.3.2 Radius of Gyration (R_g) :

Radius of gyration, being similar to the Hydrodynamics Radius, is the measure of distribution of components (in our case atoms) from the center of mass of the system (protein). It indicates the compactness of the protein and is defined as,

$$R_g = \sqrt{\sum_i \frac{m_i (\vec{r}_i - \vec{r}_c)^2}{M}} \quad (2.4)$$

where, m_i is the mass and \vec{r}_i is the position vector of the i^{th} atom, M is the total mass and \vec{r}_c is the position vector of the center of mass of the protein. Higher the value of R_g lesser is the protein's compactness and vice versa. Typically folded proteins have lesser R_g than there unfolded counterparts and intrinsically disordered proteins.

2.4 Calculation of Minimum Free Energy Path :

For most chemical reactions lowest free energy path depicts the most probable mechanism for a reaction. It is, therefore of fundamental interest to calculate the lowest free energy path representing the reaction. The method, developed by Ensing *et al.*^[5] is built over the Well-Tempered Metadynamics^[2] and calculates the minimum free energy path along the free energy surface, obtained with the metadynamics.

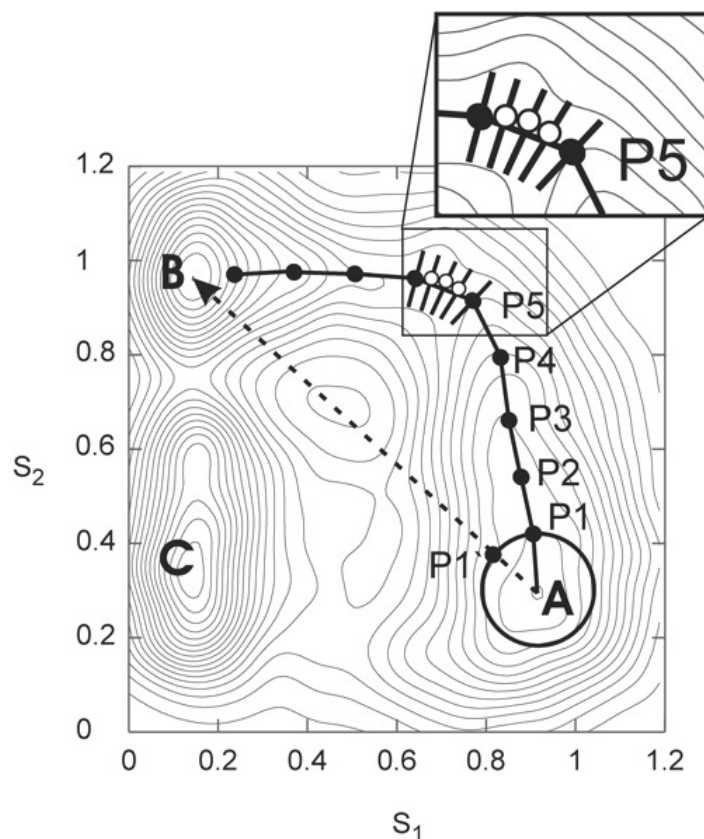


Figure 2.1: Figure representing the model free energy surface obtained using hills method. Reproduced with permission from Ensing *et al.*^[5]. Copyright © 2014 American Chemical Society

Figure 2.1 shows the example of 2D free energy surface, as a function of collective variables S_1 and S_2 , that we usually get after the metadynamics run. In the figure, A denotes the reactant well and B and C both represents the product well. As can be seen from the free energy surface, there can be many possible path joining the reactant and product wells. Locating the minimum free energy path along the free energy surface starts by identifying the product and reactant minima in the free energy surface. This can be readily obtained as collective variables spends most of its time fluctuating around these values.

The minimization of $V(s)$ is performed by choosing brackets around $V(s)$ as $[s_\alpha - \delta s_\alpha, s_\alpha + \delta s_\alpha]$ such that $V(s) < V(s_\alpha - \delta s_\alpha)$ and $V(s) < V(s_\alpha + \delta s_\alpha)$. The new random point, s'_α is chosen in the interval s_α and the bracket ends $[s_\alpha - \delta s_\alpha, s_\alpha + \delta s_\alpha]$ which has the highest potential. If $V(s_\alpha) < V(s'_\alpha)$ then s'_α replaces the bracket end with the lower potential otherwise s_α replaces the bracket end having the higher potential while s'_α becoming the new minimum. This process is repeated until the difference between $V(s_\alpha) - V(s'_\alpha)$ is below some tolerance value, for all collective variables α .

Initially, a coarse path is traced joining the wells A and B by taking longer steps in the direction of the vector \overrightarrow{AB} , using the bracketing method described above. The coarse path (shown as circles in figure 2.1) is then optimized, by optimizing the points in direction perpendicular to the path. The resulting path is the lowest free energy path joining the two wells along the free energy surface for the given choice of collective variables.

Bibliography

- [1] Christopher M. Dobson. Protein folding and misfolding. *Nature*, 426 (6968):884–890, 2003.
- [2] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, 2008.
- [3] M. Souaille and B. t Roux. Extension to the weighted histogram analysis method: Combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.*, 135:40–57, 2001.
- [4] Giovanni Bussi, Alessandro Laio, and Michele Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *PRL*, 96:090601, 2006.
- [5] B. Ensing, A. Laio, M. Parrinello, and M. L. A Klein. Recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B*, 109:6676–6687, 2005.
- [6] Fugao Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 2001:2050, 86.
- [7] Simone Marsili, Alessandro Barducci, Riccardo Chelli, Piero Procacci, and Vincenzo Schettino. Self-healing umbrella sampling: a non-equilibrium approach for quantitative free energy calculations. *The Journal of Physical Chemistry B*, 110:14011–14013, 2006.
- [8] Robert B Best, Gerhard Hummer, and William A Eaton. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences*, 110:17874–17879, 2013.

Chapter 3

Ion Aggregation In Sulphate Solutions

3.1 Introduction

Sulphate salts are of paramount significance in vast number of chemical processes ranging from biochemical charge transfer reactions to industrial process such as detergent manufacturing^[1]. Sulphates', being kosmotropic in nature posses a large hydration shell and are shown to be extremely effective in salting out and stabilization of proteins^[2]. In environmental context, sulphate has been shown to effectively increase the earth's albedo by about 0.4 W/ m^2 partially aiding in global warming by offsetting the effect of other green-house gases therefore making sulphate salts of particular interest for researchers^[3]. Molecular dynamics simulation present and fast and scalable approach in characterizing the local water structure making/breaking properties of such anions simultaneously aiding in classifying the origin of ion specific effects. For this reasons, an accurate representation of sulphate solution in simulation using classical force fields in necessary to correctly reproduce interaction between sulphate and it's environment.

Given it's importance, there've been many attempts since last decade to develop the forcefield parameters for sulphate dianion^[4;5]. However, there seems to be few discrepancies associated with the preexisting non-polarizable potentials. Several studies have shown that simulation of sulphate salts using several of the available non-polarizable force fields leads to extensive ion clustering of the sulphate anions even at the concentration of 0.5 M, whereas simulations using polarizable model of sulphate correctly reproduces many thermodynamical quantities such as hydration free energy, chemical potentials and solution activity data^[6]. Werensson and Jungwrith group^[6], studied this sulphate aggregation and attributed the extensive degree of ion pairing to the in-capabilities observed in the non-polarizable force fields to model the polarization of solvation shell waters induced by the sulphate anion. Nonetheless, the extensive ion pairing observed at 0.5 M concentration is clearly not the observed behaviour in the aqueous sulphate solutions.

To correctly reproduce the experimental solution activity and chemical potential data, optimum balance of ion pairing interactions is of crucial significance perticulary for solutions of strong and multivalent electrolytes. In this chapter, we show that the observed ion aggregation in sulphate solutions for non-polarizable force fields is independent of the water and sodium model used. Moreover, we state and show that reason for such extensive ion association is that, most sulphate ion force fields have been parametrized intensively with a clear focus on reproducing the experimental solvation energy while completely overlooking need for the optimization

cation-ion interaction parameters which leads to observed aggregation in case of divalent anions. Furthermore, we exemplify on the inadequacy of Lorentz-Bertholet^[7;8] combination rules in representing the interactions for multivalent ions and emphasize on the significance of specific parametrization strategy of cation-anion interaction parameters, to prevent aggregation.

3.2 Design and Methodology :

We've performed 8 simulations of sulphate solution using various force fields summarized in table 3.1. Each simulation consisted of 19 sulphate and 38 sodium ions using the force fields as described in the table 3.1. Each system was then solvated using the associated water models to yield the concentration of 0.3 M. System was then energy minimized using the steepest-descent^[9] algorithm implemented in GROMACS^[10] followed by heating upto 300 K using Berendsen thermostat^[11] under NVT conditions with a coupling constant of 0.4 ps. The system was then equilibrated for 1 ns under NPT conditions using Berendsen thermostat^[11] (300K) and Parrinello-Rahman barostat^[12] (1bar), both coupled with a coupling constant of 0.4 ps. Equilibration was immediately followed by 10 ns of production dynamics under NPT conditions using Nose-Hoover chains^[13] (300K) and Parrinello-Rahman barostat^[12], each with a coupling constant of 0.2 ps. All the Simulations were performed using GROMACS MD^[10] package, using leap-frog integrator with a time step of 2 fs. The Particle Mesh Ewald^[14] treatment for long range electrostatics and Van Der Walls scheme for non-bonded interactions were used, each with a cutoff of 1 Å. Intermolecular Van Der Walls interactions, unless otherwise stated, were modelled using the standard Lorentz-Bertholet^[7;8] combination rules. All bond lengths and bond angles were constrained using LINCS^[15] algorithm.

To understand the mechanism of sulphate aggregation we have used several forcefields available in the literature. Sulphate-Water interactions, if modelled insignificantly, could result in altered dynamics of ions, which leads to aggregation. Therefore, in order to discern the role of water model in sulphate ion association, we've performed simulations using TIP3P^[16], TIP4P-Ew^[17] as well as SPC/E^[18] water model using sulphate potentials published by Cannon *et al.*^[4] and using the standard sodium ion model^[19] in the AMBER force fields. Furthermore, apart from the sulphate-water interaction, sulphate-sodium interactions are also of vital significance in correctly reproducing the experimental activity for sodium sulphate solutions. Therefore, in order to judge the effect of sodium ion potentials on sulphate aggregation, we've carried out simulation using sodium ion parameters developed by Joung and Chetham^[20] in SPC/E^[18] water. The model developed by Joung and Chetham^[20] is well tested and is shown to perform excellently in reproducing many different properties of alkali solutions. Moreover, to remove the bias of sulphate ion force field from our discussion and perhaps to address the issue more generally, we've also carried out simulations using various different sulphate potentials available in literature and are best summarized in the following table.

Note that, SIM7 and SIM6 differ only in the representation of their inter-molecular potentials between cation and anion. SIM7 was carried out using the optimised sulphate-sodium interaction parameters, as proposed by Verde *et al.*^[22], whilst SIM8 was performed without the optimized parameters and the inter-molecular potentials were calculated within the simulation using the Lorentz-Bertholet^[7;8] combinations rules, as for SIM1 to SIM6.

Simulation no	Force field details			Aggregation
	Sulphate-model	Water model	Sodium model	
SIM1	Cannon <i>et al.</i> ^[4]	SPC/E ^[18]	Amber ^[19]	Yes
SIM2	Cannon <i>et al.</i> ^[4]	TIP3P ^[16]	Amber ^[19]	Yes
SIM3	Cannon <i>et al.</i> ^[4]	TIP4P-Ew ^[17]	Amber ^[19]	Yes
SIM4	Cannon <i>et al.</i> ^[4]	SPC/E ^[18]	JC ^[20]	Yes
SIM5	Reimer <i>et al.</i> ^[21]	SPC/E ^[18]	Reimer <i>et. al.</i> ^[21]	Yes
SIM6	Williams <i>et al.</i> ^[4]	SPC/E ^[18]	Amber ^[19]	Yes
SIM7	Verde <i>et al.</i> ^[22]	TIP3P ^[16]	JC ^[20]	No
SIM8	Verde* <i>et al.</i> ^[22]	TIP3P ^[16]	JC ^[20]	Yes

Table 3.1: Table Summarizing various force fields and the aggregation behaviour of sulphate ions

3.3 Results & Discussions :

We here discuss the unrealistically large degree of ion association observed in aqueous sulphate solutions while highlighting the importance of specific optimization of sulphate-sodium interaction parameters and discuss the role of polarizability. All the runs, with the exception of model developed by the Verde *et al.*^[22] resulted in a persistent clustering of sulphate ions. Moreover the clustering, once formed (≈ 5 ns) was persistent even after 100 ns of simulation time. The figure 3.1 represents the snapshot of the observed clustering in sulphate solutions.

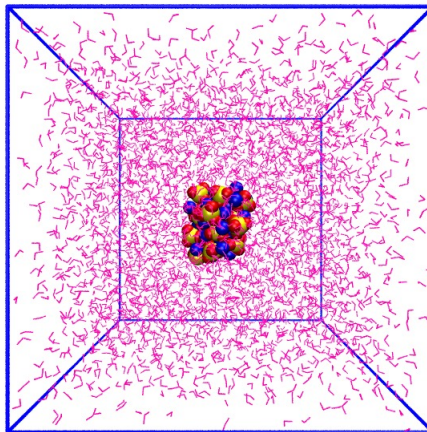


Figure 3.1: Schematic representing the observed aggregation of sulphate anions across all force fields.

3.3.1 Ion association in sulphate solutions :

We calculated the cumulative number radial distribution function of sulphate-sulphate ions as a function of simulation time to quantify the aggregation of sulphate ions. The cumulative

number radial distribution represents the variation in the number of ions around the reference atom/molecule as a function of the radial distance. The following figure represent the calculated distribution function for all the simulations.

As can be seen from the figure 3.2, all the runs with the notable exception SIM7 shows the higher sulphate-sulphate coordination number within their first hydration shells ($\approx 5 \text{ \AA}$ -Appendix A) which suggests that sulphate ions are indeed aggregating. Note that despite of using the several different water models in simulations from SIM1 to SIM3 there's is still prevalent degree of ion association present in the sulphate solutions, which shows that the aggregation is not particular to a water model. Identically, SIM1 and SIM4, which differ only in their representation of sodium ion potentials, shows the similar sulphate-sulphate coordination number within the first hydration shell which emphasizes that sulphate ion association is not arising due to the use of particular sodium ion potentials.

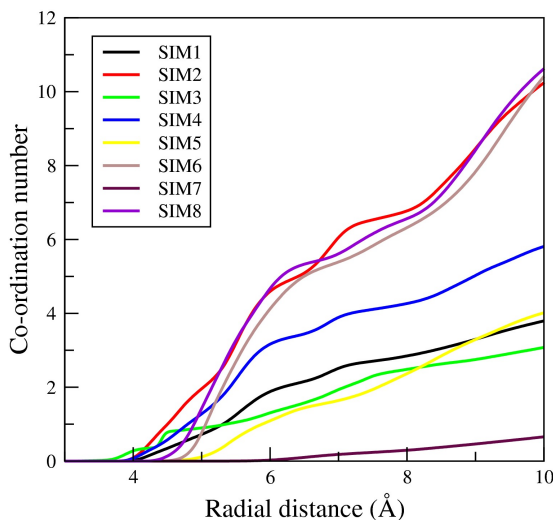


Figure 3.2: Cumulative number radial distribution function for sulphate-sulphate ion pairs for all the force-fields used in the study

As stated earlier, to remove the bias introduced by molecular mechanics force field, we have done simulation using a set of different potential parameters for sulphate ions published in the literature. Interestingly, Reimer *et al.*^[21] and Williams *et al.*^[4], in their original paper mentioned the robustness of their sulphate potentials and showed explicitly that their sulphate potentials do not form nanometer scale aggregates. However, by contrast, our results from SIM5 and SIM6 shows that there is a presence of quite substantial amount of homo and hetero-ion pairing present. Verde *et al.*^[22], recently published the potentials for several oxoanions, sulphate being one of them. As the figure-3.2 shows, SIM7 has the lowest sulphate-sulphate coordination number, which indicates that unlike for SIM1-SIM6, sulphate ions in SIM7 are not involved in any cluster formation even after ≈ 200 ns of simulation time.

The sulphate potential parameters developed by Verde and co-workers^[22], were extensively parametrized against the sodium ion model of Joung and Chetham^[20] for TIP3P^[16] water model, focused primarily on the experimental hydration free energies and activity coefficients. By contrast, other previously published sulphate-potentials were majorly focused only on reproducing the hydration free energy for sulphate anions for a particular water model whilst completely ignoring the significance of anion-cation interactions in reproducing the experimental activity data.

Note that, in all of the simulations which led to aggregation (SIM1-SIM6 & SIM8), optimised sulphate-sodium interactions were not present and instead were modelled using the standard Lorentz-Bertholet [7;8] combination rules. We thus think that, it is this specific optimization of inter-molecular interaction parameters that inhibits the sulphate anions among aggregating in themselves in the model of Verde *et al.* [22] and might be the necessary step in calibration of other polyvalent ion potentials e.g. CO_3^{2-} .

3.3.2 Inter-molecular interactions :

To illustrate our point, SIM8 was carried out using the same simulation set-up only differing in representation of inter-molecular anion-cation potentials, as discussed in previous section. Following figure represents the cumulative number radial distribution function for the SIM8, in comparison with SIM7.

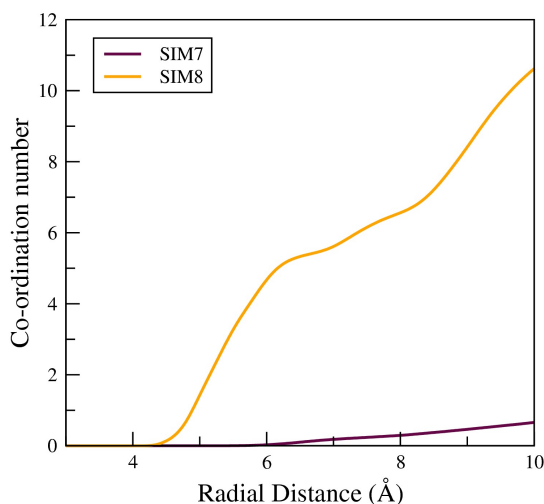


Figure 3.3: Cumulative number radial distribution function for sulphate-sulphate ion pairs for SIM7, SIM8

As can be seen from the Figure-3.3, removal of optimised interactions between sulphate and sodium leads to the extensive homo and hereto ion clustering. Particularly high values of the coordination number in SIM8, in comparison to SIM7, as shown in the Figure-3.3, proves that the standard set of Lorentz-Bertholet [7;8] combination rules when used along with non-polarizable forcefield models, not only represents the sulphate-sodium interactions inadequately but often overestimates it causing the unrealistically large degree of ion association. Moreover, this inadequacy gets particularly pronounced in the simulation of polyvalent and densely charged ions, when modelled using standard non-polarizable models, highlighting the need of the specific optimization of anion-cation interaction parameters for solutions containing strong electrolytes.

To elaborate more on the magnitude of the overestimation of interactions between anion-cation in standard non-polarizable force fields, we've calculated the interaction energies between sulphate and sodium on the initial 1 ns of simulation data. This was necessary in order to avoid the spurious drifts in the energy values after the process of aggregation. Figure below represents the interaction energies between sulphate and sodium for the range of force fields used.

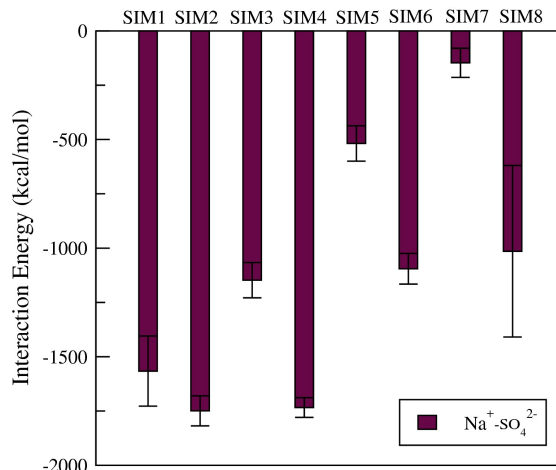


Figure 3.4: Figure representing the interaction energies between sulphate and sodium ions

As can be seen from the figure, the interaction energy between sulphate and the sodium for SIM1-SIM6 and SIM8 is nearly order of magnitude greater than in SIM7, which backs our observations suggesting that the overestimation of interactions between cation and anion might be the underlying cause of observed clustering in sulphate solutions. We think that, this effect might be general and present across all the high charge density ions when modelled without optimised cation-anion interactions. Note that, some amount of transient clustering was still seen across SIM1-SIM6 and further sophisticated techniques viz. calculation of partial dehydration energy, binding energy for ion pairs might be needed in order to fully understand the mechanism of aggregation and is beyond the scope of this investigation. Since, the model proposed by Verde *et al.* [22] (SIM7) describes the sodium sulphate solution most accurately amongst all the pre-existing non-polarizable forcefields, we've thus used it to study the unfolding dynamics and thermodynamics of Ubiquitin in sulphate solution and is discussed in subsequent chapters.

3.4 Conclusion :

Above discussion, clearly indicates the fact that, sulphate aggregation phenomena is not associated with any particular water or sodium model and is present across all different water and sodium models. Moreover, we show that pre-existing forcefields for sulphate anion, available in the literature fail to correctly reproduce the interaction between sulphate and sodium ions while simultaneously overestimating it. By using the same setup as SIM7 in SIM8 and removing only the optimised interaction potentials between sulphate and sodium ions, we show that this aggregation arises from the lack of specific optimisation of cation-anion interaction parameters and emphasizes on the inadequacy of Lorentz-Bertholet [7;8] combination rules in describing the cation-anion interactions specially for multivalent and high charge density ions. From our results, it is safe to conclude that the similar ion aggregation observed for CO₃²⁻ ions might also be the result of incorrect description of interactions within cation and anion thus highlighting the need for cation-anion optimization in the next generation force fields.

Bibliography

- [1] F. Albert Cotton and Geoffrey Wilkinson. *Advanced Inorganic Chemistry*. Wiley, New York, NY, 2nd ed. edition, 1966.
- [2] Collins KD. Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods*, 34:300–311, 2004.
- [3] S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller. *Intergovernmental Panel on Climate Change Chapter 2: Changes in Atmospheric Constituents and Radiative Forcing*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
- [4] William R. Cannon, B. Montgomery Pettitt, and J. Andrew McCammon. Sulfate anion in water: Model structural, thermodynamic, and dynamic properties. *J. Phys. Chem.*, 98:62256230, 1994.
- [5] Christopher D. Williams and Paola Carbone. A classical force field for tetrahedral oxyanions developed using hydration properties: The examples of perchlorate and sulfate. *The Journal of Chemical Physics*, 143:174502, 2015.
- [6] Erik Wernersson and Pavel Jungwirth. Effect of water polarizability on the properties of solutions of polyvalent ions: Simulations of aqueous sodium sulfate with different force fields. *J. Chem. Theory Comput.*, 6:3233–3240, 2010.
- [7] H. A. Lorentz. Ueber die anwendung des satzes vom virial in der kinetischen theorie der gase. *Annalen der Physik*, 248 (1):127–136, 1881.
- [8] Daniel Berthelot. Sur le melange des gaz, comptes rendus hebdomadaires des seances de. *Academie des Sciences*, 126:1703–1855, 1898.
- [9] W. H. Press. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge university press, 2007.
- [10] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, J. C. Shirts, M. R. and Smith, P. M. Kasson, and al. vander Spoel, D. et. Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29:845–54, 2013.
- [11] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [12] M. Parrinello and A Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys*, 52:7182–7190, 1981.
- [13] S. I. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 100:191–198, 2002.
- [14] T. Darden, D. York, and L. J. Pedersen. Particle mesh ewald: An $n \cdot \log(n)$ method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.
- [15] Berk Hess, Henk Bekker, Herman J.C.Berendsen, and Johannes G.E.M.Fraaijey. Lincs:a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18:1463–1472, 1997.

- [16] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.
- [17] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: Tip4p-ew. *J. Chem. Phys.*, 120:9665, 2004.
- [18] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.
- [19] J. J. Aqvist. Ion-water interaction potentials derived from free energy perturbation simulations. *Phys. Chem.*, 94:8021, 1990.
- [20] In Suk Joung and Thomas E. Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112:9020–9041, 2008.
- [21] Joachim Reimer, Matthew Steele-MacInnis, Jörg M. Wambach, and Frédéric Vogel. Ion association in hydrothermal sodium sulfate solutions studied by modulated ft-ir-raman spectroscopy and molecular dynamics. *J. Phys. Chem. B*, 119 (30):9847–9857, 2015.
- [22] Sadra Kashfolgheta and Ana Vila Verde. Developing force fields when experimental data is sparse: Amber/gaff-compatible parameters for inorganic and alkyl oxoanions. *Phys. Chem. Chem. Phys.*, 19:20593–20607, 2017.

Chapter 4

Ubiquitin Unfolding Under Native Conditions

4.1 Introduction :

The function of protein is closely associated with its tertiary structure. Water at the protein-water interface has been shown to play a major role governing the protein dynamics and its conformational fluctuations^[1]. An accurate representation of the interactions between the two in MD simulations, is therefore necessary to get the molecular level picture of protein folding/unfolding reactions. However, the current state of the art force fields for both protein and water models has been proven to be insufficient in sampling experimentally observed conformations for IDP's and unfolded proteins^[2;3]. Accuracy of the MD calculation depends on the underlying potential energy functions and presence of such discrepancies in the energy functions could seriously hamper the advent of MD simulations. In this study, we therefore assess the accuracy of modern protein and water models used for bimolecular simulations from the thermodynamical contexts.

The significance of water model in influencing protein dynamics and thermodynamics has been studied extensively in literature^[3;4]. Ranging from most widely used force fields such as AMBER99sb^[5], CHARMM-22*^[6] till the most recently proposed ones such as CHARMM-36^[7] and ff14sb^[8] are all parameterised against three decade old TIP3P^[9] water model, which produces many thermodynamical properties of pure water poorly compared with the more accurate ones such as TIP4P-Ew^[10], TIP4P-2005^[11;4] and TIP4P-D^[3] making its use as a protein solvent rather questionable. Moreover, these water models have been shown better to sample the more expanded conformations of unfolded proteins. Despite evidences illustrating the accuracy of newer water models in sampling the correct conformational ensembles and its suitability as a protein solvent^[3;4], the thermodynamics of protein in these solvents is still poorly understood and is the prime motivation behind this study.

In this study, we demonstrate the effect of water model on the unfolding thermodynamics of Ubiquitin whilst simultaneously highlighting the importance of choice of the correct water for bimolecular simulations. We show that TIP4P-Ew^[10] and TIP4P-2005^[11] water model shifts the protein folding/unfolding equilibrium on the unfolded side of the protein which results in unfolded states that are more stable than native state of the protein^[12;13]. We argue on the incompetence to TIP4P-Ew^[10] and TIP4P-2005^[11] water models in accurately describing the protein-water thermodynamics and associate it to the hydrogen bonding strength and overestimation of polar interaction energies^[12]. Moreover, we show that the TIP3P^[9] and the dispersion

corrected TIP4P-D^[3] water model yields the most realistic picture of protein unfolding.

4.2 Design and Methodology :

To gain insight into the chemical unfolding pathway of Ubiquitin and to find out the barrier separating the folded and unfolded basins of the Ubiquitin under physiological conditions (at 0.154M NaCl concentration) we've performed five Well-Tempered Metadynamics simulations, which constitutes of over 10 μ s of simulation time.

4.2.1 Preparation of Initial Configuration :

The crystal structure of the native state of the Ubiquitin was obtained from the RSCB-PDB (PDB ID: 1UBQ)^[14] data base^[15;16]. The initial coordinates and topology for Ubiquitin were prepared using GROMACS MD software^[17]. We have performed the simulations using five different force fields and are best summarized in the table below.

Force Field Details			
Simulation	Force Field	Water Model	Ion Model
SIM1	AMBER-99sb ^[18]	TIP4P-Ew ^[10]	Joung & Chetham ^[19]
SIM2	CHARMM-36m ^[20]	TIP4P-Ew ^[10]	Joung & Chetham ^[19]
SIM3	ff03ws ^[4]	TIP4P-2005s ^[11;4]	Joung & Chetham ^[19]
SIM4	RSFF(2+) ^[21]	TIP4P-D ^[3]	Joung & Chetham ^[19]
SIM5	AMBER-99sb ^[18]	TIP3P ^[9]	Joung & Chetham ^[19]

Table 4.1: Table summarizing the force-fields used across this study

Simulation Details :

We report five simulation of Ubiquitin under native conditions using force fields and associated water and ion (NaCl) models, as described in table 4.1. The initial structure of the protein was put in a cubic box of length 100 Å. The system was then solvated using associated water model mentioned in the table 4.1 with each force field and ions were added upto physiological salt concentration (0.154 M of NaCl). The system was then energy minimized using steepest-descent^[22] algorithm implemented in GROMACS^[17] followed by heating upto 300K using Berendsen thermostat^[23] under NVT conditions with a coupling constant of 0.4 ps. The heavy atoms of the protein was restrained during heating using a force constant of 25 Kcal/mol/Å², using harmonic potential. The force constant was then reduced to 0.5 Kcal/mol/Å² in six steps under NPT conditions followed by 5 ns unrestrained equilibration using Berendsen thermostat^[23](300K) and Parrinello-Rahman barostat(1 bar)^[24]. The unbiased 10 ns MD simulation was then performed on the equilibrated structure using Nose-Hoover thermostat^[25] and Parrinello-Rahman barostat^[24], each with a coupling constant of 0.4 and integration time step of 2 fs. The leap-frog integrator was used to integrate the equations of motions. The Particle Mesh Ewald (PME)^[26] treatment for long range electrostatics was used with a long range cutoff of 10 Å in addition to 10 Å cutoff for Van Der Walls interactions. All bond lengths and bond angles were constrained using LINCS algorithm^[27].

Metadynamics Simulations :

After production dynamics, we've performed multiple Well-Tempered Metadynamics (WTMD)^[28] simulations to explore the free energy surface of Ubiquitin. After careful investigation of several CV's, we've chosen Native contacts (N_c) and Radius of gyration (R_g) as our collective variables for reasons discussed in chapter 2. A constant hill height of 0.2 kJ/mol along with a constant hill width of 6 and 0.6 Å for N_c and R_g respectively was used throughout all our simulations. The biasfactor of 15 was chosen to bias the dynamics along the collective variables. Hills deposition rates was different in all cases and are best summarized in the following table.

Metadynamics Details	
Simulation no.	Deposition rate (ps)
SIM1	2
SIM2	0.4
SIM3	0.4
SIM4	4
SIM5	3

Table 4.2: Table summarizing the simulation settings for metadynamics

All our simulations was carried out using GROMACS(version 4.5.5)^[17] MD package patched with PLUMED program^[29] (version 1.3) for metadynamics simulations. Convergence of each metadynamics run was ensured using the sum-hills utility of the program PLUMED (version 1.3)^[29].

4.3 Results and Discussions :

We here, discuss the metadynamics simulations and the associated free energy surfaces for Ubiquitin unfolding with each simulation one by one, while simultaneously demonstrating the significance of the water model and it's effect on the protein unfolding dynamics and thermodynamics. Figure below depicts the free energy surface of Ubiquitin obtained using AMBER99sb and CHARMM-36m^[20] as force fields along with the widely used TIP4P-Ew^[10] water model.

Figures 4.1 and 4.2 depicts the free energy surfaces obtained with simulations using AMBER-99sb^[18] (SIM1) and CHARMM-36m^[20] (SIM2) force fields along with the TIP4P-Ew^[10] water model. As can be seen from the figures, each free energy surface shows two well defined minima. In free energy surface (FES) constructed for SIM1 and SIM2, the first minimum, with the characteristic value of $N_c = 300$, $R_g = 12$ Å and $N_c = 321$, $R_g = 12$ Å belongs to the native state structure of the Ubiquitin, for SIM1 and SIM2 respectively. We recognize, the second minimum, appearing around the $N_c = 150$, $R_g = 13$ Å and $N_c = 170$, $R_g = 12.1$ Å, for SIM1 and SIM2 respectively, as unfolded states of the protein and thus designate them as U1. Note that the native state is the state with the highest number of native contacts in the free energy surface as per the definition of Native contacts (N_c). The dashed lines in each free energy surface represents the calculated minimum free energy path using the algorithm of Ensing *et al.*^[30] as discussed in the chapter-2.

The one interesting feature of the free energy surfaces represented in figures 4.1 and 4.2 is the large difference in free energy of the folded and unfolded basins. The free energy of the

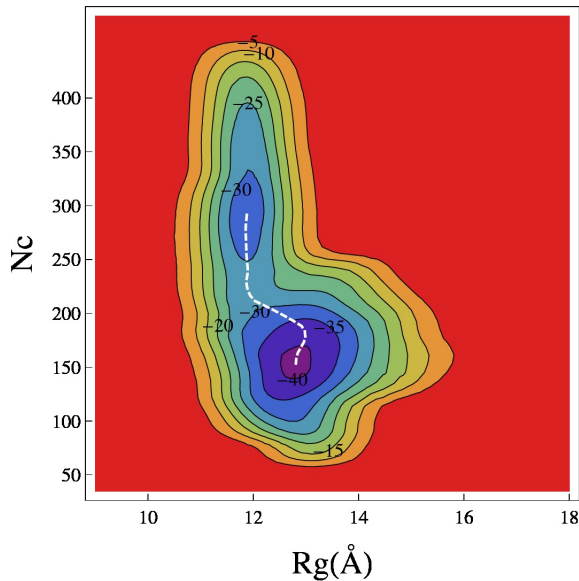


Figure 4.1: FES for SIM1

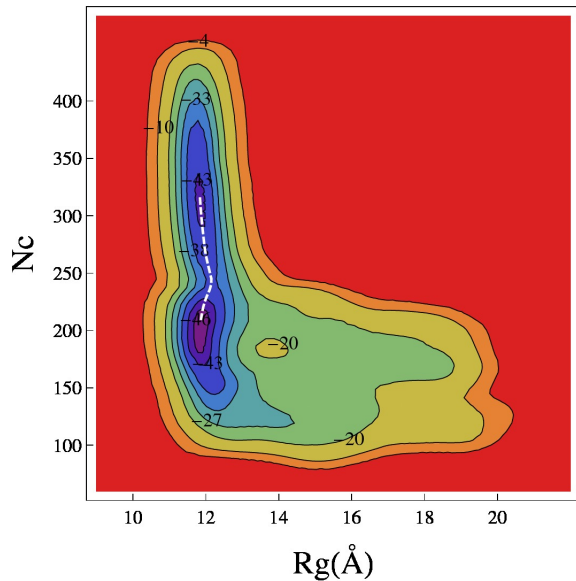


Figure 4.2: FES for SIM2

unfolded state in SIM1 and SIM2 is -40 kcal/mol and -46 kcal/mol respectively, which is way higher than the native state free energy of -33 kcal/mol and -43 kcal/mol. We attribute the large deviations among the free energies of the native and unfolded basins, in SIM1 and SIM2 to the difference in force fields used and simulation settings used for metadynamics and remains to be verified. Despite the large deviations amongst the free energies of the two basins in SIM1 and SIM2, one common fact is that the unfolded (U1) state is highly stable free energetically than the native state of the protein, which is rather contradictory to the prevailing belief that the native state of the protein under physiological conditions sits at the global minimum of the free energy landscape.

There are often increasing evidences in the literature that current state of the art protein and water force fields often overestimates the degree of compactness of unfolded and intrinsically disordered proteins i.e. the unfolded states studied in the simulation are too compact than that obtained through the experiments^[3;4]. Such a compact unfolded states in general could overestimate the magnitude of the salt-bridge interactions within the protein and thus leading to the overstabilization of unfolded ensembles. Thus, to authenticate whether and if the Unfolded (U1) states are really stable or it is mere an artifact of the modern force fields, we've carried out another simulation using ff03ws^[4] and TIP4P-2005s^[11;4] (SIM3) water model which has been shown to be successful in sampling more open conformational ensembles of proteins^[4]. The figure-4.3 represents the free energy surface obtained for Ubiquitin unfolding using the aforementioned force fields.

The combination of ff03ws^[4]+ TIP4P-2005s^[11;4] has been extensively validated against the experimental data and is proven to be accurate in sampling conformations of IDP's having R_g values similar to that of experiments. As is evident from the free energy surface in figure-4.3, the combination of ff03ws^[4]+ TIP4P-2005s^[11;4] does samples the extended conformations of proteins which is proved by the R_g values ranging from $12 \text{ \AA} \leq R_g \leq 36 \text{ \AA}$. The free energy surface in figure 4.3 clearly shows 3 well defined minima characterized by $N_c = 300, 173, 40$ and $R_g = 11.8 \text{ \AA}, 12 \text{ \AA}, 13 \text{ \AA}$, hereafter termed as N, U1 and U2 respectively. As is evident from the free energy profile of SIM3, figure 4.3, the unfolded states viz. U1 and U2 are way more stable than native state of the protein Ubiquitin (\approx stable by 5 and 6 Kcal respectively).

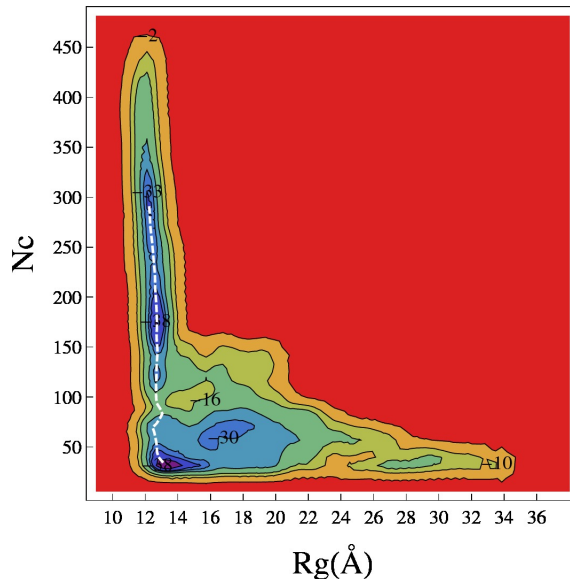


Figure 4.3: FES for SIM3

On the account that the combination of ff03ws^[4] + TIP4P-2005s^[11;4] is tuned to sample more open conformations of proteins, the likelihood of the discrepancy leading to the overestimation of stability of unfolded regions through salt bridge interactions is rather low, which brings our attention to the water model being employed.

As discussed earlier, the water governs the folding unfolding equilibrium of the proteins and can contribute significantly towards the stability of a particular structure of proteins. To examine whether the highly stable unfolded states in SIM1, SIM2 and SIM3 represent reality or mere an artifact arising out of the water model, we've performed simulations using TIP4P-D^[3] (SIM4) and TIP3P^[9] water models. The recently proposed Residue Specific Force Field (RSFF)^[21] is built over AMBER-99sb^[18] and was used insted of standard AMBER-99sb^[18] in SIM4 to ensure the compatibility of protein-water force fields. Moreover, TIP4P-D^[3] water model is known to destabilize the native state of the protein (≈ 2 Kcal) when paired with standard AMBER force fields^[3] while RSFF2(+)^[21] with TIP4P-D^[3] is shown to reproduce unfolding dynamics and thermodynamics of Trp-cage and β -hairpin quite realistically^[21]. The following figures shows the free energy profiles obtained using RSFF2(+)^[21] with TIP4P-D^[3] water model (SIM4) and AMBER99sb with TIP3P^[9] water model.

The free energy profile obtained using combination of RSFF2(+)^[21] and TIP4P-D^[3] water model is shown in figure 4.4. It is apparent that the free energy surface clearly shows three distinct minima centered around $N_c = 340, 200, 101$ and $R_g = 11.8 \text{ \AA}, 12 \text{ \AA}, 12.1 \text{ \AA}$, indicating the native and two unfolded states, hereafter labelled as N, U1 and U2 respectively . Correspondingly, the free energy surface for the SIM5 (figure 4.5), obtained using AMBER-99sb^[18] and TIP3P^[9] water model, shows two distinct minima at $N_c = 320, 101$ and $R_g = 1.17 \text{ \AA}, 1.25 \text{ \AA}$ termed hereafter as N and U1. Quite remarkably, the native state sits at the global free energy minimum with a free energy ≈ -39 Kcal/mol in both the free energy surfaces. The U1 and U2 states, in SIM4, have free energies ≈ -30 and -25 Kcal/mol respectively, whilst, in SIM5, the free energy of the unfolded U1 states is nearly about -32 Kcal/mol.

From the approach adopted. it is quite conspicuous that the water model has the pronounced effect on governing protein folding-unfolding equilibrium. At this stage, we formulate that the,

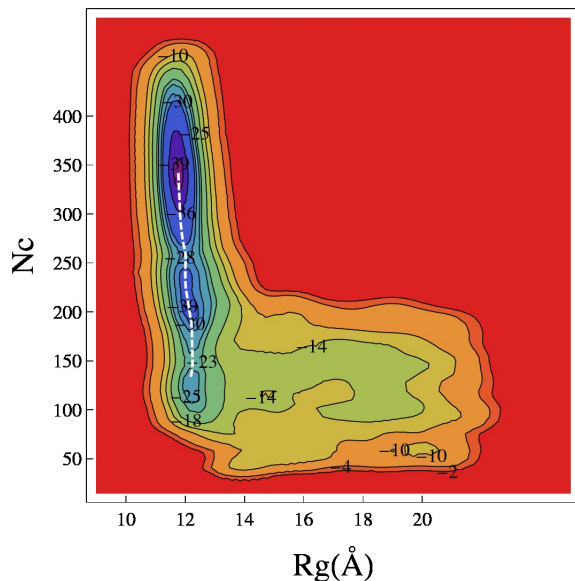


Figure 4.4: FES for SIM4

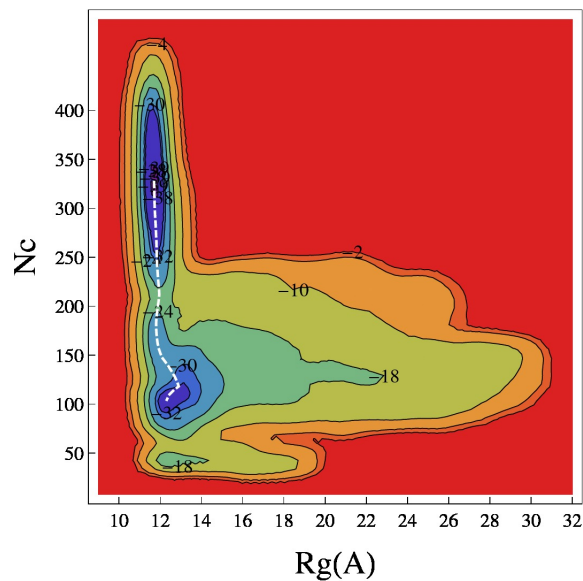


Figure 4.5: FES for SIM5

widely acknowledged TIP4P-Ew^[10] and TIP4P-2005s^[11;4] water models strongly favour the unfolded state ensemble of protein, whilst underestimating the native state stability and on the contrary, TIP3P^[9] and TIP4P-D^[3] water models represent the unfolding-folding thermodynamics quite realistically. To answer precisely the reason of stable unfolded states in SIM1, SIM2 and in SIM3, we explicitly state the barrier involved along the lowest free energy path, in the unfolding process of Ubiquitin, along with the hydrogen bond and energy analysis to further endorse our formulation and is discussed in the next sections.

4.4 Minimum free energy path and barriers of unfolding :

The most probable reaction mechanism, for any reaction, could well be approximated by the lowest free energy path connecting the reactant and product. To explain the over-stability of unfolded state ensembles, as predicted by TIP4P-Ew^[10] and TIP4P2005s^[11;4] water model and perhaps understand the unfolding mechanism of Ubiquitin, we calculated the minimum free energy path along both the free energy surface joining the Native and unfolded states, using the algorithm developed by Ensing *et al.*^[30] as discussed in the chapter-2. The figures below represent the relative free energy as a function of steps along the lowest free energy path, thus the MFEP index 1 represents the native state of the protein and corresponds to the state N in each free energy surface. Note that, a point in the MFEP index for one simulation need not correspond to it's counterpart in another free-energetically since all the simulations performed and the associated free energy surfaces are different.

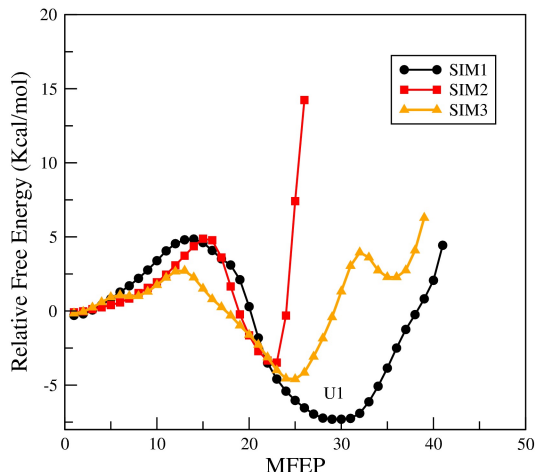


Figure 4.6: FE along MFEP for SIM1, SIM2, SIM3

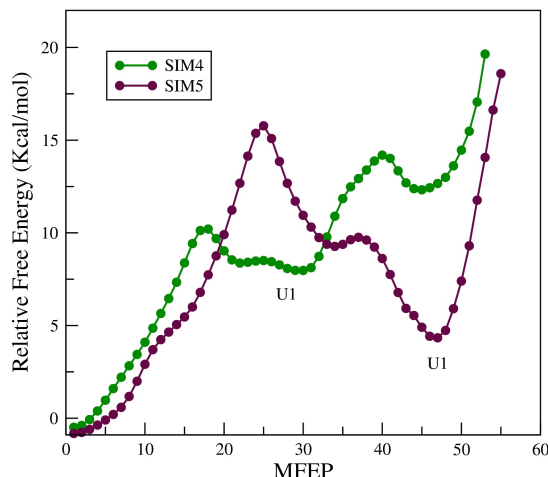


Figure 4.7: FE along MFEP for SIM4, SIM5

Figures 4.6 and 4.7, shows the varying stability of the unfolded states in each simulations and arises, as discussed later, due to the difference in hydrogen bonding strength and protein-water interactions. As the figures 4.6 and 4.7 shows, the folded and unfolded basin are separated by the free energy barrier along the lowest free energy path that the system must overcome in order to transition into the unfolded state of the protein. Table-4.3 summarizes the free energy barrier involved in all of the simulations.

Comparison of the free energy barriers			
Simulation	Native state F.E	Transition State F.E	Unfolded state F.E
No.	(kcal/mol)	(kcal/mol)	(kcal/mol)
SIM1	0	5	-5
SIM2	0	5	-5
SIM3	0	2	-7
SIM4	0	10	12.5
SIM5	0	15	4

Table 4.3: Table summarizing the relative free energies, barriers and free energy differences for SIM1-SIM5

Several studies report the barriers involved in the unfolding of Ubiquitin under different physical conditions. Piana *et al.*^[31] reported the unbiased temperature induced unfolding of Ubiquitin (at $T = 380$ K), which is near to the melting temperature for Ubiquitin (≈ 373 K), using the CHARMM-h force field for the protein and TIP3P^[9] water model. The free energy barrier separating folded and unfolded regions was reported to be around ≈ 5 Kcal/mol. Another study, performed by Mukopadhyay *et al.*^[32] with the denaturing Guanidinium ions, also reports the barrier to be around ≈ 5 -6 Kcal/mol. It is widely acknowledged that presence of denaturant and high temperature conditions increasingly destabilize the native state of the protein by shifting the folding-unfolding equilibrium towards the unfolded states and thus simultaneously reducing the free energy barrier required to unfold the protein. We infer that the free energy barrier separating the folded and unfolded basins at temperatures near to ≈ 300 K and

under physiological (native) conditions should therefore be generally higher than that reported by Piana *et al.*^[31] and by other studies.

However, As the table 4.3 shows, the free energy barriers involved in the unfolding of Ubiquitin was approximately ≈ 5 Kcal/mol for SIM1 and SIM2 while it was as low as ≈ 2 Kcal/mol for SIM3, which endorses the claim that the TIP4P-Ew^[10] and TIP4P-2005s^[11;4] water model significantly destabilize the native state of the protein^[12;13]. The free energy barriers for SIM4 and SIM5 are significantly high compared with other simulations (≈ 10 Kcal/mol, 15 Kcal/mol for SIM4 and SIM5 respectively) and may well represent the true physical barrier involved as Ubiquitin is normally found in monomeric folded state *in vivo*.

4.5 Hydrogen Bonding and Residue Interaction energy :

Hydrogen bonding interactions, although weak in nature, can significantly alter the folding-unfolding thermodynamics of the particular ensemble of states of the proteins. As reported, increasing peptide water hydrogen bonding could lead to either destabilization or stabilization of the native structures of the protein owing to the significant difference in hydrogen bonding energetics accompanying the modern water models^[12]. Therefore to further elucidate the origin of over-stabilization of unfolded states in SIM1, SIM2 and SIM3 , we've monitored protein water hydrogen bonding network in the folded and unfolded regions on the free energy surface. We also calculate the interaction energies of the polar and non-polar residues across the different water models studied, in unfolded (U1) ensembles of Ubiquitin to verify if the stabilization is arising out of the preferential solvation of polar/non-polar residues. Figures below depicts the above mentioned quantities.

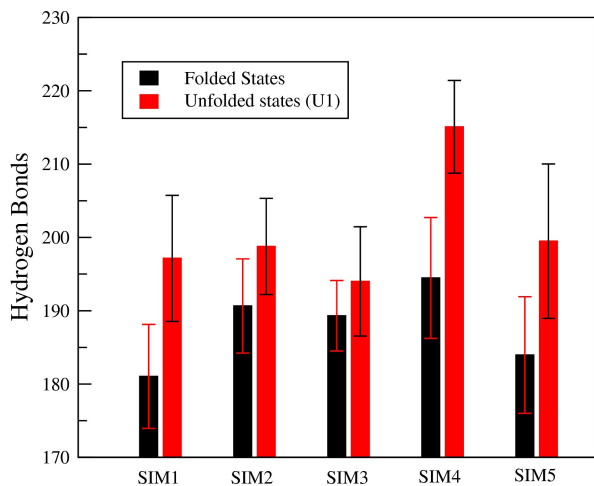


Figure 4.8: Hydrogen Bonding

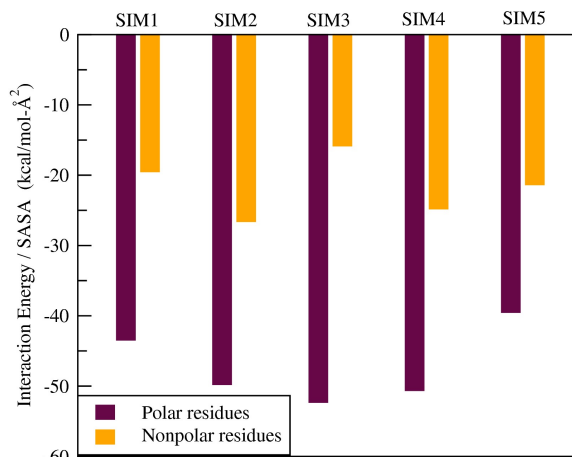


Figure 4.9: Non-polar solvation energies

Figures-4.8 represents the average protein-water hydrogen bonding in the folded (N) and the unfolded states (U1) across all the simulations while figure-4.9 depicts the average interaction energy between polar/non-polar surfaces of Ubiquitin and the water molecules in the unfolded (U1) states of the protein. Note that, the figure-4.9 does not show any error bars and is due to the fact that quantity interaction energy per unit solvent accessible surface area follows the ratio distribution, the mean and variance for which is ill-defined. As can be easily picked form the figure 4.8, unfolded state (U1) is always characterized by the higher number of hydrogen bonds than native state (N) of the protein across all the simulations performed. The larger hydrogen

bonding observed in the unfolded states could be attributed to larger solvent accessible surface area owing to the loss of tertiary interactions (N_c) in the proteins.

All of the simulations performed, with the notable exception of SIM4, have an approximately equal number of protein-water hydrogen bonds in the unfolded state (U1) of the protein. Paschek *et al.*^[12], in a seminal paper showed that the strength of protein-water hydrogen bonding differs across different water models and TIP4P-Ew^[10] water model favors the unfolded ensemble of states of proteins by an additional stabilization of -0.4 kJ/mol per hydrogen bond compared to the TIP3P^[9] water model. Since the number hydrogen bonds is approximately equal in SIM1, SIM2, SIM3 and in SIM5, we thus understand that the presence of an additional bias of -0.4 kJ/mol per hydrogen bond, in SIM1 and SIM2 (both performed with the TIP4P-Ew water model), shifts the populations towards the unfolded states of the protein. We approximate that the similar effect could well be present in case of TIP4P-2005s^[11;4] water model (SIM3) since both TIP4P-Ew^[10] and TIP4P-2005^[11;4] were derived originally from the TIP4P water model. Since such a bias is absent in TIP3P^[9] water, there is no population shift and thus native state sits at the global minimum in the free energy surface.

Protein-water interaction energies i.e Coloumbic and Lennered-Jones interaction energies, if overestimated, could also favor less structured configurations having the higher interactions with water over the structured one and such an amplification of protein-water interactions is indeed one of the strategies being used to sample more open and less structured configurations of IDP's^[3;4]. Therefore, an appropriate balance of protein-water interactions is absolutely necessary to correctly monitor the folding-unfolding transitions of proteins. Proteins are composed of hydrophobic and hydrophilic patches (amino acids) and therefore monitoring their interactions with water could as well explain the shift in population towards the folded/unfolded states of proteins. Figure 4.9 represents the protein-water interaction energies for polar and non-polar residues, in the unfolded (U1) state across all the simulations. Note that the interaction energies had to be normalized with the solvent accessible surface area (SASA), since the compactness of the unfolded state (U1) differs across all the simulations.

As the figure 4.9 shows, polar surfaces, as expected, has the higher interaction energy per unit solvent exposed surface area than their non-polar counterpart, across all simulations owing to the fact that polar patches of the protein are usually solvent exposed. Furthurmore, as is apparent in the figure-4.9, the interaction energy for non-polar surfaces is considerably lower compared with polar ones because the hydrophobic parts of the protein are usually buried within the hydrophobic core. Moreover, the non-polar interaction energies, do not follow any pattern (figure-4.9) and is thus less likely to be dominant factor when considering protein-water interactions. A closer look at the figure 4.9, shows that the in SIM1, SIM2 and in SIM3 polar surfaces has much negative interaction with water than in SIM5, which advocates that the polar residues are preferentially solvated^[12] in TIP4P-Ew^[10] and TIP4P-2005^[11;4] water model as compared with TIP3P^[9] water model.

Such a preferential solvation of the residues of protein, as discussed earlier, could shift the folding-unfolding equilibrium on the unfolded side by favouring the unfolded conformations of protein which solvates such residues better. We thus believe that such preferential solvation of polar groups along with the added bias of -0.4 kJ/mol per hydrogen bond, in simulations using TIP4P-Ew^[10], TIP4P-2005s^[11;4] are few of the underlying causes of the overstabilisation of the unfolded (U1) states and underestimation of the free energy barrier involved in unfolding of Ubiquitin. Note that we've constantly left SIM4 out of our discussion, since it has has highest number of protein-water hydrogen bonds and high polar-water interaction energies whilst still

yielding correct free energy profile and free energy barrier. While the mechanism of action of TIP4P-D^[3] water model still remains illusive, we think that other factors such as protein tertiary interactions, protein-ion interactions and protein conformational entropy might be playing a major role here.

As discussed in previously, simulations using AMBER99sb with TIP3P^[9] and RSFF(2+)^[21] with TIP4P-D^[3] both give quite realistic picture of Ubiquitin unfolding thermodynamics under physiological conditions. The combination of AMBER99sb with TIP3P^[9] is comparatively older than RSFF(2+)^[21] with TIP4P-D^[3] water model and both force fields along with the respective water models have been shown to represent unfolding thermodynamics of Trp-cage protein quite realistically^[21;18]. RSFF is largely built over AMBER99sb by Wu & coworkers^[21], with a modification that explicitly involves the optimization of long range n-n+4 interactions, a characteristic feature of α -helices within proteins. Conventionally, unfolded states of the protein are characterized by the loss of tertiary interactions, higher radius of gyration as well as lesser secondary structure content than the native state. However, we fear that such a modifications of the force fields could, in principle, lead to the unfolded state having the intact secondary structure content. Moreover, the combination of RSFF(2+)^[21] with TIP4P-D^[3] is rather recent and have not been yet tested while the older combination of AMBER99sb with TIP3P^[9] has been extensively validated against experimental data and shown to reproduce the unfolding thermodynamics quite realistically^[12;18] and is therefore used as a our reference for comparison to analyse the effect of other anions on Ubiquitin unfolding.

4.5.1 Mechanism of Unfolding of Ubiquitin under native conditions :

Ubiquitin unfolding has been investigated thoroughly in literature using different approaches viz. mechanical force induced unfolding experiments^[33], denaturant induced unfolding of Ubiquitin^[32] and temperature induced unfolding^[34] and all-atom MD simulation near it's melting temperature ($T_m = 373$ K)^[31]. Ubiquitin is comprised of five β strands arranged in parallel-antiparallel fashion, one α -helix and one 3_{10} -helix and stands as best model for investigating protein folding-unfolding pathways. Several studies reported the two step unfolding pathway for Ubiquitin which includes formation of stable core consisting of α -helix and $\beta_2\beta_1$ followed by the global unfolding, using NMR experiments and MD simulations^[35;36;37;38]. Our results further backs the two state unfolding of Ubiquitin and are discussed below. For our convenience, we designate the secondary structural units of the Ubiquitin and are summarized in the table below.

Summary of the labels used	
Residue Index	Designations
Res. 1-17	$\beta_2\beta_1$
Res. 11-17 & Res. 65-72	$\beta_1\beta_5$
Res. 65-72 & Res. 37-46	$\beta_5\beta_3$
Res. 22-35	α -Helix
Res. 55-60	3_{10} -Helix

Table 4.4: Table summarizing the labels for the secondary structural segments used in this study

The secondary structural units of the protein e.g. α -helix and β -sheet are usually held together by the hydrogen bonds. β -sheets are formed by hydrogen bonding interactions within

the individual β -strands of the protein while α -helical regions are characterized by their $i-i+3$ and $i-i+4$ hydrogen bonding networks. Therefore, in order to investigate the unfolding mechanism of Ubiquitin under native conditions, we've monitored the hydrogen bonding pattern along the lowest free energy path (MFEP) for each secondary structural segment and is shown in the figure-4.10.

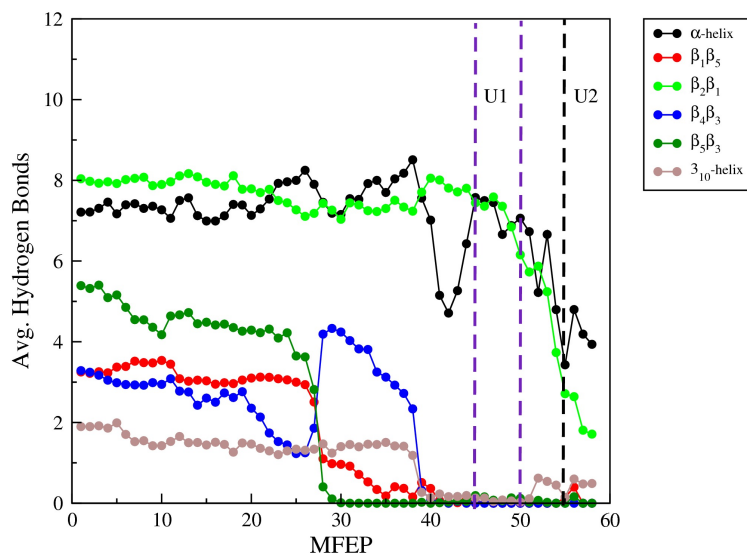


Figure 4.10: Mechanism of unfolding of secondary structural segments of Ubiquitin

As can be seen readily from the figure 4.10, the α -helix and $\beta_2\beta_1$ are perhaps the most stable and rigid secondary units of the protein as indicated by the constant number of hydrogen bonds from the native to U1 state of the protein (MFEP index 1-40). Note that, presence of this secondary structural units (one α -helix & $\beta_2\beta_1$) in U1 state is in a direct agreement with the other studies reporting unfolding mechanism of the Ubiquitin. On the other hand, as can be seen in the figure, $\beta_5\beta_3$ and $\beta_1\beta_5$ are comparatively unstable and undergoes unfolding around the transition state region (MFEP index 23-26) of the unfolding pathway. The smallest β -sheet units of the Ubiquitin i.e $\beta_4\beta_3$ unfolds via gradually decrease the number of hydrogen bonds, followed by the sudden increase along the transition state while completely dissociating in the U1 state of the protein. The smaller 3_{10} -helix, despite having lowest number of hydrogen bonds (≈ 2), maintains its structure through the transition state before breaking in the metastable U1 state of the pathway.

Overall, the transition state (MFEP index 23-26) is characterized by the abrupt loss of hydrogen bonding interactions in the secondary structural units viz. $\beta_5\beta_3$ and $\beta_1\beta_5$, and are first to unfold. The smaller secondary units $\beta_4\beta_3$ and 3_{10} -helix maintains their structure through the transition state whilst unfolding completely in the metastable U1 state. Owing to the high number of hydrogen bonding interactions within the native state structure of the $\beta_2\beta_1$ and α -helix, these units are last one to break and survive through the U1 state and undergoing unfolding only after the U1 state has occurred.

4.6 Conclusion :

The given discussion illustrates that the, dynamics and the thermodynamics associated with the protein are greatly affected by the choice of water model and such effects should not be underestimated. By specifically studying the unfolding of the Ubiquitin under native conditions and mapping its free energy surface, we illustrate that the TIP4P-Ew^[10] and TIP4P-2005^[11;4] water model with scaled protein-water interactions, populates the unfolded states of the protein by shifting the folding-unfolding equilibrium towards the unfolded side of the protein. We attribute this observed effect to the difference in the hydrogen bonding strength (-0.4 kJ/mol per hydrogen bond) and overestimation of the solvation energy for the polar side chains for TIP4P-Ew^[10] and TIP4P-2005s^[11;4] water models, as compared with the TIP3P water model^[9]. By constructing the free energy surface for unfolding, we simultaneously suggest that the combination of the RSFF(2+)^[21] along with TIP4P-D^[3] water model and AMBER-99sb^[18] with TIP3P^[9] water model yields the most realistic picture of the unfolding thermodynamics of the Ubiquitin.

The unfolding dynamics of the protein Ubiquitin is widely studied under various pathogenic condition and to the best of our knowledge, we report the unfolding dynamics of the Ubiquitin under native conditions for the first time^[31;32]. We illustrate that the Ubiquitin follows the two step unfolding pathway even under native conditions, as illustrated earlier by many other studies and suggest that metastable U1 state is characterized by the presence of secondary structural elements (α -helix and $\beta_2\beta_1$) which eventually unfolds in the U2 state encountered along the minimum free energy path.

Bibliography

- [1] Der A, Kelemen L, Fabian L, Taneva SG, Fodor E, and Pali T et. al. Interfacial water structure controls protein conformation. *J Phys Chem B*, 111:5344–5350, 2007.
- [2] R. B. Best and J. Mittal. Free-energy landscape of the gb1 hairpin in all-atom explicit solvent simulations with different force fields: Similarities and differences. *Proteins*, 79(4):1318–1328, 2011.
- [3] Stefano Piana, Alexander G. Donchev, Paul Robustelli, and David E. Shaw. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B*, 119:5113–5123, 2015.
- [4] Robert B. Best, Wenwei Zheng, and Jeetain Mittal. Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.*, 10:5113–5124, 2014.
- [5] J. M. Wang. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules ? *J. Comput. Chem.*, 21:1049–1074, 2000.
- [6] A. D. MacKerell. Importance of the cmap correction to the charmm22 protein force field: Dynamics of hen lysozyme. *Biophys. J.*, 90:L36–L38, 2006.
- [7] R.B. Best, X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, M. Feig, , and A.D. MacKerell Jr. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi1 and chi2 dihedral angles. *Journal of Chemical Theory and Computation*, 8:3257–3273, 2012.
- [8] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, and C. Simmerling. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *J. Chem. Theory Comput.*, 11:3696–3713, 2015.
- [9] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.
- [10] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, 120:9665, 2004.
- [11] J. L. F. Abascal and C. J. Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *Chem. Phys.*, 123(234505), 2005.
- [12] Dietmar Paschek, Ryan Dayb, and Angel E. Garcia. Influence of water–protein hydrogen bonding on the stability of trp-cage miniprotein. a comparison between the TIP3P and TIP4P-Ew water models. *Phys. Chem. Chem. Phys.*, 13:19840–19847, 2011.
- [13] Ferruccio Palazzesi, Matteo Salvalaglio, Alessandro Barducci, and Michele Parrinello. Communication: Role of explicit water models in the helix folding/unfolding processes. *Chem. Phys.*, 145:121101, 2016.
- [14] Vijay-Kumar Senadhi and Charles E. Bugg and William J. Cook. Structure of ubiquitin refined at 1.8Å resolution. *J. Mol. Biol.*, 194(3):531–544, 1987.

-
- [15] P. W. Rose, B. Beran, C. Bi, W. F. Bluhm, and Dimitropoulos. The rcsb protein data bank:redesigned web site and web services. *Nucleic Acids Res.*, D392-D401:2011, 39.
- [16] N. Deshpande, K. J. Address, W. F. Bluhm, J. C. Merino-Ott, W. Townsend-Merino, Q. Zhang, C. Knezevich, L. Xie, L. Chen, and Z. Feng. The rcsb protein data bank: A redesigned query system and relational database based on the mmcif schema. *Nucleic Acids Res.*, 33:D233–D237, 2005.
- [17] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, J. C. Shirts, M. R. and Smith, P. M. Kasson, and al. vander Spoel, D. et. Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29:845–54, 2013.
- [18] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, and Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, 65:712–725, 2006.
- [19] In Suk Joung and Thomas E. Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112:9020–9041, 2008.
- [20] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M Feig, de Groot, Grubmuller H. B.L., and A.D. Jr. MacKerell. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14:71–73, 2016.
- [21] Hao-Nan Wu, Fan Jiang, and Yun-Dong Wu. Significantly improved protein folding thermodynamics using a dispersion-corrected water model and a new residue-specific force field. *J. Phys. Chem. Lett.*, 8 (14):3199–3205, 2017.
- [22] W. H. Press. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge university press, 2007.
- [23] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [24] M. Parrinello and A Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys*, 52:7182–7190, 1981.
- [25] S. I. NosÉ. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 100:191–198, 2002.
- [26] T. Darden, D. York, and L. J. Pedersen. Particle mesh ewald: An n · log (n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.
- [27] Berk Hess, Henk Bekker, Herman J.C.Berendsen, and Johannes G.E.M.Fraaijey. Lincs:a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18:1463–1472, 1997.
- [28] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, 2008.
- [29] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, and R. A. Broglia. Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, 180:1961–1972, 2009.
-

- [30] B. Ensing, A. Laio, M. Parrinello, and M. L. A Klein. Recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B*, 109:6676–6687, 2005.
- [31] Stefano Piana, Kresten Lindorff-Larsen, and David E. Shaw. Atomic-level description of ubiquitin folding. *PNAS*, 110 (15):5915–5920, 2013.
- [32] Manoj Mandal and Chaitali Mukhopadhyay. Microsecond molecular dynamics simulation of guanidinium chloride induced unfolding of ubiquitin. *Phys. Chem. Chem. Phys.*, 16:21706–21716, 2014.
- [33] Anders Irbäck, Simon Mitternacht, and Sandipan Mohanty. Dissecting the mechanical unfolding of ubiquitin. *PNAS*, 102:13427–13432, 2005.
- [34] Hoi Sung Chung, Munira Khalil, Adam W Smith, Ziad Ganim, and Andrei Tokmakoff. Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin. *PNAS*, 102:612–617, 2005.
- [35] Patrick L Wintrode, George I Makhatadze, and Peter L Privalov. Thermodynamics of ubiquitin unfolding. *Proteins: Structure, Function, and Bioinformatics*, 18:246–253, 1994.
- [36] Tobin R Sosnick, Robin S Dothager, and Bryan A Krantz. Differences in the folding transition state of ubiquitin indicated by φ and ψ analyses. *PNAS*, 101:17377–17382, 2004.
- [37] E Larios, JS Li, K Schulten, H Kihara, and M Gruebele. Multiple probes reveal a native-like intermediate during low-temperature refolding of ubiquitin. *Journal of molecular biology*, 340:115–125, 2004.
- [38] Hoi Sung Chung, Ziad Ganim, Kevin C Jones, and Andrei Tokmakoff. Transient 2d ir spectroscopy of ubiquitin unfolding dynamics. *Proceedings of the National Academy of Sciences*, 104:14237–14242, 2007.

Chapter 5

Ubiquitin Unfolding In Presence of Hofmeister Anions

5.1 Introduction

Protein folding/unfolding equilibrium has been widely investigated, in recent year in literature under the various physical conditions^[1;2]. The folding/unfolding dynamics and thermodynamics of proteins are greatly affected by the factors including pH, presence of col-solutes viz. Urea and GdmCl and higher temperature^[3]. Nevertheless, despite being proven to significantly affect the protein folding/unfolding equilibrium, salt induced changes to the protein dynamics and thermodynamics has largely remain elusive compared to their counterparts viz. Urea and TMAO. As the salts are ubiquitous under physiological conditions, investigating protein thermodynamics in salt solutions and under a range of concentrations becomes rudimentary to dig further deep into the mechanism of actions of these ions, thereby elucidating the long-standing enigma of the Hofmeister series.

As previously discussed, Hofmeister effects are of central importance in governing many phenomena including protein stability, protein crystallization, bacterial growth and colloidal assembly etc.^[4] Sulphate ions, in general, are classified as *kosmotropes* and lie on the "stabilizing or salting out" end of the series^[5]. Despite that, there's ever increasing debate on the mechanism of action of sulphate anions on protein stability, as few groups suggests it's contribution to protein stability^[6] while others argues that despite being kosmotropic in nature, SO_4^{2-} behaves in chaotropic manner leading to the destabilisation of the native state of the protein^[7]. Apart from sulphate ions, ions such as ClO_4^- , SCN^- (*chaotropes*) lie on the "destabilizing and salting-in" end of the Hofmeister series^[5] and are well-known protein denaturants. Many emergent approaches attribute their denaturing capacity to their tendency to form hydrogen bonds with the protein backbone^[7;8]. However, despite the long term presence of these Hofmeister effects, the underlying mechanism of action of such salts still remains elusive and is the topic of investigation of this study.

Thus, in order to close the ever-lasting argument on the nature of the sulphate-protein interactions and perhaps shed some light on the mechanism of action of ClO_4^- and SCN^- ions, we here investigate the unfolding thermodynamics of the Ubiquitin using all-atom molecular dynamics simulation coupled with metadynamics. We report, for the first time that the sulphate anions do not have a pronounced effect on the native state stability, however increases the unfolding free energy barrier by destabilizing the unfolded state ensembles of proteins. We attribute the observed instability of the unfolded states to the hydrophobic residues-water interactions and

suggest that sulphate anions solvates the hydrophobic residues of the protein. Moreover, we also discuss the sulphate ion induced changes in the unfolding dynamics of the Ubiquitin in comparison with the unfolding dynamics under native conditions (chapter-4) and report that secondary structural fragments are increasingly unstable in the sulphate solutions. Lastly, we discuss the unfolding of Ubiquitin under the effect of ClO_4^- and SCN^- ions.

5.2 Design and Methodology

To understand the mechanism of action of Hofmeister salts and to gain insight into the barrier separating the unfolded regions from folded regions thereby measuring the extent of stabilization/destabilization of Ubiquitin in salt solutions, we've carried out a total of three simulations using the ions from the both ends of Hofmeister series viz. SO_4^{2-} , ClO_4^- , SCN^- .

Preparation of Initial Configuration :

The crystal structure of native state of Ubiquitin was obtained from the RSCB-PDB^[9;10] data base (PDB ID: 1UBQ^[11]). The initial coordinates and topology were prepared using GROMACS-MD^[12] software. In each of the simulation, protein was modeled using AMBER-99sb^[13] parameter set while TIP3P^[14] water model was used across all simulations for reasons discussed in chapter-4. The sodium ions, by Joung-Cheatham^[15], were used across all the systems to neutralize the net negative charge on the system. The anionic force-field parameter set for SO_4^{2-} was obtained from Ana-Villa-Verde *et al.*^[16] (for reasons discussed in chapter-3) while parameters for ClO_4^- and SCN^- were adopted from Badden *et al.*^[17] and Vincze *et al.*^[18] respectively.

Simulation Details :

We report three simulation of Ubiquitin using various anions as described in previous section. The initial structure of the native Ubiquitin was put in a cubic box of length 10 nm. Protein was solvated using TIP3P water model followed by ion insertion to yield the concentration of 0.3 M. The system was then energy minimized using the Steepest-descent algorithm^[19] implemented in GROMACS^[9;10]. Energy minimization was followed by simulated-annealing upto 300 K under NVT conditions using Berendsen thermostat^[20] and Parrinello-Rahman barostat^[21]. The heavy atoms of the protein was restrained during heating using force constants of $25 \text{ kcal/mol}/\text{\AA}^2$, using harmonic potential. The force constant was then reduced $0.5 \text{ kcal/mol}/\text{\AA}^2$ in six steps under NPT conditions followed immediately by 5 ns unrestrained equilibration using Berendsen thermostat^[20] and Parrinello-Rahman barostat^[21] with a coupling window of 0.4 ps each. 10 ns of normal unbiased MD simulation was then performed using the equilibrated structure under NPT contions using Nose-Hoover thermostat^[22] and Parrinello-Rahman barostat^[21], each with a coupling constant of 0.2 ps and using intergation time step of 2 fs. The leap-frog integrator was used to integrate the equations of motion. Long range electrostatics and Van Der Walls interactions were calculated according to the Particle Mesh Ewald (PME)^[23] and cutoff scheme respectively, each using a cutoff value of 10\AA . All bond lengths and bond angles were constrained using LINCS algorithm^[24].

Metadynamics Details :

The Well-Tempered Metadynamics simulations (WTMD)^[25] were performed after 10 ns of unbiased MD simulation in order to explore the associated free energy surface with each anion. As discussed earlier in chapter-2, N_c and R_g were used as collective-variables for describing the

unfolding dynamics of Ubiquitin. A constant Gaussian hill height of 0.2 kJ/mol along with a constant hill width of 6 and 0.6 Å for N_c and R_g respectively was used throughout all our simulations. After trying out several values, the biasfactor of 15 was then chosen to bias the dynamics along N_c and R_g . The hills deposition rate of 3 ps was kept constant across all the simulations. All metadynamics simulations were performed using PLUMED(version-1.3.0)^[26] code. The minimum free path (MFEP) was calculated, according to the algorithm proposed by Ensing *et al*^[27], as discussed in chapter-2.

5.3 Results and Discussion

In this section, we discuss effect of sulphate anions on the unfolding thermodynamics of Ubiquitin while simultaneously highlighting the salt induced changes in the unfolding dynamics of Ubiquitin. We also address the ongoing debate on the nature of the sulphate ion (*kosmotropic* or *chaotropic*) in influencing the stability of the proteins and discuss the corresponding mechanism whilst closing the contest of observed reversal of Hofmeister series near charged side-chains once and for all.

5.3.1 Direct Ion-Protein Interactions :

Earlier studies often regarded sulphates as kosmotropic in nature and thus being largely excluded from interacting directly with the protein surface^[5]. While this assumption has largely been proven incorrect, increasing attention has been paid on the direct interaction mechanism concerning ion and protein. Jungwirth *et al.*^[7], using the reductionist approach suggested that the kosmotropic anions might be involved in the direct interaction with the positively charged side chains residues of the protein. Gokran *et al.*^[28], using charge density measurements, illustrated that sulphate ions despite being kosmotropic, are involved in the direct interaction with the lysozyme even at the concentration range as low as ≤ 0.1 M. Therefore, to understand the sulphate ion distribution around protein, we calculate the hydrogen bonding and interaction energies between protein and sulphate anions on the normal unbiased MD trajectories. Our results indicate the sulphate ions are indeed involved in a direct interaction between protein surface are discussed below.

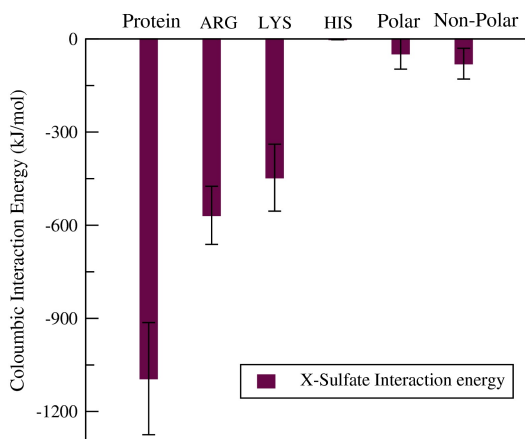


Figure 5.1: Interaction Energies

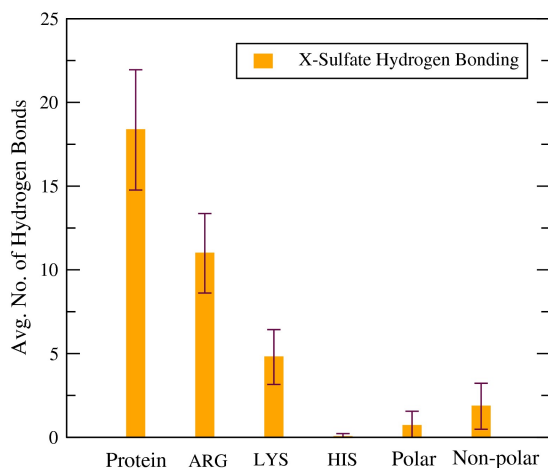


Figure 5.2: Hydrogen Bonding

To elucidate the interaction mechanism between sulphate and Ubiquitin, we monitored the hydrogen bonding and the interaction energy between the sulphate ions and the positively

charged residues on the surface of the Ubiquitin viz. Arginine, Lysine, Histidine and polar-nonpolar regions of the protein. Inspection of the calculated coulombic interaction energies (figure 5.1) clearly suggest that the sulphate ions despite being kosmotropic in nature are involved in the direct interaction with protein surface which is well in agreement with the previous studies^[7:28]. Furthermore, as shown in the figure 5.1, sulphate ions interacts most prominently with the positively charged side chains of the Ubiquitin, which constitutes of $\approx 15\%$ of the residues of the protein while accounting for nearly 83% of total protein-sulphate interaction energy. Moreover, Arginine side chains (total ARG = 4) of the protein interacts more preferentially with the sulphate ions than the corresponding Lysine (total LYS = 7) residues while the Histidine residue (total HIS = 1) does not react at all due to it's position which lies within the core of the protein Ubiquitin. Note that the sulphate ions largely remain excluded from the interaction with polar or non-polar surfaces as indicated by their substantially lower interaction energy value.

As the figure 5.2 depicts, the average number of hydrogen bonds between the sulphate ions and the protein residues is ≈ 18 out of which 11 hydrogen bonds are contributed by Arginine residues while Lysine residues contribute to ≈ 5 hydrogen bonds. Note that Histidine residues, as previously stated, do not interact much with sulphate ions and hence do not contribute much to the hydrogen-bonding interactions. Moreover, sulphate ions are not involved in the hydrogen bonding with any polar and non-polar residues figure (5.2) which forces us to think that the coulombic interactions might be the key in governing observed high degree of hydrogen bonding for Arginine and Lysine residues of the protein. To investigate the lifetime of the interactions of the sulphate ions with the Arginine and the Lysine side chains, we monitored the minimum distance between oxygen on the sulphate atom and the nitrogen atoms on the arginine and lysine side chains on a 10 ns unbiased MD trajectory.

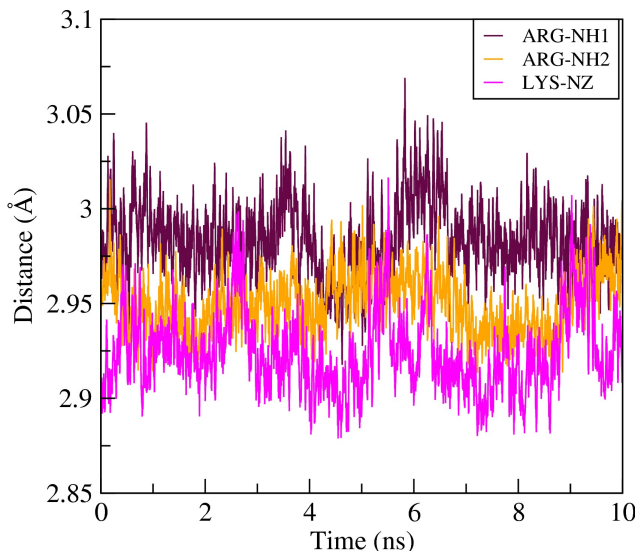


Figure 5.3: Minimum distance between "N" of protein residues and "O" of sulphate ions as a function of time

Figure 5.3 plots the running averages of the distance between the oxygen on the sulphate and the nitrogen on the side chains. As can be seen from the figure, the distance between the oxygen of sulphate and nitrogen of the side-chains, remains constant throughout the simulation, which suggest that sulphate ions are almost always involved in the interactions with the positively charged residues of the protein. Note that, there may be the exchange of sulphate ions and

figure only plots the distance to the closest oxygen on the sulphate and says nothing about the residence time distribution of sulphate ions around the side chains.

5.3.2 Free energy surface of Ubiquitin unfolding in sulphate solutions :

Figure 5.4 depicts the free energy surface of Ubiquitin unfolding in sulphate solution. The constructed free energy surface, figure 5.4, shows three well defined minima and the dashed line connecting them corresponds to the calculated lowest free energy path using algorithm developed by Ensing *et al*^[27], along the free energy surface, as discussed in chapter-2. We label the first minimum with the characteristic value of $N_c = 330$ and $R_g = 12.1 \text{ \AA}$ as "N" since it corresponds to the native state of the protein. Second and third minimum, corresponding to $N_c = 105$, $R_g = 15.8 \text{ \AA}$ and $N_c = 30$, $R_g = 14 \text{ \AA}$ corresponds to the unfolded states of the protein and are hereafter tagged as "U1" and "U2" respectively. As can be seen in the figure 5.4, the free energies of the N, U1 and U2 states are ≈ 39.5 , 26 and 32 kcal/mol respectively and the native state of the protein lies on the absolute minimum in the free energy surface.

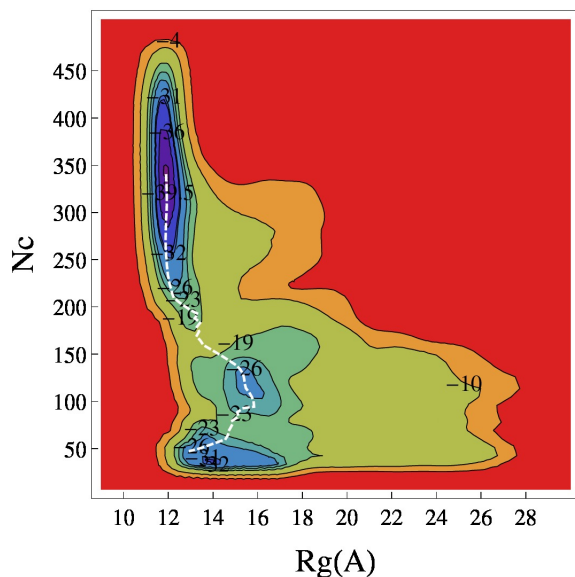


Figure 5.4: FES for SO_4^{2-}

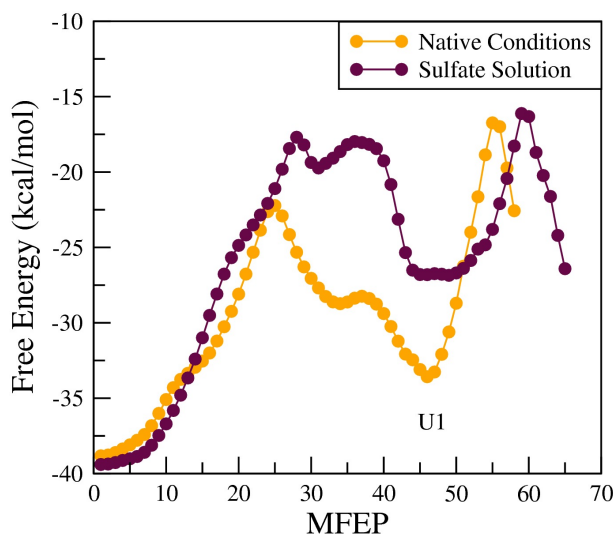


Figure 5.5: MFEP for SO_4^{2-}

To further perceive the role of sulphate ions in stabilization/destabilization of the native state, we've calculated the lowest free energy path (figure 5.5) describing the unfolding reaction of the Ubiquitin. As is evident from the figure 5.5, the free energy barrier involved in separating the folded and the unfolded basins from each other under the influence of sulphate ions is ≈ 5 kcal/mol greater than the free energy barrier for unfolding of Ubiquitin under native conditions. The increased barrier height clearly indicates that sulphate salts indeed favour the native structure of the protein by increasing the unfolding free energy barrier involved. Interestingly, the free energy value for the native state (N) of the Ubiquitin in sulphate solution (≈ 39.5 kcal/mol) corresponds well with the free energy value of state N under the native conditions (≈ 39 kcal/mol in SIM5 of chapter-4) which suggest that, despite the presence direct and persistent ion-protein interactions, sulphate salts do not appreciably alter the stability of the native state of the protein, as opposed to few earlier studies^[7;28].

A closer look at the figure 5.5 suggest that, the U1 state involved for Ubiquitin unfolding in sulphate solutions has higher free energy than the corresponding U1 state encountered under native conditions. The peculiar increase in the unfolded (U1) state free energy for Ubiquitin

unfolding with sulphate solution, clearly indicates that sulphate ions tend to increase the involved unfolding free energy barrier, not by stabilizing the native state of the protein but by destabilizing the unfolded (U1) states of the protein Ubiquitin, as opposed to the commonly used osmolytes and chemical chaperones which stabilizes the native state of the proteins. Nonetheless, our results clearly indicates the fact the sulphate anions stabilize the protein by increasing it's unfolding free energy through destabilization of the the unfolded (U1) states, contrary to some recent reports suggesting that the sulphate ions even though *kosmotropic* in nature, behaves in *chaotropic* way and results in destabilization of the native proteins.

5.3.3 Hydrophilic/Hydrophobic-water Interactions :

As illustrated in chapter-4, interaction between hydrophilic/hydrophobic patches on the protein and water could significantly dictate the stability/instability of proteins. Moreover, from figures-5.2 and 5.3, we know that the sulphate ions are involved in the direct interaction with the positively charged residues of the protein. Therefore, in order to address the observed instability of the unfolded (U1) state when compared with the U1 state under native conditions and perhaps to determine, whether the presence of sulphate ions and their direct interaction mechanism influences the solvation of hydrophilic/hydrophobic parts of the protein, we calculate the solvent accessible surface area (SASA) for hydrophilic/hydrophobic residues of the Ubiquitin along the unfolding pathway (MFEP) and compare it with the solvent accessible area for Ubiquitin unfolding under the native conditions.

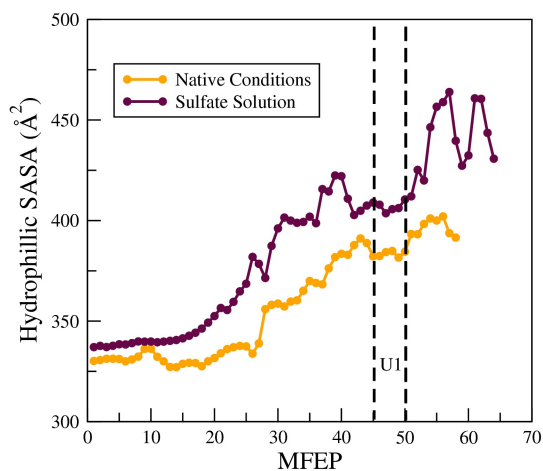


Figure 5.6: Hydrophilic SASA

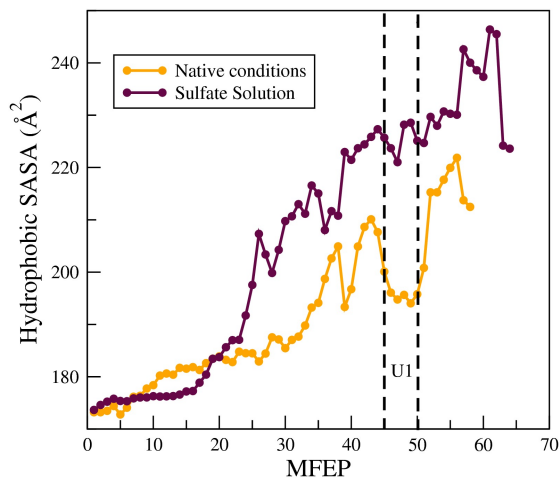


Figure 5.7: Hydrophobic SASA

The figure-5.6 shows the solvent accessible surface area for polar residues of the Ubiquitin, in native and in sulphate salts solution along an unfolding pathway. As is apparent from the figure, the hydrophilic surface area for Ubiquitin remains quite similar in the unfolded (U1) state of the protein, in both native as well as sulphate salt solution and is this likely to be less important when determining the stability of the U1 state ensemble of Ubiquitin. However, a glance at the figure-5.7 suggests that, the hydrophobic surface area in the U1 state differs quite substantially for sulphate solution and native conditions case.

Higher hydrophobic surface area implies that the hydrophobic parts of the protein are more exposed for interaction with solvent molecules. Since such an interaction, by definition is unfavourable in nature, it thus leads to the observed destabilisation of the unfolded (U1) state ensemble of the Ubiquitin thereby shifting the folding-unfolding equilibria on the folded side

of the protein and increasing the unfolding free energy barrier. At this stage, we think that sulphate ions exposes the hydrophobic parts of the Ubiquitin by destabilising the native salt bridges within Ubiquitin by forming the stronger and persistent salt bridges with the positively charged residues of the protein (figure-5.3). Such a behaviour, although can be true across all classes of wild-type proteins, remains to be verified and is a matter of the further study.

5.3.4 Mechanism of Unfolding of Ubiquitin :

Many studies have reported osmolytes or chemical chaperons significantly alters the unfolding dynamics of the protein^[3]. We previously studied the two step mechanism of unfolding of Ubiquitin under native conditions (chapter-4). In this section, we illustrate the sulphate ion induced changes in unfolding mechanism of the native Ubiquitin in comparison with the unfolding mechanism under native conditions. Following figure, summarizes the observed unfolding pathway along the lowest free energy path for each secondary structural element within the Ubiquitin. Note that, we follow the naming conventions introduced in chapter-4.

A glance on figure 5.8, indicates that the sulphate ions clearly assists in unfolding of the secondary structural fragments within native structure of the Ubiquitin molecule. A closer look at the figure 5.8(a), clearly indicates that the the α -helix becomes increasingly unstable in the simulation containing sulphate ions. α -helix within native Ubiquitin is comprised of $\approx 43\%$ of total charged residues of the protein. These charged residues are responsible for maintaining the 3D structure of the α -helix through salt-bridge interactions. We believe that, presence of sulphate ions weakens the involved salt-bridges by forming a stronger, more prevalent and persistent salt bridges with the positively charged residues on the helical region of the Ubiquitin, and could be direct consequence of the persistent interactions present (Figure 5.3), thus leading to the instability of the helical region of the Ubiquitin.

As is evident from the figure 5.8(b) that, the $\beta_1\beta_5$ follows the similar unfolding pattern in both cases, undergoing the substantial loss in hydrogen bonding interactions near the transition state (MFEP index 25-28) followed by a complete loss of the $\beta_1\beta_5$ sheet in the unfolded U1 region (MFEP index 45-50). On the contrary, $\beta_2\beta_1$ becomes exceedingly unstable in the sulphate solutions following the transition state of the protein as illustrated by the increased loss of hydrogen bond interactions (Figure 5.8(c)). Note that $\beta_2\beta_1$, however, do not completely unfold and there's still presence of some secondary structure in the U1 state. Moreover, figure 5.8(d) indicates that $\beta_1\beta_5$ sheet is clearly more stable in sulphate solution than in native conditions as suggested by the persistent hydrogen bonding interactions till MFEP index 58 and loses it's secondary structure only in U2 state (MFEP index 64-66) in case of sulphate simulations.

Feature-wise, $\beta_4\beta_3$ sheet is the smallest sheet amongst all the β sheets and undergoes a gradual unfolding under sulphate conditions as opposed to oscillatory behaviour observed in the case with native conditions(5.8(e)). Note that $\beta_4\beta_3$ has some remnant residual structure present in the U1 state and unfolds completely only in U2 state of the unfolding pathway of the protein. Remarkably, the only 3_{10} -helix present within native Ubiquitin gets heavily stabilized by the sulphate ions present in the solution as indicated by the presence of constant number of hydrogen bonds throughout the unfolding pathway for the Ubiquitin, as compared to it's counterpart wherein it unfolds within U1 state.

Note that, in all of cases discussed in the figure-5.8(a-f), there's a clear and an abrupt dip in the number of hydrogen bonds for the case of Ubiquitin unfolding with sulphate ion around the transition state region, while such a dip is largely absent in the simulation under native

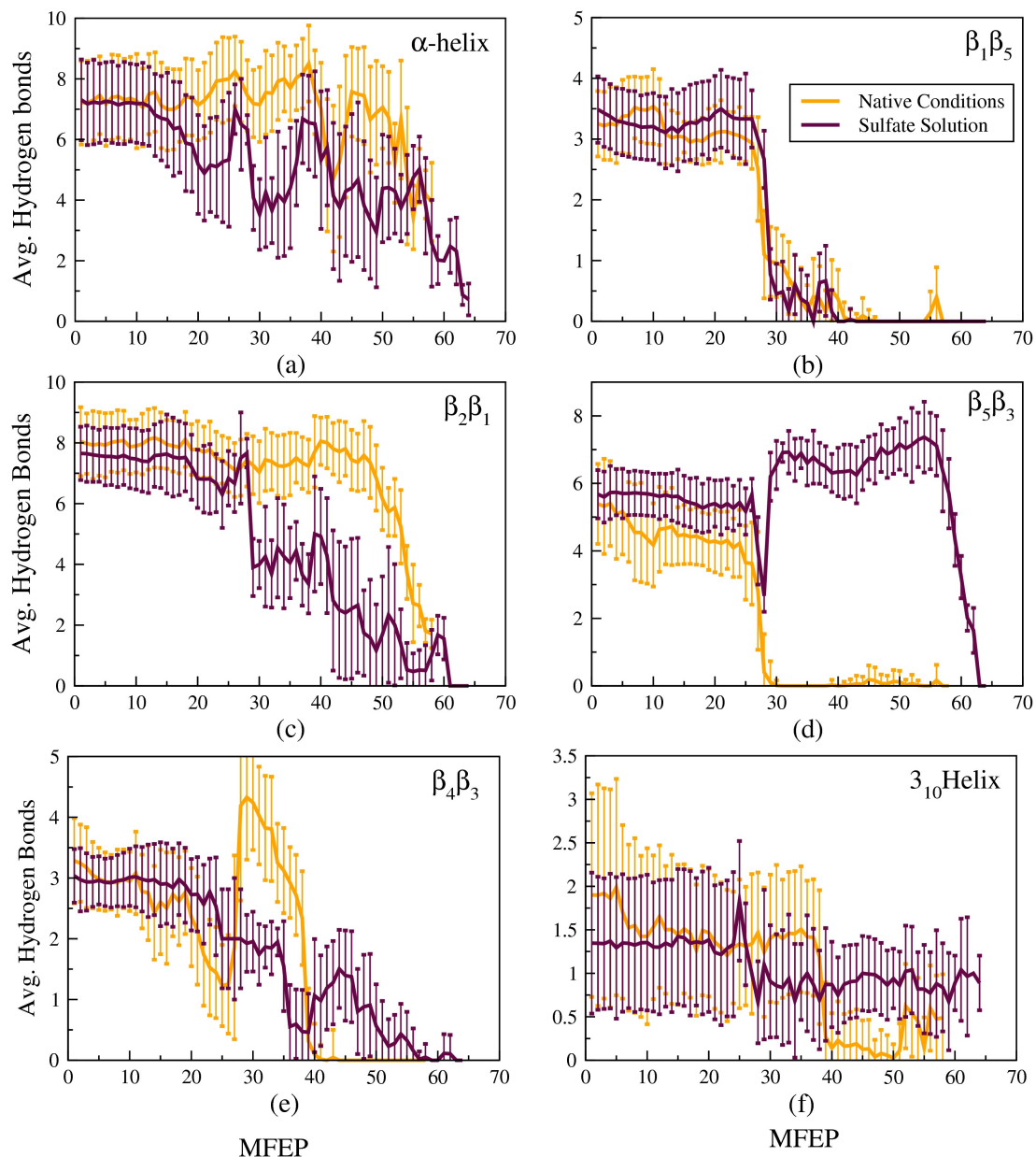


Figure 5.8: Figure representing unfolding pathways

conditions. We thus conclude that this sudden loss of hydrogen bonds within the secondary structural units of the protein is also one of the reasons behind the observed instability of the unfolded (U1) state in the sulfate simulation.

5.4 Outlook

To decipher the mechanism of action of the denaturing salts within the Hofmeister series, quiet similarly to the case of SO_4^{2-} anions, we are currently carrying out the simulation with ClO_4^- and SCN^- ions. Note that, despite being structurally analogous to the SO_4^{2-} ions, ClO_4^- occupies the position within the negative end of the Hofmeister series from the SO_4^{2-} ions. Both the anions considered, ClO_4^- and SCN^- , are chaotropic in nature and has been known to destabilize the native state of the protein. There's substantial amount of evidence in literature describing the instability of the protein in ClO_4^- and SCN^- salts arises out of their preferential interaction with protein backbone than the side chains of the protein. We here, aim to verify these conclusions and perhaps restructure Hofmeister series according the the effect of ions on the free energy of unfolding of Ubiquitin.

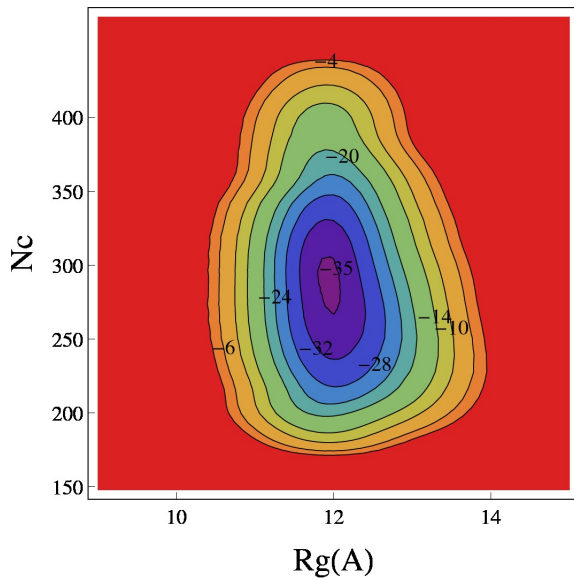


Figure 5.9: FES for ClO_4^- ions

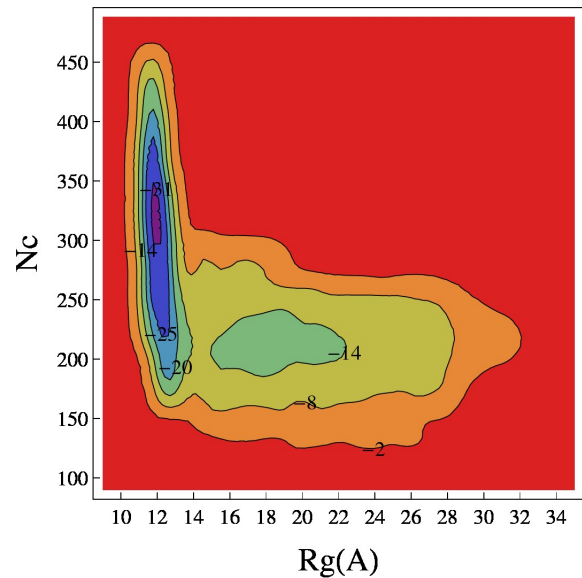


Figure 5.10: FES for SCN^- ions

The figure above represents the calculated free energy surface for Ubiquitin unfolding using both of the aforementioned ions. As can be seen from the figure, the native state is clearly unstable compared with case of native conditions or sulphate solution (figure 5.9). However, since the simulation is yet converge, it is practically wrong to draw any conclusions at this stage.

5.5 Conclusion

Our results indicate that the sulphate ions despite *kosmotropic* in nature, are involved in the direct-ion protein interactions that are persistent even on the time-scale of several nanoseconds and are well in agreement with the previous studies. Furthermore, we report for the first time that, the sulphate ions increases the unfolding free energy of the Ubiquitin by destabilizing the unfolded (U1) state of the protein, contrary to the more obvious and prevalent explanation describing the increased stability of the native state. By calculating the solvent accesible surface area for hydrophilic/hydrophobic parts of the Ubiquitin, we further illustrate that, such destabilization of the unfolded state ensemble arises because of the greater solvation of the hydrophobic parts of the Ubiquitin in sulphate solution as compared with native conditions. Moreover, we show that the sulphate ions drastically alter the observed two step mechanism in native conditions and almost all the native structures are lost only in the U2 state of the protein with the exception of $\beta_1\beta_5$ sheet.

Bibliography

- [1] Christopher M. Dobson. Protein folding and misfolding. *Nature*, 426 (6968):884–890, 2003.
- [2] Reman K. Singh, Neharika G. Chamachi, Suman Chakrabarty, and Arnab Mukherjee. Mechanism of unfolding of human prion protein. *J. Phys. Chem. B*, 121 (3):550–564, 2017.
- [3] David B. Kony Philippe H. Hünenberger Wilfred F. van Gunsteren. Molecular dynamics simulations of the native and partially folded states of ubiquitin: Influence of methanol cosolvent, ph, and temperature on the protein structure and dynamics. *Protein Science*, 16 (6):1101–1118, 2007.
- [4] Quinn Alexander Besford, Maoyuan Liu, and Angus Gray-Weale. Pair correlations that link the hydrophobic and hofmeister effects. *Phys. Chem. Chem. Phys*, 18:14949–14959, 2016.
- [5] Collins KD. Ions from the hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods*, 34:300–311, 2004.
- [6] Jordan W. Bye and Robert J. Falconer. Three stages of lysozyme thermal stabilization by high and medium charge density anions. *J. Phys. Chem. B*, 118:4282–4286, 2014.
- [7] Jana Paterova, Kelvin B. Rembert, Jan Heyda, Yadagiri Kurra, Halil I. Okur, Wenshe R. Liu, Christian Hilty, Paul S. Cremer, , and Pavel Jungwirth. Reversal of the hofmeister series: Specific ion effects on peptides. *J. Phys. Chem. B*, 117:8150–8158, 2013.
- [8] Soohaeng Yoo Willow and Sotiris S. Xantheas. Molecular-level insight of the effect of hofmeister anions on the interfacial surface tension of a model protein. *J. Phys. Chem. Lett*, 8:1574–1577, 2017.
- [9] P. W. Rose, B. Beran, C. Bi, W. F. Bluhm, and Dimitropoulos. The rcsb protein data bank:redesigned web site and web services. *Nucleic Acids Res.*, D392-D401:2011, 39.
- [10] N. Deshpande, K. J. Address, W. F. Bluhm, J. C. Merino-Ott, W. Townsend-Merino, Q. Zhang, C. Knezevich, L. Xie, L. Chen, and Z. Feng. The rcsb protein data bank: A redesigned query system and relational database based on the mmcif schema. *Nucleic Acids Res.*, 33:D233–D237, 2005.
- [11] Vijay-Kumar Senadhi and Charles E.Bugg abd William J. Cook. Structure of ubiquitin refined at 1.8Å resolution. *J. Mol. Biol.*, 194 (3):531–544, 1987.
- [12] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, J. C. Shirts, M. R.and Smith, P. M. Kasson, and al. vander Spoel, D. et. Gromacs 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29:845–54, 2013.
- [13] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, and Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, 65:712–725, 2006.
- [14] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

-
- [15] In Suk Joung and Thomas E. Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112:9020–9041, 2008.
- [16] Sadra Kashefolgheta and Ana Vila Verde. Developing force fields when experimental data is sparse: AMBER/GAFF-compatible parameters for inorganic and alkyl oxoanions. *Phys. Chem. Chem. Phys.*, 19:20593–20607, 2017.
- [17] M. Baaden, F. Berny, G. Wipff, and C. Madic. A molecular dynamics and quantum mechanics study of m3+ lanthanide cation solvation by acetonitrile: the role of cation size, counterions and polarization effects investigated. *J. Phys. Chem. A*, 104:7659, 2000.
- [18] A. Vincze, P. Jedlovszky, and G. Horvai. The l/l interface and adsorption of SCN⁻ anions as studied by different molecular simulation techniques. *Analytical Sciences/Supplements*, 17icas:i317–i320, 2002.
- [19] W. H. Press. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge university press, 2007.
- [20] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [21] M. Parrinello and A Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys*, 52:7182–7190, 1981.
- [22] S. I. NosÉ. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 100:191–198, 2002.
- [23] T. Darden, D. York, and L. J. Pedersen. Particle mesh ewald: An n · log (n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.
- [24] Berk Hess, Henk Bekker, Herman J.C.Berendsen, and Johannes G.E.M.Fraaijey. Lincs:a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18:1463–1472, 1997.
- [25] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, 2008.
- [26] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, and R. A. Broglia. Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, 180:1961–1972, 2009.
- [27] B. Ensing, A. Laio, M. Parrinello, and M. L. A Klein. Recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B*, 109:6676–6687, 2005.
- [28] Gokarn YR, Fesinmeyer RM, Saluja A, Razinkov V, Chase SF, and Laue TM et al. Effective charge measurements reveal selective and preferential accumulation of anions, but not cations, at the protein surface in dilute salt solutions. *Protein. Sci*, 20:580–587, 2011.
-

Appendix A

The sulphate ion, as previously mentioned is *kosmotropic* in nature and thus possesses the large hydration shell. Therefore, to get the idea of length of the first hydration shell of the sulphate anions, we've calculated the radial distribution function between sulphate ions and water. Figure below depicts the calculated radial distribution function between sulphur (S) of SO_4^{2-} ions and water oxygen (OW).

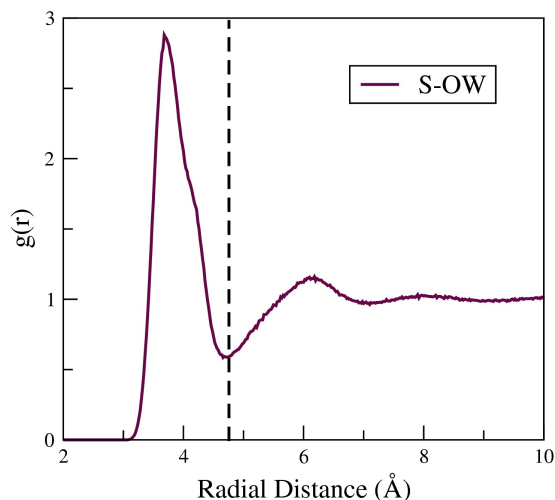


Figure 11: Radial distribution function between sulphur (S) of SO_4^{2-} ions and water oxygen (OW)

The radial distribution function, as the name suggests, is an indicator of the variation of the density of atoms/molecules from the reference atom/molecule radially or in simpler terms it refers to the probability of finding an atom/molecule at a distance r from the the reference atom. Thus, the first minimum in the radial distribution function corresponds to the first hydration shell (in case of molecule-water rdf) while the second minimum corresponds to the second hydration shell of the atom/molecule and so on. As can be seen from the figure-11, the first minimum in the sulphate-water rdf lies around $r \approx 5 \text{\AA}$ which tells us that, the first hydration shell of sulphate anions lies around $\approx 5 \text{\AA}$.

Note that, the sulphate-water distribution function shown in the figure-11 is calculated from the SIM7 owing to the fact that only the SIM7 does not led to the aggregation of sulphate ions. Moreover, the calculated rdf matches well with previous studies done using the polarizable force fields for sulphate and water and thus emphasizes the accuracy of the sulphate potentials used in SIM7.