

ASSESSMENT OF LOCAL HYDROPHOBICITY AND ITS EFFECT IN
MEDIATING PROTEIN RELATED ASSOCIATIONS

by
Di Cui

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry and Biochemistry

Summer 2015

© 2015 Di Cui
All Rights Reserved

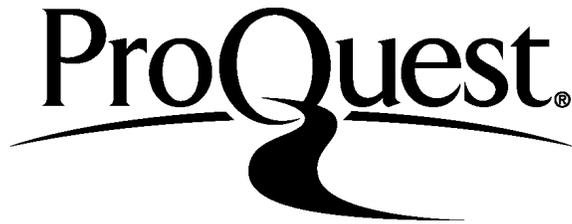
ProQuest Number: 3730249

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 3730249

Published by ProQuest LLC (2015). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ASSESSMENT OF LOCAL HYDROPHOBICITY AND ITS EFFECT IN
MEDIATING PROTEIN RELATED ASSOCIATIONS

by

Di Cui

Approved: _____
Murray V. Johnston, Ph.D.
Chair of the Department of Chemistry and Biochemistry

Approved: _____
George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____
James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____
Sandeep Patel, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____
Catherine L. Grimes, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____
Zhihao Zhuang, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____
Edward R. Lyman, Ph.D.
Member of dissertation committee

ACKNOWLEDGEMENTS

As a starting point for the dissertation, I would like to express my upmost thankfulness to my advisor of Ph.D. program, Dr. Sandeep Patel. In my four years study as his graduate student, he is like a father to guide me progressing in my academic. It is Dr. Sandeep Patel who initiated my interest in science; it is Dr. Sandeep Patel who guided me to become a proper Ph.D. that should have independent thinking ability; it is Dr. Sandeep Patel who enlightened me to think, talk and perform like a scientist. What I learn from him is not only knowledge on statistic thermodynamics and molecular dynamic simulation, but also an attitude toward career and success. I will always remember his instruction which appear in the gateway of our cluster. "Success is easy: JUST DO IT!" and "With any endeavor in life, to have success, you have to own it." I learn from my advisor not only from his talk, but also from his doing. Every night, he will work with us on the cluster. I could always communicate with him about my research in the first time. It makes me feel that I was not alone in the path of scientific research and knowledge discovery.

The people who next come to my mind to thank are my father, Lei Cui and my mother, Yanfeng Qi. They are the most important persons to me, who love me and support me without any reasons. I will always love them just like they love me. I could also feel kinship care from my other family members. Here, I would also like to express my gratitude to them: Yan Cui, Wenzhe Qi, Yanling Qi, Yanjie Qi, Yanhong Qi, Bo Cui, Zhiyao Zhang, Qiming Shi, Yiming Zhao, Xin Qi, Yunqing Cui, Guolan Liu, Hong Cui and Zhenduo Qi. It always makes me warm as talking to these family members. Specifically, I would like to thank my grandmother, Shulan Sui, who has passed away 10 years ago. I want to tell her that your young grandson has grown up. I will always miss her.

No one can succeed without the help of others. Besides my advisor, I would like to thank some helpful faculty members. I would like to thank Dr. Edward R. Lyman, Dr. Zhihao Zhuang and Dr. Catherine L. Grimes for serving me as committee members and giving their insightful comments on my research. I also appreciate the help from my colleagues in Dr. Patel's group: Dr. Shuching Ou, Dr. Sudipta Kumar Sinha, Dr. Yuan Hu, Hongbin Wang and Eric Peters. It is a great pleasure to work with these decent and intelligent people. It is a good memory of me to discuss with them about research.

Finally, let me express my thankfulness to Meng Zhang. Thank you for encouraging me to move forward in my life. Ernest Hemingway once wrote, "The world is a fine place and worth fighting for."

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS AND CONSTANTS	xxii
ABSTRACT	xxvi

Chapter

1 INTRODUCTION AND MOTIVATION	1
1.1 Introduction	1
1.1.1 Hydrophobic Effect	1
1.1.2 Protein Hydrophobicity	2
1.1.3 Ion-Specificity around Protein Surface	7
1.1.4 Denaturants around Protein Surface	9
1.1.5 Protein-Protein Interactions Mediated by Hydrophobic Effect	11
1.2 Objective	13
2 FORCE FIELDS AND COMPUTATIONAL METHODS	19
2.1 Molecular Dynamics Simulation and Force Fields	19
2.2 Radial Distribution Function	21
2.3 Density Profiles	22
2.4 Surface Tension	23
2.5 Vapor Pressure	24
2.6 Dipole Moment Profile	25
2.7 Interfacial Potential	26
2.8 Free Energy and Potential of Mean Force	27
2.8.1 Adaptive Biasing Force	28
2.8.2 Umbrella Sampling	29
2.9 Instantaneous Protein Interface and Interface Fluctuations	30

3	EVALUATION OF EFFECTIVE HYDROPHOBICITY AROUND PROTEIN SURFACE AND IMPLICATIONS FOR BINDING PATCH THROUGH HYDROPHOBIC EFFECTS	32
3.1	Introduction	32
3.2	Materials and Method	36
3.2.1	Selection of Proteins	36
3.2.2	Simulation Details	38
3.3	Results	41
3.3.1	Water Networks around Ubiquitin	41
3.3.2	Surface Water Density around Ubiquitin	43
3.3.3	Cluster Analysis to Identify the Effective Hydrophobic Interface	46
3.3.4	Another Example: HFBII	51
3.3.5	Applications to Other Proteins	59
3.4	Summary	67
4	ION-SPECIFICITY AROUND EFFECTIVE HYDROPHOBIC REGIONS OF PROTEIN SURFACE	72
4.1	Introduction	72
4.2	Method	77
4.2.1	Simulation Details	77
4.2.1.1	Umbrella Sampling Potential of Mean Force Calculations: Ion Translocation Across Aqueous Liquid-vapor Interface	77
4.2.1.2	Protein in KCl/KI Aqueous Solution	81
4.2.1.3	Aqueous Protein Interfaces	82
4.3	Result and Discussion	88
4.3.1	Liquid-Vapor Interface	88
4.3.2	Ion Distributions Around Protein in 1 Molal Aqueous Environment	92
4.3.3	Potential of Mean Force	93

4.3.4	Less Hydrophobic and Hydrophilic Protein Interface	110
4.4	Summary and Conclusions	117
5	ORIENTATIONAL PREFERENCE OF GUANIDINIUM CATION AND UREA DENATURANTS AROUND EFFECTIVE HYDROPHOBIC REGIONS OF PROTEIN SURFACE	120
5.1	Introduction	120
5.2	Method	124
5.3	Result and Discussion	130
5.3.1	Liquid-Vapor Interface	130
5.3.2	Aqueous Protein Interface	136
5.4	Summary and Conclusions	154
6	A BIOCHEMICAL MODEL FOR BINDING MEDIATED BY HYDROPHOBIC INTERACTION - ASSOCIATION BETWEEN UBIQUITIN AND UBIQUITIN BINDING DOMAINS	162
6.1	Introduction	162
6.2	Materials and Methods	166
6.2.1	Simulation Details	166
6.2.2	Potential of Mean Force	170
6.3	Results and Discussion	174
6.3.1	Free Energy, Enthalpy and Entropy Changes	174
6.3.2	Interaction Energies from Different Components	180
6.3.3	System Component Contributions to Potential of Mean Force	181
6.3.4	Water Density	186
6.3.5	Free Energetics and UIM Helix Orientation Relative to Ubiquitin	189
6.3.6	Entropy Analysis of Protein	194
6.4	Conclusion	196
7	CONCLUSION	200
7.1	Summary of Key Results	200
7.2	Future Work	204

BIBLIOGRAPHY	208
------------------------	-----

Appendix

A CLUSTER ANALYSIS OF LOW-HYDRATED RESIDUES OF VARIOUS OF PROTEINS	230
B PERMISSIONS	245

LIST OF TABLES

3.1	Comparison of identified effective hydrophobic regions based on surface water density and based on water number fluctuation. . . .	60
3.2	Critical hydration level to identify effective hydrophobic regions for each protein.	64
3.3	Identified effective hydrophobic patches for various kinds of proteins.	65
3.4	Coverage and accuracy of various approaches of predictions. Based on the reference, [1] the coverage is defined as $Cov = TP/RI$ and the accuracy is defined as $Acc = TP/(TP+FP)$. TP is the number of true prediction; RI is the number of real residues involving in the hydrophobic association according to the literature; FP is the number of false positives in the prediction.	68
4.1	LJ parameters for ions applied in this work and verification. Note: for the case of single water-anion binding case, there are two geometries for the binding structure, one is C_S , another is C_{2v} ; the binding distances and binding energies for Cl^- and I^- shown here are from C_S geometry.	79
4.2	NAMD input parameters for the simulations.	83
4.3	Angle between positive z vector and the line connecting central position of the patch (0,0,12) with each of the heavy atom position on the patch.	86
6.1	NAMD input parameters for the simulations.	170
6.2	Duration of the simulated trajectory of various orientations. . . .	174
6.3	Association free energies for ubiquitin with UIM at different orientations.	191

6.4	Enthalpic contribution and entropic contribution for different orientations.	194
6.5	Fitting constants for entropy extrapolation. The numbers in the brackets represent asymptotic standard errors upon fitting.	196
6.6	Decoupled entropies from translational, rotational and vibrational contributions.	196

LIST OF FIGURES

3.1	(A) Representative snapshot of ubiquitin solvated by 600 water molecules (B) Representative snapshot of ubiquitin solvated by 600 water molecules (rotated by 180° of Panel A).	40
3.2	(A) Distribution of distance between water oxygen and the nearest heavy atom of protein ubiquitin (B) Distribution of distance between water oxygen and the nearest heavy atom of protein HFBII.	42
3.3	(A) Arrangement of water molecules on the surface of ubiquitin solvated by 600 water molecules. The water molecules that belong to the largest cluster are colored in blue; those of all other water molecules in red (B) Evolution of the largest water cluster size (C)-(H) Probability distribution $P(S_{max})$ for the largest cluster size S_{max} of the water molecules around ubiquitin protein surface at various hydration levels from $N_w = 400$ to $N_w = 900$	44
3.4	Surface area water density around each amino acid residue of ubiquitin protein at hydration level of $N_w = 600$	47
3.5	Water density map around ubiquitin with $N_w = 600$. (A) and (B) represent two side of the protein respectively with a rotation of 180 °. Red colour indicates a lower water density; blue colour indicates a higher water density.	48
3.6	(A) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 500$ (B) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 600$ (C) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 700$	52
3.7	(A) Representative of the cluster of low-hydrated residues of ubiquitin with $N_w = 500$ (B) Representative of the cluster of low-hydrated residues of ubiquitin with $N_w = 600$ (C) Representative of the cluster of low-hydrated residues of ubiquitin with $N_w = 700$ (D) Representative of the reported binding patch for ubiquitin. In all cases, the residues involved are shown in blue colour.	53

3.8	(A)-(E) Probability distribution n_S of water clusters with size S around ubiquitin at various levels from $N_w = 400$ to $N_w = 800$ (F) Fisher exponent (from fit of the cluster size distribution graph) τ of ubiquitin at various hydration levels.	54
3.9	(A) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 620$ (B) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 640$ (C) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 660$ (D) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 680$	55
3.10	Probability distribution $P(S_{max})$ for the largest cluster size S_{max} of the water molecules around HFBII protein surface at various hydration levels. (A) $N_w = 400$ (B) $N_w = 500$ (C) $N_w = 600$ (D) $N_w = 700$ (E) $N_w = 800$ (F) $N_w = 900$	56
3.11	Cluster analysis of low-hydrated residues of HFBII at critical hydration level.	57
3.12	Water density map around HFBII with $N_w = 500$. (A) and (B) represent two side of the protein respectively with a rotation of 180° . Red colour indicates a lower water density; blue colour indicates a higher water density.	58
3.13	Map based on the fluctuation of number of water around ubiquitin surface at critical hydration level.	61
3.14	(A) Water density map around protein Ubl-domain of HHR23B at critical hydration level $N_w = 600$ with long tail around C-terminal (B) Water density map around protein Ubl-domain of HHR23B at critical hydration level $N_w = 600$ without long tail around C-terminal. . .	63
3.15	Prediction coverage vs prediction accuracy of different approaches.	69
4.1	Ion-water radial distribution function (RDF) for anions in TIP3P.	80

4.2	(A) Representative snapshots of the system used in the study (B) representation of the hydrophobic interface defined in this study. Including residues L7, V18, L19, L21, I22, V24, V54, A61, L62, L63. Dash orange line roughly select the region of interest (C) representation of the less hydrophobic interface defined in this study. Including residues I31, A32, D34, I38, A41, H42, S45. (D) representation of the hydrophilic interface defined in this study. Including residues D25, C26, K27, T28, A58, D59, Q60.	85
4.3	(A) PMF of single Cl^-/I^- approaching the liquid-vapor interface in TIP3P water (B) Normalized liquid-vapor interface fluctuation at ($x = 0, y = 0$) as a function of anion restrained position for Cl^- and I^- .	90
4.4	Number density distribution of Cl^-/I^- around the hydrophobic interface of HFBII in 1.0 m KCl/KI aqueous solution. (A) Cl^- density distribution (B) I^- density distribution. X axis represents the lateral distance $r = \sqrt{x^2 + y^2}$. We define $r > 0$ means the signs of x component and y component are the same; while $r < 0$ means the signs of x component and y component are different.	94
4.5	Number of bins that display Cl^-/I^- densities above certain threshold values around the hydrophobic patch of HFBII in 1.0 m KCl/KI aqueous solution. (A) above threshold value 3 (B) above threshold value 4 (C) above threshold value 5 (D) above threshold value 6 (E) above threshold value 7 (F) above threshold value 8.	95
4.6	(A) PMF for single Cl^-/I^- approaching the hydrophobic protein-solvent interface (B) Coordinate water numbers around single Cl^-/I^- as a function of the reaction coordinate.	97
4.7	(A) Evolution of PMF for Cl^- approaching hydrophobic interface (B) Evolution of PMF for I^- approaching hydrophobic interface (C) Evolution of PMF for Cl^- approaching less hydrophobic interface (D) Evolution of PMF for I^- approaching less hydrophobic interface (E) Evolution of PMF for Cl^- approaching hydrophilic interface (F) Evolution of PMF for I^- approaching hydrophilic interface.	98
4.8	(A) PMF for single Cl^-/I^- approaching the hydrophobic protein-solvent interface with fixed protein and fixed anion (B) Hydrophobic protein-solvent interface fluctuation at ($x = 0, y = 0$) as a function of anion restrained position for Cl^- and I^- with fixed protein and fixed anion.	99

4.9	Protein-solvent mean interface $\langle h(x, y) \rangle$ and interface fluctuations $\langle \delta h^2(x, y) \rangle$ in single Cl^-/I^- solution. The color scale represents the interface fluctuations (A) Cl^- resides at $z = 24 \text{ \AA}$ (B) I^- resides at $z = 24 \text{ \AA}$ (C) Cl^- resides at $z = 18 \text{ \AA}$ (D) I^- resides at $z = 18 \text{ \AA}$ (E) Cl^- resides at $z = 16 \text{ \AA}$ (F) I^- resides at $z = 16 \text{ \AA}$	102
4.10	(A) Average aqueous protein surface fluctuation $\langle \delta h^2 \rangle$ as a function of θ and ϕ (A) Cl^- locates at $z = 18 \text{ \AA}$ (B) I^- locates at $z = 18 \text{ \AA}$	103
4.11	Average water oxygen density around (A) Cl^- at position $z = 18 \text{ \AA}$ (B) I^- at position $z = 18 \text{ \AA}$ (C) Cl^- at position $z = 19 \text{ \AA}$ (D) I^- at position $z = 19 \text{ \AA}$. X axis represents the lateral distance $r = \sqrt{x^2 + y^2}$ and Y axis represents the distance from positive z direction.	105
4.12	(A) Hydrophobic interface fluctuation at $(x = 0, y = 0)$ as a function of anion restrained position for Cl^- and I^- (B) Normalized interface fluctuation at $(x = 0, y = 0)$ as a function of anion restrained position for Cl^- and I^-	107
4.13	Differences between ion position and surface position (surface position is defined by the surface height at position $x = 0, y = 0$) for I^- and Cl^- at various positions from 16.5 (black), 18 (red), 19 (green) and 22 (blue).	108
4.14	Differences between ion position and surface position (surface position is defined by the surface height around position $x = 3, y = 0.6$) for I^- and Cl^- at various positions from 16.5 (black), 18 (red), 19 (green) and 22 (blue).	109
4.15	Evolution of protein backbone RMSD as (A) Single Cl^- locates at $z = 16 \text{ \AA}$ (B) Single I^- locates at $z = 16 \text{ \AA}$ (C) Single Cl^- locates at $z = 25 \text{ \AA}$ (D) Single I^- locates at $z = 25 \text{ \AA}$	111
4.16	Hydrophobic interface fluctuation at $(x = 0, y = 0)$ as a function of anion Z-position for Cl^- and I^- in the case of flexible protein.	112
4.17	(A) PMF for single Cl^-/I^- approaching the less hydrophobic protein-solvent interfaces (B) PMF for single Cl^-/I^- approaching the hydrophilic protein-solvent interfaces.	114

4.18	Inherent interface fluctuations of HFBII. For A, B, C and D, each one depicts one side of the protein interface with a rotation of 90° respectively. Red colour represents larger fluctuations, while blue colour represents smaller fluctuations. The highlight regions in A, C and D corresponds to the hydrophobic, less hydrophobic and hydrophilic regions that we define in this study.	116
4.19	(A) Less hydrophobic interface fluctuation at $(x = 0, y = 0)$ as a function of anion restrained position for Cl^- and I^- (B) Hydrophilic interface fluctuation at $(x = 0, y = 0)$ as a function of anion restrained position for Cl^- and I^-	118
5.1	(A) Representative snapshot of single Gdm^+ with parallel orientation to the liquid-vapor interface (B) Single Gdm^+ with perpendicular orientation to the liquid-vapor interface (C) HFBII protein in 1.0 molal concentration of GdmCl aqueous solution (D) Single Gdm^+ with parallel orientation to the HFBII protein-solvent interface (E) Single Gdm^+ with perpendicular y orientation to the HFBII protein-solvent interface (F) Single Gdm^+ with perpendicular x orientation to the HFBII protein-solvent interface.	127
5.2	(A) Orientationally resolved probability map of single Gdm^+ around liquid-vapor interface (B) PMF of single Gdm^+ from bulk transporting through liquid-vapor interface with parallel orientation, perpendicular orientation and no orientational restraint (C) Orientationally resolved probability map of single urea around liquid-vapor interface (D) PMF of single urea from bulk transporting through liquid-vapor interface with parallel orientation, perpendicular orientation and no orientational restraint. For clarity, in (B) and (D), no orientational restraint profiles are shifted by 1 kcal/mol; perpendicular profiles are shifted by 2 kcal/mol. The GDS positions are denoted as orange dash lines in (B) and (D).	133
5.3	(A) Surface height fluctuation for liquid-vapor interface at $(x = 0, y = 0)$ as a function of position of single Gdm^+ (B) Normalized surface height fluctuation for liquid-vapor interface at $(x = 0, y = 0)$ as a function of position of single Gdm^+ (C) Surface height fluctuation for liquid-vapor interface at $(x = 0, y = 0)$ as a function of position of single urea (D) Normalized surface height fluctuation for liquid-vapor interface at $(x = 0, y = 0)$ as a function of position of single urea.	137

5.4	(A) Representative of sampling volume for probing orientational resolved probability of solute around certain region of protein interface (B) Orientational resolved probability distribution of Gdm^+ around hydrophobic protein interface in 1.0 molal GdmCl solution (C) Orientational resolved probability distribution of urea around hydrophobic protein interface in 1.0 molal urea solution (D) Orientational resolved probability distribution of Gdm^+ around hydrophilic protein interface in 1.0 molal GdmCl solution.	140
5.5	(A) Representative of sampling volume for probing orientational resolved probability of solute around certain region of protein interface with the range of $-6 \text{ \AA} \leq x \leq 6 \text{ \AA}$, $-6 \text{ \AA} \leq y \leq 6 \text{ \AA}$ and $12 \text{ \AA} \leq z \leq 25 \text{ \AA}$ (B) Orientational resolved probability distribution of Gdm^+ around hydrophobic protein interface in 1.0 molal GdmCl solution (C) Orientational resolved probability distribution of urea around hydrophobic protein interface in 1.0 molal urea solution (D) Orientational resolved probability distribution of Gdm^+ around hydrophilic protein interface in 1.0 molal GdmCl solution.	141
5.6	(A) Representative of sampling volume for probing orientational resolved probability of solute around certain region of protein interface with the range of $-7 \text{ \AA} \leq x \leq 7 \text{ \AA}$, $-7 \text{ \AA} \leq y \leq 7 \text{ \AA}$ and $12 \text{ \AA} \leq z \leq 25 \text{ \AA}$ (B) Orientational resolved probability distribution of Gdm^+ around hydrophobic protein interface in 1.0 molal GdmCl solution (C) Orientational resolved probability distribution of urea around hydrophobic protein interface in 1.0 molal urea solution (D) Orientational resolved probability distribution of Gdm^+ around hydrophilic protein interface in 1.0 molal GdmCl solution.	142
5.7	(A) PMF for single Gdm^+ with parallel orientation, perpendicular y and perpendicular x orientation from bulk approaching the hydrophobic protein-solvent interface (B) PMF for single urea with parallel orientation, perpendicular y and perpendicular x orientation from bulk approaching the hydrophobic protein-solvent interface. .	144
5.8	(A) RDFs between central carbon of Gdm^+ and oxygen atoms of water molecules (B) Coordinate water numbers within the first hydration shell of Gdm^+ restraining at different orientations as a function of Z position of central carbon of Gdm^+	145
5.9	PMF profiles for Gdm^+ approaching another Gdm^+ with parallel orientation (red), perpendicular orientation (blue) and no orientational restraint (green).	146

5.10	(A) Mean height of liquid-vapor interface as single Gdm ⁺ locates at z = 14 Å (B) Surface height fluctuation of liquid-vapor system as single Gdm ⁺ locates at z = 14 Å (C) Mean protein-solvent interface height as single Gdm ⁺ locates at z = 18 Å (D) Height fluctuation of protein-solvent interface as single Gdm ⁺ locates at z = 18 Å, the color scales representing the magnitude of fluctuation.	149
5.11	(A) Surface height fluctuation for hydrophobic protein interface at (x = 0, y = 0) as a function of restrain z position of single Gdm ⁺ with parallel orientation, perpendicular y and perpendicular x orientation (B) Surface height fluctuation for hydrophobic protein interface at (x = 0, y = 0) as a function of restrain z position of single urea with parallel orientation, perpendicular y and perpendicular x orientation.	150
5.12	(A) Probability distributions of Δ for Gdm ⁺ with parallel restrained orientation locating at various positions close to the peak of largest fluctuation. (B) Probability distributions (log scale) of Δ for Gdm ⁺ with parallel restrained orientation locating at various positions close to the peak of largest fluctuation.	153
5.13	(A) Gdm ⁺ with perpendicular y orientation (orange) around the hydrophobic protein patch (B) Gdm ⁺ with perpendicular x orientation (orange) around the hydrophobic protein patch, residue I22 and L63 are shown in green (C) Gdm ⁺ with perpendicular y orientation (orange) around the hydrophilic protein patch, residues D25 is shown in green.	155
5.14	(A)-(C) Differences between single Gdm ⁺ position and surface position (surface position is defined by the surface height at position x = 0, y = 0) Δ for Gdm ⁺ with different restrained orientations locating at various positions, (A) parallel orientation (B) perpendicular y orientation (C) perpendicular x orientation (D)-(F) Probability distributions of Δ for Gdm ⁺ with different restrained orientations locating at various positions, (D) parallel orientation (E) perpendicular y orientation (F) perpendicular x orientation.	156
5.15	(A) PMF for single Gdm ⁺ with parallel orientation, perpendicular y and perpendicular x orientation from bulk approaching the hydrophilic protein-solvent interface (B) Surface height fluctuation for hydrophilic protein interface at (x = 0, y = 0) as a function of restrain z position of single Gdm ⁺ with parallel orientation, perpendicular y and perpendicular x orientation.	157

5.16	(A) Gdm ⁺ number density map around HFBII protein (hydrophobic side). Blue represents higher number density, while red represents lower number density (B) Gdm ⁺ number density map around HFBII protein (opposite side) (C) Representation of hydrophobic protein patch of HFBII with orange highlighting each hydrophobic residues on the patch.	159
6.1	Representative snapshots of the system used in the study (A) representation of the hydrophobic side of the UIM helix. Non-polar residues, white; basic residues, blue; acidic residues, red; uncharged hydrophilic residues, green (B) representation of the hydrophilic side of the UIM helix (c) representation of the ubiquitin and UIM binding. The distances change from 15 Å to 34 Å. The orientations change from -150° to 180°.	168
6.2	Evolution of protein backbone RMSD (A) Ubiquitin-UIM complex with 0° orientation (B) Ubiquitin-UIM complex with -30° orientation (C) UIM (D) Ubiquitin.	169
6.3	Number density profiles of water (A) associated state window [15:16] using timestep 0.8 fs (B) associated state window [15:16] using timestep 1.0 fs (C) separated state window [32:33] using timestep 0.8 fs (D) separated state window [32:33] using timestep 1.0 fs.	171
6.4	Evolution of z component of the center of mass of ubiquitin (A) associated state window [15:16] (B) separated state window [32:33].	172
6.5	Evolution of the free energy for different orientations (A) 0° orientation (B) -30° orientation (C) -60° orientation (D) -90° orientation.	176
6.6	Thermodynamic contributions along the reaction coordinate. Relative Gibbs free energy, G(red), enthalpy, H(blue) and entropic term, -TS (green), and their uncertainties (vertical bars).	178
6.7	Decomposed interaction energies: (A) protein-water H _{PW} (B) domain-water H _{DW} (C) water-water H _{WW} (D) protein-domain H _{PD} .	182

6.8	(A) Average separation between center of mass of R74 on ubiquitin and center of mass of E5 on the UIM as a function of reaction coordinate (B) Interaction energy between R74 and E5 as a function of reaction coordinate. (C) Interaction energy between R42, R72, R74 on the ubiquitin and E3, E5, E6, E7 on the UIM as a function of reaction coordinate (D) Interaction energy between R42, R72, R74 on the ubiquitin and all the residues on the UIM. In all cases, the VDW component is shifted by 40 kcal/mol and the electrostatic component is shifted by -40 kcal/mol for clarity.	183
6.9	Comparison of interaction energies between different components using PME and different non-bonded cutoff values (A) Protein-domain interaction energy (B) Domain-water interaction energy (C) Protein-water interaction energy.	184
6.10	PMF contribution from (A) protein (B) water.	186
6.11	Water density distribution maps for a select slice along the z direction (A) representation of the select slice to consider water density (B) water density distribution map for $z_0 = 0 \text{ \AA}$ at window [32:33] (C) water density distribution map for $z_0 = 32 \text{ \AA}$ at window [32:33]. The density is normalized such that $\rho = 1$ corresponds to the bulk water density of 0.0334 \AA^{-3}	190
6.12	Free energies of restrained UIM with different orientations association with ubiquitin. (A) Mapping the free energies of association at various orientations and separations; (B),(C),(D),(E) Free energy profiles for all the orientations. For clarity, red lines are shifted by 10 kcal/mol; green lines are shifted by 20 kcal/mol.	192
6.13	Representation of hydrophobic residues involved in ubiquitin-UIM binding (A) UIM restrained at 0° orientation (B) UIM restrained at -120° orientation. Color scheme: red, hydrophobic residues L8, I44, H68 and V70 on ubiquitin; blue, hydrophobic residues L8, I9, A12, I13, L15 on UIM.	193
6.14	Evolution of vibrational entropy for (A) Ubiquitin-UIM complex with 0° orientation (B) Ubiquitin-UIM complex with -30° orientation (C) UIM (D) Ubiquitin.	197
A.1	Cluster analysis of low-hydrated residues of CUE domain	231
A.2	Cluster analysis of low-hydrated residues of UBA of DSK2	232

A.3	Cluster analysis of low-hydrated residues of GGA3 GAT domain . .	233
A.4	Cluster analysis of low-hydrated residues of UBA of Human BMSC-Ubp	234
A.5	Cluster analysis of low-hydrated residues of Ubl-domain of HHR23A	235
A.6	Cluster analysis of low-hydrated residues of Ubl-domain of HHR23B	236
A.7	Cluster analysis of low-hydrated residues of NEDD8	237
A.8	Cluster analysis of low-hydrated residues of Pinch-1 LIM4 domain .	238
A.9	Cluster analysis of low-hydrated residues of Sla1 SH3-3 domain . .	239
A.10	Cluster analysis of low-hydrated residues of Nck-2 SH3 domain . . .	240
A.11	Cluster analysis of low-hydrated residues of CIN85 SH3-3 domain .	241
A.12	Cluster analysis of low-hydrated residues of Crk SH2 domain	242
A.13	Cluster analysis of low-hydrated residues of Abl SH3 domain	243
A.14	Cluster analysis of low-hydrated residues of HPR	244

LIST OF ABBREVIATIONS AND CONSTANTS

ABF	Adaptive Biasing Force
ABP	Arabinose Binding Protein
ACC	Accuracy
ALA	Alanine, 1 Letter Code: A
AMBER	Assisted Model Building with Energy Refinement
ARG	Arginine, 1 Letter Code: R
ASN	Asparagine, 1 Letter Code: N
ASP	Aspartic acid, 1 Letter Code: D
CHARMM	Chemistry at HARvard Molecular Mechanics
COM	Center of Mass
COV	Coverage
CUE	Coupling of Ubiquitin Conjugation to Endoplasmic Reticulum Degradation
CV	Collective Variable
CYS	Cysteine, 1 Letter Code: C
FP	Number of False Positives
Gdm⁺	Guanidinium Cation
GDS	Gibbs Dividing Surface
GGA3	Golgi-localized, γ -ear-containing, ADP-ribosylationfactor-binding Protein-3
GLN	Glutamine, 1 Letter Code: Q
GLU	Glutamic acid, 1 Letter Code: E

GLY	Glycine, 1 Letter Code: G
HFBII	Hydrophobin II
HHR23	Human Homologs of Saccharomyces Cerevisiae Rad23
HIS	Histidine, 1 Letter Code: H
HP_r	Histidine Phosphocarrier Protein
ILE	Isoleucine, 1 Letter Code: I
ITC	Isothermal Titration Calorimetry
k_B	Boltzmann Constant, $1.3806504 \times 10^{-23}$ J/K = 0.00198721 kcal/mol/K
K_d	Dissociation Constant
LEU	Leucine, 1 Letter Code: L
LJ	Lennard Jones
LV	Liquid-Vapor
LYS	Lysine, 1 Letter Code: K
MD	Molecular Dynamics
MET	Methionine, 1 Letter Code: M
MUP-1	Mouse Major Urinary Protein-1
NAMD	Not (just) Another Molecular Dynamics Program
NEDD8	Neuronal Precursor Cell Expressed Developmentally Downregulated Protein-8
NMA	N-methylacetamide
NMR	Nuclear Magnetic Resonance Spectroscopy
NPT	Constant Particle, Pressure and Temperature Ensemble
NVT	Constant Particle, Volume and Temperature Ensemble
OP	Order Parameter
PBC	Periodic Boundary Condition
PDB	Protein Data Bank

PHE	Phenylalanine, 1 Letter Code: F
PINCH	Particularly Interesting New Cysteine-Histidine-Rich Protein
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
PRO	Proline, 1 Letter Code: P
PTS	Phosphotransferase System
RDF	Radial Distribution Function
RI	Number of Real Residues
Rxn. Coord.	Reaction Coordinate
SAMs	Self-assembled Monolayers
SASA	Solvent Accessible Surface Area
SER	Serine, 1 Letter Code: S
SH2	Src Homology 2
SH3	Src Homology 3
SPC/E	Simple Point Charge Extended
SUMO	Small Ubiquitin-like Modifier
SWM4-NDP	Simple Water Model with 4-Sites and Negative Drude Polarizability
THR	Threonine, 1 Letter Code: T
TI	Thermodynamic Integration
TIP3P	Three-Point Transferable Intermolecular Potential
TIP4P-FQ	Four-Point Transferable Intermolecular Potential with Fluctuating Charges
TIP4P-QDP	Four-Point Transferable Intermolecular Potential with Charge-Dependent Polarizability
TRP	Tryptophan, 1 Letter Code: W
TP	Number of True Predictions

TYR	Tyrosine, 1 Letter Code: Y
UBA	Ubiquitin Associated
UBD	Ubiquitin Binding Domain
UBL	Ubiquitin Like Protein
UIM	Ubiquitin Interacting Motif
US	Umbrella Sampling
VAL	Valine, 1 Letter Code: V
vdW	van der Waals
WHAM	Weighted Histogram Analysis Method

ABSTRACT

Hydrophobic effects play a key important role in mediating the biological association and self-assembly processes. Among them, a prime example where hydrophobic effects have profound implications is from the protein related associations. In the context of protein related interactions, such as protein-ion interaction, protein-ligand interaction and protein-protein interaction, a prior knowledge of relevant binding interfaces, which are defined as clusters of residues involved directly with binding interactions, is difficult. In the binding events that mainly driving by hydrophobic effects, a routinely and widely used approach to predict the binding residues is simply based on the hydropathy value of single residue. However, recent studies suggest that consideration of hydrophobicity for single residues on a protein surface require accounting of the local environment dictated by neighboring residues and local water. Therefore, in the case of hydrophobic mediated association, it is the effective hydrophobicity with the consideration of neighboring effect and context dependency that determines whether the residue would involve in the binding patch. In this dissertation, I first use a method derived from percolation theory to evaluate spanning water networks in the first hydration shells of a series of small proteins in order to locate a critical hydration level to best distinguish the effective hydrophobic and hydrophilic region around protein surface. Further, residue based water density could be applied to scale the effective hydrophobicity at such a critical hydration level. Finally, single-linkage clustering methods were applied to cluster the effective hydrophobic residues in a well defined patch that are putatively involved in binding interactions. This simple method is able to predict with sufficient accuracy and coverage the binding interface residues of a series of proteins. The approach is competitive with automated servers. The

results of this study highlight the importance of accounting of local environment in determining the hydrophobic nature of individual residues on protein surfaces.

With the identified effective hydrophobic patch that is extensively involved in the protein binding, it is possible to further explore the ion specificity around the region. Umbrella sampling molecular dynamics simulation approach was applied to study the potentials of mean force along an order parameter bridging the state where the ion is fully solvated and one where it is biased via harmonic restraints close around the protein-water interface. Specifically, the protein hydrophobin-II (HFBII) with 71 amino acid residues expressed by filamentous fungi was the target protein. Such a choice is due to the fact that HFBII has an amphiphilic structure character with a well defined hydrophobic patch and several hydrophilic patches. Therefore, it is possible to compare the ion-specific effect around the hydrophobic and hydrophilic region of the protein. Two representative ions, Cl^- and I^- , which have been shown previously by simulations as displaying specific-ion behaviors at aqueous liquid-vapor interfaces, were considered in the study. We further explore anion-induced interface fluctuations near protein-water interfaces using coarse-grained representations of interfaces. As in the case of a pure liquid-vapor interface, at the hydrophobic protein-water interface, the larger, less charge-dense iodide anion displays a marginal interfacial stability compared with the smaller, more charge-dense chloride anion. Furthermore, consistent with the results at aqueous liquid-vapor interfaces, iodide induces larger fluctuations of the protein-water interface compared to chloride, which is an indication of the possible connection between the surface stability of the ion and the induced fluctuation of protein-water interfacial height of the ion. The correlation is further confirmed in the case of denaturant guanidinium cation and urea with different configurations as they approach the hydrophobic protein patch. Finally, hydrophobic effective was discussed in the context of protein-protein interaction. Using a rigid body model, the thermodynamic signatures of the association between ubiquitin and ubiquitin interaction motif was explored. Much like in the case of a purely hydrophobic solute, association is favored by entropic contributions from release of water from the interprotein

regions and association is disfavored by loss of enthalpic interactions. This is a further demonstration of the signature of the hydrophobic effect mediated association from the computational approach.

Chapter 1

INTRODUCTION AND MOTIVATION

1.1 Introduction

1.1.1 Hydrophobic Effect

In terms of the hydrophobic effect, the first impression for most of people is about the famous adage "oil and water don't mix". The significance of hydrophobic effect can be found in a wide region of biological process, such as membrane formation, protein folding and aggregation, binding of a substrate to the enzyme. [2, 3, 4, 5, 6, 7] In general, it is an effective force caused by the nonpolar molecules that have a tendency to minimize the aqueous solvent exposed surface area.

Traditional argument about the hydrophobic effect stems from the iceberg model by Frank and Evans. [8] The hydrophobic solute will disrupt the structure of bulk water and since it is incapable of hydrogen bonding with water, water-water hydrogen bonds are reoriented along such a surface in order to minimize disruption of the three-dimensional hydrogen bonded network of water molecules. This leads to a structured water cage formation around the hydrophobe surface and it has the orientation constraints on water molecules in the hydration shell of nonpolar solutes, resulting in a decrease in the entropy of water. This structure character of the hydration water around hydrophobic solute has been further proved by the neutron and x-ray scattering techniques. [9, 10, 11, 12] Such unfavorable effects from the loss of configurational entropy of water molecules can be minimized if hydrophobe molecules aggregate. Upon aggregation, water molecules form one larger cage surrounding the hydrophobic aggregate and the surface area of such aggregate is smaller than the sum of surface areas of individual solutes. This makes the entropic contribution less unfavorable and, hence,

makes the free energy more favorable. Besides this view, there are some other views on the hydrophobic hydration. One of them is based on the scaled particle theory. As the hydrophobic solute dissolves in the fluid, it requires a creation of a spherical cavity with radius λ . If the amount of reversible work in this process is expressed as $W(\lambda)$, [13] then the probability $P(\lambda)$ of finding a point outside the exclusion volume of the sphere is:

$$P(\lambda) = \exp[-W(\lambda)/kT] \quad (1.1)$$

Although with the limitation of the assumption of water molecules as rigid spheres, this theory is quite successful in prediction of the heats and entropies of nonpolar gases of aqueous solutions. [14] It has been further pointed out that this hydrophobic solvation process in aqueous depends on the size of the solute. [15, 16] For small solute case such as methane, the exclusion volume is small enough so that the solute can be accommodated in water without the breakage of hydrogen bonds. [17, 18] The solvation free energy in this case scales well with the volume of the solute. In contrast, the hydrophobic solvation is quite different in the large solute case with scale beyond nanometer. The larger exposed hydrophobic surface of the solute leads to unavoidable breakage of hydrogen bonds at the surface so that water molecules have a tendency to escape away from the large hydrophobic solute surface. As a result, it forms an interface between water and hydrophobic solute, which is similar to the liquid-vapor interface. In such large solute case, the solvation free energy cost scales better with the solute surface area instead of solute volume. [19]

1.1.2 Protein Hydrophobicity

As it comes into the protein case, hydrophobic effect is more complicated. Due to the strong heterogeneity of protein surfaces in terms of both topography (local, as well as global, geometry and shape) and chemical composition, different regions on protein surface may have distinct hydrophobicity. In order to characterize the hydrophobicity differences among amino acid residues with different chemical compositions, several ways of hydrophobicity scales have been developed. The most common method is

based on the measurement of free energies of transfer for the side chains of each type of amino acid between two immiscible phases. In Wolfenden hydrophobicity scale, one phase is selected as water and the other phase is selected as vapor, which is the simplest nonpolar phase. [20] Based on this, the order of the 20 amino acids from most hydrophobicity to least hydrophobicity is GLY, LEU, ILE, VAL, ALA, PHE, CYS, MET, THR, SER, TRP, TYR, ASP, LYS, GLN, GLU, HIS, ASP and ARG. Different from Wolfenden hydrophobicity scales that only consider the contributions of the sidechains, the Wimley-Whilte whole residue hydrophobicity scales also included the contributions of the peptide bonds. This consideration is especially important for the membrane proteins since the effect from the H-bonded peptide bonds would influence the position selection of transmembrane helix. [21, 22, 23, 24] In another Kyte and Doolittle hydrophobicity scale, the final scale values were not only determined by the water-vapor transfer free energies of side chains of each type of amino acid, but also by the interior-exterior distribution of amino acid. [25] This adjustment lowers the hydrophobicity scale of GLY that has high hydration free energy but low frequency of distribution in the interior of the protein, which makes more sense. In these considerations, the hydrophobicity scale of each type of amino acid is considered separately, which only determines by the chemical compositions of the residues. Using these absolute hydrophobicity scale to directly evaluate the protein surface hydrophobicity may cause problems due to the local environment effect. The local environment here is defined by a collection of protein residues nearby. Initially, this local environment effect is discussed in a simpler nanoscale plate system by Giovambattista et al.[26] They found that surface water density in the first hydration layer of a hydrophobic region with hydrophilic borders around is significantly higher than that of an identical hydrophobic region surrounded by hydrophobic borders, reflecting the effect that a canonically-defined hydrophobic region may represent a more or less hydrophobic environment right in the vicinity of its spatial location due to perturbations from its neighboring components. Later, these arguments and observations were found also applied to the self-assembled monolayers (SAMs) and heterogeneous protein surfaces, [27]

which lead to the perspective of effective hydrophobicity of a protein surface residue and warrants the view that the hydrophobicity of a group is context-dependent and thus a reflection of multiple effects of surrounding moieties on protein surfaces. [28] In this sense, these effective hydrophobic regions around protein are not isolated. Instead, they could form a cluster based on the chemical and topographical context. [29]

Due to the nanometer scale of protein surface, it is difficult to dissect the effective hydrophobic patch experimentally. In order to scale this context dependent effective hydrophobicity around protein surface, several approaches from simulation studies have been applied. Since the nature of hydrophobicity involves the disfavor of water molecules nearby, hydration of such effective hydrophobic region should display a high cost of free energy. Based on this, Beuming et al [30] performed thermodynamic analysis of water molecules around the protein surface and evaluated the hydration free energy around several hydration sites based on the inhomogeneous solvation theory [31, 32], which calculated the enthalpic and entropic contributions of each hydration site. Enthalpic contribution can be computed by considering the nonbonded interaction energy between hydration site water and the other components in the system. Entropic contribution can be estimated purely from protein-water correlation entropy by the overlook of water-water and other higher order correlation terms. From their calculations, hydration sites near the aromatic and aliphatic side chains manifested a higher average hydration free energy, which could be described as effective hydrophobic regions around protein surface. They further connected these high hydration free energy sites with the binding sites of the protein. Therefore, based on the free energetic characterization of water around protein surface, they could scale the effective hydrophobicity with application to binding sites prediction.

Besides the approximate calculation from inhomogeneous solvation theory, hydration free energy can further be obtained from density fluctuation around the probe solute as shown in the following equation: [33]

$$\Delta G_{dehydration} = -k_B T (\ln P_N - \ln P_0) \quad (1.2)$$

where $\Delta G_{dehydration}$ is the dehydration free energy, which has the same magnitude but opposite sign as the hydration free energy. P_N is the probability of observing N water molecules in the probe volume. P_0 is the probability of observing no water molecules in the probe volume, which is the reference state in this case. Based on the above equation, effective hydrophobic region is associated with lower dehydration free energy. As a result, the distribution profile of number of water molecules in the probe volume around effective hydrophobic region should display a fat low- N tail character as shown in the previous publication. [28, 34] In this sense, a larger fluctuation of the number of water molecules can be observed around the effective hydrophobic regions compared with the effective hydrophilic regions. This enhanced fluctuation around the effective hydrophobic region can be further characterized by the local compressibility [27]:

$$\chi_{fl}(z) = \frac{V}{kT} \frac{\langle N(z)^2 \rangle - \langle N(z) \rangle^2}{\langle N(z) \rangle^2} \quad (1.3)$$

where $\langle \rangle$ represents the ensemble average, k is the Boltzmann's constant, and $N(z)$ denotes the probability distribution of number of water molecules in an observation volume V that has the specific separation with the defined surface along z direction. Recently, Patel et al [35] shows another way to assess the context dependent hydrophobicity around a representative hydrophobin II protein surface based on the free energy of forming a cavity around the surface.

$$\mu = \int_0^\infty \langle N_v \rangle_\Phi d\Phi \quad (1.4)$$

where $\langle N_v \rangle_\Phi$ is the force to create a cavity with volume v and with a total number of N water molecules. With a factor Φ in the equation, it can couple the average number of water in probe volume linearly to the external biasing potential to empty the region.

Before applying these hydration free energy calculation method to estimate the effective hydrophobicity, a simple thought based on the water density around different regions has been attempted to apply. Godawat et al [36] monitored the water density near the surfaces of fully solvated self-assembled monolayers (SAMs) with different functional groups exposed in aqueous solution. However, it was found that in this case

water density shows a poor distinction around hydrophobic head groups ($-CF_3$, $-CH_3$) and hydrophilic head groups ($-OH$, $-CONH_2$). Only by considering the water density fluctuation around these two regions can give rise to distinct differences around effective hydrophobic and hydrophilic region as we pointed out before. Instead of considering the water density, Acharya et al [27] actually use the density of probe hydrophobic solutes to scale the effective hydrophobicity around protein surface. Here, they performed the molecular dynamic simulation of the representative hydrophobin II protein again in aqueous solution with limited amount of small hydrophobic solutes. In this way, the effective hydrophobicity around different regions can be evaluated from the average local density of hydrophobic probes in the vicinity. A limitation in this approach is that in order to prevent the probe hydrophobic solutes from aggregating during the simulation, the number density of the solutes must be careful chosen to be small enough. As a result, a sufficient long time simulation was required in order to obtain the converged local density around protein surface. Compared these two approaches to scale the effective hydrophobicity based on the density, the later one could give a better distinction of effective hydrophobicity due to the application of limiting amount of “probes” in the system.

Inspired by this, we propose a somewhat complementary, or “in like spirit” protocol to identify the effective hydrophobic region around protein surface based on the limiting amount of water molecules in the system. In this case, the protein in our study should be surrounded by a finite water shell with free boundaries. Unlike previous studies with fully solvated protein in the simulation box, water in our system should be able to arrange dissimilarly between hydrophobic and hydrophilic protein surface regions. There is no constraint that water density be equivalent at all positions around the protein surface. Based on the hydrophobicity nature, water molecules should manifest more tendencies to locate around the effective hydrophilic regions than the effective hydrophobic ones. Besides, applying water molecules as probes could avoid the issue of aggregation in the case of hydrophobic solutes at high concentration, so the sampling efficiency could improve. In this approach to deal with a partially solvated

protein surface, in theory, there should be an existence of a critical hydration level at which water molecules coverage could manifest a distinct variance around effective hydrophobic groups and effective hydrophilic ones. In this sense, we could determine the effective hydrophobicity around protein surfaces at this critical hydration level.

1.1.3 Ion-Specificity around Protein Surface

Evaluation of effective hydrophobicity is a key step towards the decent understanding of ion-specific effect around protein surface. Previously, interactions between ions and protein in aqueous solution have been widely studied. [37, 38, 39] Hofmeister effects or ion-specific effects, related to the modulation of surface tension and protein solubility by additive salts that influence the strength of direct and water-mediated interactions in solution have been intensely explored with the ultimate aim of extracting basic physical insights into the above mentioned processes[40, 41, 42, 43]. According to the ability to salt in or salt out proteins, the Hofmeister series has been proposed for different cations and anions. [44] For the anions, it has such a order: $F^- > HPO_4^{2-} > CH_3CO_2^- > Cl^- > NO_3^- > Br^- > I^- > SCN^-$. While for the cations, it has the following order: $NH_4^+ > K^+ > Na^+ > Li^+ > Mg^{2+} > Ca^{2+} > \text{guanidinium}$. In both of these two series, the ions on the left decrease the solubility of hydrophobic molecules by strengthening the hydrophobic effect; while the ions on the right increase the solubility of hydrophobic molecules through the weakening of hydrophobic effect. Therefore, modulation of effective hydrophobicity around aqueous protein surface through the addition of the ions can be considered as the microscopic origins and molecular mechanisms of ion-specific effects. A detailed understanding of the modulation could give us the insight about further design of interactions between protein and solute.

Among the vast discusses on ion-specific effects, halide ions stabilities around liquid-vapor interfacial regions have been paid great attention. [45, 46, 47, 48, 49, 50, 51, 52] It has been widely shown that larger halide ions such as I^- and Br^- , which locate at the right end of Hofmeister series, tend to bind to liquid-vapor interfaces more strongly and with lower transfer free energies than the early members of the series such as Cl^-

and F^- anions. Several factors, including ion size, ion polarizability, ion hydration properties and solvent polarizability [53] have been considered in order to describe a unifying explanation of the molecular underpinnings of such behaviors. Previously, polarizability was considered as a key issue to modulate the ion-water interaction and it was critical to apply polarizable force fields in order to accurately describe the halide specific effect around the interfacial region. [54] Later, it was pointed out that this argument was debatable since a careful parameterized non-polarizable force fields could also give a consistent halide adsorption behavior around interface. Therefore, a fully description regarding this is still required.

Recent studies [55, 56, 57] have begun to consider the halide specific effect around liquid-vapor interface from the perspective of perturbations of interfacial water molecules as the anions approach the interface. It has been suggested that for the two chemically distinct anions Cl^- and I^- , which represent the neutral and chaotropic positions in the Hofmeister series, the more surface stable I^- anion induces larger interfacial fluctuations compared to the non-surface active species Cl^- , thus demonstrating a strong correlation with induced interfacial fluctuations and anion surface stability as observed from molecular simulations. Further, the differences in induced interfacial fluctuations by Cl^- and I^- could be related to the nature of the hydration environment around the anions; water molecules in the hydration shells of I^- are shown to be more dynamic and less persistent compared to those in proximity to Cl^- . When approaching the liquid-vapor interface, coupling of local solvent around anions with solvent further away and near an interface leads to different perturbations of the interface by the two anions, and thus different contributions to interface height fluctuations, and ultimately surface stability via contributions from interfacial entropy arising from surface fluctuations correlations[55, 57, 56]. This provides a new insight to interpret the ion-specificity around liquid-vapor interface. Since there is implied a connection of the behaviors of ions at aqueous liquid-vapor interfaces to those of biochemically relevant interfaces such as protein-water interface and bilayer-water interface [58], it is natural to further explore the halide specificity around a more general hydrophobic

surface region. Heyda et al.[59] found that larger halides, I^- and Br^- , displayed a preferential spatial correlation with the hydrophobic methyl groups in the molecules of N-methylacetamide (NMA). It is also shown that the free energetics of transferring I^- and Br^- from bulk aqueous solution to a hydrophobic self-assembled monolayer-water interface is lower compared with Cl^- and Na^+ cases. Furthermore, Lund et al. probed the distribution of F^- and I^- around a spherical macromolecule in an uncharged case. [60] Around this hydrophobic particle, F^- shows a repulsion nature while I^- tends to be weakly attracted to it. Jungwirth and coworkers also provided volumes of data on the nature of differential binding of anions to protein surfaces. [61, 62, 63] Also, it has been suggested that ion-specific effects are dissimilar around hydrophobic and hydrophilic surfaces, with large I^- showing a stronger affinity than the smaller halide ions to the hydrophobic surfaces while the reverse trends of size-dependence of halide ions are realized at the hydrophilic surfaces. [64, 65, 66, 60] Their explanation is based on the point that more charge-dense Cl^- tends to have stronger direct electrostatic interaction with the hydrophilic region of the protein compared with less charge-dense I^- ; on the other hand, larger and partially-hydrated I^- would have a larger extent of solvent-assisted attraction with the hydrophobic patch of the protein. A detailed understanding of the mechanism of the so-called solvent-assisted attraction was required in this sense.

1.1.4 Denaturants around Protein Surface

The investigation of interactions between protein and simple halides can move a step further by considering the more complicated solutes guanidinium cation and urea. Both of the guanidinium cation (Gdm^+) and urea can serve as the protein denaturants. Seeking for a deep and fundamental understanding of this denaturation process has enjoyed a long history. [67, 68, 69, 70, 71, 72, 61, 73, 74, 75, 76, 77, 78, 79, 80, 43, 81] A main controversy about the denaturation is whether the denaturants affect the protein structure through a direct way or indirect way. In a direct way mechanism, the denaturants would interact with protein backbones or side chains directly. The interaction

may involve electrostatic interaction, van der Waals and hydrophobic interactions. In an indirect mechanism, the denaturants would modulate the solution properties to affect the protein denaturation. We notice that the detailed mechanism of this process is dependent on the denaturant investigated. Here, we are particularly interested in the denaturants guanidinium chloride (GdmCl) and urea. Both of them belong to the weak denaturants, so a significant high concentration was required for denaturation the proteins. (For urea, it usually requires a concentration of 5 M; for GdmCl, it usually requires a concentration of 7 M.) In such a high concentration, the notion of the existence of direct interaction is generally acceptable.

In terms of direct interactions, one of the major concerns for denaturation involves the lessening of the hydrophobic effect as it involves the formation of a compact "pre-folded" ensemble of states where protein hydrophobic surface exposure to solvent is reduced in relation to the purely unfolded ensemble of states. The idea is that by associating with hydrophobic regions around the protein surface, denaturant molecules can shield the hydrophobic surface area even in unfolded or extended configurations of the peptide/polymer. This idea has been proved by the molecular dynamic simulations. [77, 75] Recently, from an experimental approach of chemical force microscopy measurements, Ma et al [82] showed that in a certain range of pH values, addition of guanidinium groups could diminish the measurable hydrophobic interactions, which is a further verification of this mechanism. This chemical denaturation mechanism naturally involves direct interaction of the cosolvent molecule with regions of the protein surface. A particular aspect of this interaction deals with the precise nature of association geometries and the associated free energetics; specifically, molecules such as urea, and more so guanidinium cation (Gdm^+), can present several predominant relative orientations to the protein surface through which the interaction is mediated. In general, it is proposed that a dominant interaction of urea with surface groups in protein simulations involves hydrogen bonding with polar side-chain functions [83, 84], while the unique hydration properties of the Gdm^+ [85] support alternative interaction modes involving stacking with side-chain planar and hydrophobic groups.

It is worth to notice that both Gdm^+ and urea molecule possess planar structures. Previous studies have shown that in both the 1 M and 5 M GdmCl solutions with presence of liquid-vapor interfaces, Gdm^+ has tendencies to locate around the interfacial region by adopting a parallel orientation to the liquid-vapor interface. [86] A later study further verified this by considering the free energies of transferring single Gdm^+ with various of orientations from bulk to the liquid-vapor interface. [87] There is a little free energy minimum around liquid-vapor interface with parallel orientation Gdm^+ while it shows a repulsive nature for the perpendicular orientation Gdm^+ . Besides the liquid vapor interface, the orientation preference for the Gdm^+ can be also found in the case of flat hydrophobic plate [75, 88] and hydrophobic polymer surface [77]. The orientation preference of Gdm^+ is derived from the anisotropic hydration characters of the molecule. Within the plane of Gdm^+ , the N-H group can serve as hydrogen bond donor, interacting with water molecules. Therefore, it is fully solvated and can be considered as a hydrophilic molecule. Above or below the planar face, Gdm^+ is inadequate to serve either as hydrogen bond donor or acceptor. In this case, the molecule is only partial solvated and can be considered as a hydrophobic molecule. The different solvation patterns in these two configurations of orientation result in their different adsorption behaviors. As the Gdm^+ approaching the hydrophobic surface region with its planar surface parallel, it mimics the hydrophobe-hydrophobe association, which is more free energetically favorable.

1.1.5 Protein-Protein Interactions Mediated by Hydrophobic Effect

Protein-protein association always occur in the aqueous solution. More and more arguments [6, 89] have pointed out that water should not just be considered as environment. More importantly, it can serve as an active player during the association. In the context of water mediated protein-protein association, hydrophobic effect is crucial.

A major issue concerning hydrophobic mediated binding processes is the driving force for the association process. In a general case, the characteristic thermodynamic

signature of association depends on nature of protein surface involved in the association. In the case of association between hydrophilic binding pockets and hydrophilic binding ligands, usually it is enthalpy-driven. A representative example of this is the binding of galactose to Arabinose Binding Protein (ABP). The thermodynamic quantity of this reaction has been measured by isothermal titration calorimetric (ITC) experiments with a favorable change of enthalpy around -95 kJ/mol. [90] When it turns into the case of association between the hydrophobic binding patch of protein and hydrophobic ligand, we would expect to observe a favorable contribution from entropy from a classical view of view. [15, 91, 89] That is, the overall stabilizing contribution to the complex state is an increase in total entropy (the origins of which generally are associated with the release of water degrees of freedom upon association, thus allowing more configurational states). In this model, proteins are fully solvated in the dissociation state. As the binding partners approaching each other, water molecules around the surface regions towards to binding are squeezed out, resulting in a dewetting transition. The expelled water to the bulk has more translational degree of freedom, which is an increase of entropy that would drive the association. In fact, previous study involved the modeled hydrophobic solute would give a support of this concern. Investigation of hydrophobic associations between hydrophobic plates, [92, 93] graphene plates [94] and carbon nanotubes [95] have all shown that the process is highly entropically favorable. It is important to realize that although the real hydrophobic protein surface shares some of the characters of purely nanoscale hydrophobic solutes, the heterogeneity of protein surfaces in terms of both topography and chemical composition complicated the issue. Investigation of the binding process of nonpolar ligand to the poorly solvated pocket of the mouse major urinary protein-1(MUP-1) indicates that despite the apparent hydrophobic character of the binding partner, the binding was enthalpy-driven and accompanied by an unfavorable entropy change. [96, 97, 98] A detailed check suggests that by virtue of poor solvation of the binding pocket, the gain of protein-protein interaction arising from complexation is larger than the lost of protein-solvent interaction

prior to the association. Therefore, it could compensate favorably. In another example, the binding event between substrate protein trypsin, which exposes the binding sites outside in an easily solvated way, and a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin is studied. It is found that the interaction is strongly entropy driven in a wide range of temperatures. [99] Based on these, it suggests that the characteristic thermodynamic signature of hydrophobic association in solution will depend on the degree of solvation of the binding pocket.

To further verify this point, molecular dynamics simulation study has been applied to investigate the enthalpy driven hydrophobic association by Setny. [100] A model for nonpolar cavity-ligand association is used in their molecular dynamics simulation. Thermodynamic contributions, including free energy, entropy and enthalpy along the binding coordinate have been investigated. The results show that the favorable driving force for this process is from enthalpy change among the release of water molecules from the hydrophobic environment to the bulk water. Although there are some controversial issues like the origin of the unfavorable entropic component for the hydrophobic association process, it is generally accepted that the enthalpy driven hydrophobic association usually involves the receptor protein that has poorly solvated binding sites. This reflects the notion that the water is not able to recoup energetically favorable water-water interactions via sufficient orientation restriction. More importantly, the above experimental and simulation results suggest that the underlying signatures of hydrophobic interactions are by no means absolute. A detailed understanding of this requires the investigation of protein surface property and solvation around. Therefore, if we could have a prior knowledge of protein binding site, it will be helpful for us to uncover the underlying thermodynamic signature.

1.2 Objective

In this dissertation, initially, molecular dynamic simulation was applied to evaluate the effective hydrophobic patches for proteins with known three-dimensional structures. With introducing of limiting amount of water molecules as probes in the system,

it is possible to take into account of the hydrophobic effect explicitly around the protein surface. Hydrophobic effect plays a vital role in driving interactions between protein and other solutes, such as ion, binding ligand and other protein. In our understanding, this hydrophobic effect can also be understood from a binding partner substitution point of view. Water molecules have quite weak interactions with the effective hydrophobic region around protein surface. As a result, when the binding partner of the protein approaches, the water molecules around effective hydrophobic region are easily displaced to leave the region exposed to bind. Therefore, identification of effective hydrophobic region on protein surface can serve as an indication of the possible binding sites for the protein. These binding sites information is quite essential for us to further manipulate and design protein related interaction.

Therefore, the objective of my first project involves the characterization of effective hydrophobicity scale around various protein surface with known three-dimensional structures. We propose a solution to study a single partial hydrated protein with coverage of different numbers of hydration water. With consideration of the hydrogen-bonded water network distributions around the protein surface at these hydration levels, we would like to first locate a proper hydration level at which water molecules as probes would give the best distinction between the effective hydrophobic region and the effective hydrophilic region. For different proteins, such a critical hydration level should be dissimilar since it may depend on the size, geometry and chemical composition of protein. With an identified critical hydration level, it is possible to further scale the effective hydrophobicity around protein surface based on the local water number density around each residue. During the binding event, several residues that are closely packed in space usually involve collectively. Considering this, we could apply the single linkage clustering method, which is a way to assign some points with know positions in space into groups according to their distances, to the selected low-hydrated residues for defining a continuous effective hydrophobic patch for the protein. This identified effective hydrophobic patch can serve as an implication of binding sites of the protein. We could further verify this by comparison of our predicted binding sites with the

experiment determined ones. The meaning of this work is that the located effective hydrophobic region could putatively involve in the binding interactions of the protein with other solutes. Therefore, this could serve as an independent way to predict the binding patch for proteins mediated through hydrophobic effect. This project is presented in Chapter 3.

With the identified effect hydrophobic patch, it is possible for us to further delve into the underlying mechanism of the association between the protein and binding partner. Starting from the simple case, initially I would like to consider the binding of the monovalent anion to the protein surface. The difference in the adsorption of two representative halide, Cl^- and I^- around liquid-vapor interface has been widely studied as mentioned in Section 1.1.3. From a novel point of view involving long-range perturbation of interfacial water, we have shown that as each of these two types of ion approaching the liquid-vapor interface, they would have different magnitudes of induced interfacial height fluctuation. I^- with more malleable solvation shell could easily couple with the water around liquid-vapor interface, leading to a higher interfacial height fluctuation and providing a favorable contribution in the association in terms of larger surface entropy. Acknowledging the intrinsic connection between the liquid-vapor interface and aqueous hydrophobic protein interface, we would like to seek a similar trend of adsorption and induced fluctuation behaviors for Cl^- and I^- as each of them approaching the effective hydrophobic patch of a rigid protein in aqueous environment. It will be helpful if we could establish a correlation between the free energetics (probabilities) of the two types of anions near the hydrophobic protein region and their induced interfacial height fluctuation, which happened in the case of liquid-vapor interface system. The behaviors of these two types of anions around the hydrophilic region of a protein were also studied for a comparison to further understand the origin of the ion-specific effect around protein surface. The particular protein we focus on in this study is hydrophobin-II (HFBII), which is a small protein with 71 amino acid residues expressed by filamentous fungi. The protein is known for its ability to form a hydrophobic coating on the surface of an object and it can self-assemble into a

monolayer on hydrophobic/hydrophilic interfaces such as a water/air interface. These functions are mainly determined by the amphiphilic structural characterization. From the method developed in Chapter 3, we could identify an effective hydrophobic patch of HFBII consisting of residues V18, L19, L21, I22, V24, V54, A61, L62 and L63. Besides, this protein also possesses several well-defined hydrophilic patches. Considering of this, this protein is an ideal candidate to compare the characters between hydrophobic and hydrophilic interfaces as ions approach. This whole project is discussed in Chapter 4.

With a better understanding of the association between simple halide and the protein surface with different effective hydrophobicity, we could further move to more complicated solute cases with consideration of the stability of Gdm^+ /urea around protein surface. Such an exploration of the stability and interaction would possibly provide views of denaturation mechanisms regarding the two solutes. It is suggested that Gdm^+ display orientation preference around the hydrophobic surface, including the liquid-vapor interface, hydrophobic plate and hydrophobic polymer surface. However, it still lacks a direct evidence for similar orientation behavior of Gdm^+ upon approaching the aqueous hydrophobic protein interfaces. The inherent chemical and topographical heterogeneity of protein surface makes it difficult to find a qualitatively rigorous approach to evaluate the relative orientation between the surface of Gdm^+ and the protein. To fill this gap, we apply molecular dynamics simulations investigating the association of Gdm^+ cation with a specific protein, HFBII, with a relatively flat surface region consisting of effective hydrophobic residues. Another denaturant urea, which shares the structure similarity with Gdm^+ , will also be explored as it approaches the same surface region of the protein. In the context of chemical denaturation via direct association, we ask here about the orientations that Gdm^+ /urea adopt when interacting with hydrophobic regions of proteins. The combination of this analysis addresses ideas of direct interaction as well as hydrophobic effects as they pertain to the denaturation process. Besides, in our previous discussion of Gdm^+ orientation preference around the liquid-vapor interface, we found that there is an interesting correlation between

the interfacial stabilities and induced interfacial height fluctuations of Gdm^+ with different orientations. Interfacial more stable Gdm^+ with parallel orientation displays a higher level of induced interfacial height fluctuation compared with less surface stable perpendicular configuration showing a lower level of perturbation of the interfacial water. We would like to further extend this idea to the case of Gdm^+ approaching a real protein surface with well-defined effective hydrophobic region. This work could be considered as a further extension of the second project discussed in Chapter 4. As parallel oriented Gdm^+ approaching the protein patch, it is partial solvated similar to the I^- case; as perpendicular oriented Gdm^+ approaching the protein patch, it is fully solvated similar to Cl^- case. If parallel orientation Gdm^+ could display more surface stability and induce larger extent of interfacial height fluctuation compared with the perpendicular one, then the whole result will be self-consistent. This part of work is presented in Chapter 5.

Hydrophobic effect not only plays a key role in the association between protein and small solute, but also may be responsible for the protein-protein association. Previous discussion has suggested that underlying signatures of the hydrophobic association in a system heavily rely on the extent of solvation around the hydrophobic interfacial regions. Therefore, if we could have a prior knowledge of protein binding site, it will be helpful for us to uncover the underlying thermodynamic signature. Ubiquitin is a protein with well characterized structure and known binding patch consisted of residues L8-I44-H68-V70. This patch can be considered as an effective hydrophobic patch that largely involves in the hydrophobic association with a bunch of ubiquitin interacting motif, which has been identified from our protocol developed in Chapter 3. In part, a goal of this study is to extend the analysis and discussion of underlying signatures of the hydrophobic association between ubiquitin and one of its binding partners - ubiquitin interacting motif (UIM). Through molecular dynamic simulation, we could like to connect the solvation situations around the binding sites of the two partners with the characterized thermodynamic signature of the association. Besides, the binding partner UIM of Vps27 adopts a helical conformation. The helix is markedly

amphiphilic with a hydrophobic stripe along one side which interacts with the complementary hydrophobic Leu8-Ile44-Val70 region of ubiquitin as proved by experiment. On the other side of the helix, it is quite hydrophilic, which may be unlikely to bind with ubiquitin. We aim to use molecular dynamics simulations in conjunction with free energy sampling methods to calculate the potential of mean forces (PMF) for reversible association of the two proteins taken to be semi-rigid bodies when the helix is restrained at different orientations to approaching the ubiquitin. We would like to observe significant free energetics differences when UIM binding with hydrophobic side and hydrophilic side. This piece of study is discussed in Chapter 6.

In the next chapter, I will start with a general discussion on the force field and routine analysis in the molecular dynamic simulation.

Chapter 2

FORCE FIELDS AND COMPUTATIONAL METHODS

2.1 Molecular Dynamics Simulation and Force Fields

Molecular dynamics simulation is a molecular modeling approach with the aid of computer to understand the properties of assemblies of molecules. In such a computer simulation, the physical movements of all the atoms, which are the basic building blocks for the molecules, are governed by the Newton's laws of motion. In order to obtain the positions and velocities of the particles in the system, we need to solve the classical equations of motion at each time step:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \quad (2.1)$$

Here, r_i represents the coordinate of the particle and m_i represents the mass of the particle at time t with total force F_i acting on the particle. Furthermore, one can obtain the total force F_i based on the potential function:

$$F_i = -\frac{\partial U}{\partial r_i} \quad (2.2)$$

In molecular modeling, based on Born-Oppenheimer approximation, U is the interatomic potentials defined by a set of parameters that is a function of the nuclear positions only. These parameters are derived from quantum calculations and experimental data. Usually, U is termed as force field in the molecular dynamic simulation and it involves two parts:

$$U = U_{\text{bonded}} + U_{\text{nonbonded}} \quad (2.3)$$

U_{bonded} is the contribution from the covalent bonded interactions while $U_{\text{nonbonded}}$ is the contribution from the nonbonded interactions.

The covalent bonded interactions include the following parts:

$$U_{\text{covalent}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{improper}} \quad (2.4)$$

U_{bond} is the energy function of bond stretching. U_{angle} is the energy function of angle bending. U_{dihedral} is the energy function of torsion angle or dihedral. U_{improper} is the energy function of improper torsions, which is from out of plane bending. The details of these terms are shown below:

$$U_{\text{bond}} = \sum_{\text{bond}} K_b (b - b_0)^2 \quad (2.5)$$

$$U_{\text{angle}} = \sum_{\text{angle}} K_\theta (\theta - \theta_0)^2 \quad (2.6)$$

$$U_{\text{dihedral}} = \sum_{\text{dihedrals}} K_\phi (1 + \cos(n\phi - \delta)) \quad (2.7)$$

$$U_{\text{improper}} = \sum_{\text{improper}} K_\omega (\omega - \omega_0)^2 \quad (2.8)$$

K_b , K_θ , K_ϕ and K_ω represent the bond force constant, the angle force constant, the dihedral force constant and the improper force constant respectively. b_0 and θ_0 represent the equilibrium bond length and angle. In Equation 2.7, n is the multiplicity; ϕ is the dihedral angle and δ is the shift of phase. In Equation 2.8, $\omega - \omega_0$ is the out of plane angle. For the nonbonded interactions, it is the summation of electrostatic interaction and van der Waals (VDW) interaction:

$$U_{\text{noncovalent}} = U_{\text{electrostatic}} + U_{\text{VDW}} \quad (2.9)$$

In the fixed-charge force field, the electrostatic interaction can be treated as the Coulomb potential:

$$U_{\text{electrostatic}} = \frac{q_i q_j}{\epsilon r_{ij}} \quad (2.10)$$

For the VDW interaction, usually, it is modeled via the Lennard-Jones (LJ) potential. In one of the most common force field, Chemistry at Harvard Molecular Modeling (CHARMM), [101, 102] it can be expressed as the follow:

$$U_{\text{VDW}} = \varepsilon_{ij} \left(\frac{R_{\text{min},ij}^{12}}{r_{ij}^{12}} - 2 \frac{R_{\text{min},ij}^6}{r_{ij}^6} \right) \quad (2.11)$$

where ε is the energy parameter, representing the depth of the potential well; R_{min} is the distance parameter, representing the distance at which the potential reaches its minimum. For each type of atom in the molecule, it has one set of these parameters. In the actual calculation of interactions between two different atom sites, the Lorentz-Berthelot (LB) combining rules are applied in the CHARMM force field.

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}, \quad R_{\text{min},ij} = \frac{R_{\text{min},i} + R_{\text{min},j}}{2} \quad (2.12)$$

2.2 Radial Distribution Function

The radial distribution function (RDF) in statistical mechanics gives the probability of identifying a particle in the certain distance of another particle. Usually, we denote this function as $g(r)$ with r representing the distance away from a reference particle. Distances between all pairs of reference particle and considered particle are calculated and binned into a histogram to count the number of particle at each separation. In order to obtain the probability, the number of particle needs to be divided by the volume it occupies as the separation from the reference varies. Consider a spherical sampling region with distance r from the reference and Δr as the thickness. The occupying volume of the considered particle is given by:

$$V = \frac{4}{3}\pi[(r + \Delta r)^3 - r^3] \approx 4\pi r^2 \Delta r \quad (2.13)$$

Divided the number of particle at each separation by this volume factor gives the number density at each separation. Furthermore, if we consider the mean number density in the whole system as ρ , $g(r)$ can be obtained by normalizing the number density at each separation with the mean number density.

$$g(r) = \frac{n(r)}{\rho 4\pi r^2 \Delta r} \quad (2.14)$$

As $g(r)$ is greater than one, it indicates an enhancement of distribution of the considered particle in this region; as $g(r)$ is less than one, it indicates a decrease of probability of finding the considered particle in this region. $g(r)$ provides useful information about the structure of liquid and it can also be obtained by experiment such as X-ray diffraction. Therefore, it is a useful piece of information to verify the force field in the calculation.

We note that the discussion of RDF usually applies to the bulk system without interfaces in any dimensions. In a system possessing the interface, such as the liquid-vapor interface system, one can consider the depth-dependent transverse distribution functions. [103, 104] In the calculations, the system is first divided into several slabs paralleling to the interface with finite-width along the normal vector of the interface. Then, in each slab, one-dimensional transverse distribution function can be obtained. Such a consideration can reflect the structural changes of the liquid particle as a function of distance from the interface.

2.3 Density Profiles

Besides the radial distribution function, sometimes it is also important to know the exact density of particles along a specific direction, which is the density profile. Usually, the adopted density here is the number density of one species in the system. Considering a liquid-vapor interface system with the normal vector along z direction for an example. The density profile along this z direction can be expressed as:

$$\rho(z) = \frac{\langle N(z) \rangle}{L_x \times L_y \times \Delta z} \quad (2.15)$$

where $\rho(z)$ is the number density of particles at a specific z position and $\langle N(z) \rangle$ is the corresponding average number of particles in the slice at z position. In order to obtain a smooth density profile, the computed density profile can be further fit to an error function with the following form: [105, 106, 107]

$$\rho(z) = \frac{1}{2}(\rho_L + \rho_V) - \frac{1}{2}(\rho_L - \rho_V)\text{erf}\left(\frac{z - z_0}{\delta_e}\right) \quad (2.16)$$

where ρ_L is the density of liquid phase and ρ_V is the density of the vapor phase, z_0 is the position called Gibbs dividing surface (GDS) where water density is around

half of the bulk density, δ_e is the intrinsic interfacial thickness. Besides fitting to the error function, it is also possible to apply a hyperbolic tangent function to smooth the computed density profile based on the form: [108, 109, 110, 111]

$$\rho(z) = \frac{1}{2}(\rho_L + \rho_V) - \frac{1}{2}(\rho_L - \rho_V)\tanh\left(\frac{z - z_0}{\delta_t}\right) \quad (2.17)$$

δ_t is the intrinsic interfacial thickness from hyperbolic tangent fit. There are significant differences between the intrinsic interfacial thickness obtained from error function fit δ_e and hyperbolic tangent function fit δ_t . In order for a direct comparison, interfacial thickness based on “10-90” definition δ usually applies. Base on this definition, the interfacial thickness is the width over the region density changing from 10% to 90%. There is a direct relationship between “10-90” thickness and intrinsic interfacial thickness: $\delta_e^{10-90} = 1.8124\delta_e$ and $\delta_t^{10-90} = 2.1972\delta_t$. Usually δ_e^{10-90} should be quite close to δ_t^{10-90} . These interfacial thickness can be applied to estimate the critical temperature. With increasing of the temperature, the thickness of the interface increases and at the critical temperature it reach to the infinite. A linear relationship can be found between temperature and the reciprocal of the thickness. Therefore, the intercept of the line between T and $1/\delta$ can be an estimate of the critical temperature.

2.4 Surface Tension

Water molecules around the surface bear the imbalanced forces. Water-water interaction is much stronger than water-air interaction, resulting in a net inward force for water molecules around the surface and this is the origin of surface tension. Surface tension can be calculated from the average difference in the normal and tangential elements of the internal pressure tensor by the following equation: [112]

$$\gamma = \frac{L_z}{2} \left(P_{zz} - \frac{P_{xx} + P_{yy}}{2} \right) \quad (2.18)$$

In the calculation, z direction is considered as the direction normal to the surface, x and y directions are considered as the directions tangential to the surface. Therefore, P_{xx} and P_{yy} are the tangential elements of internal pressure tensor, P_{zz} is the normal

element of internal pressure tensor, L_z is the simulation cell length in the z direction. The magnitude of surface tension depends on the temperature. At higher temperature, the interfacial water molecules have less interactions with each other, resulting in a decrease of surface tension. As the temperature keep increasing to the critical point, a uniform fluid phase is reached. One can predict the critical temperature T_c from the surface tension values at different temperatures based on the following equation: [113, 114]

$$\gamma(T) = c_1(1 - \frac{T}{T_c})^{11/9}(1 - c_2(1 - \frac{T}{T_c})) \quad (2.19)$$

T_c can be obtained based on the nonlinear fitting from the above equation, where unknown parameters c_1 and c_2 can be obtained simultaneously. Here, an exponent constant 11/9 was applied, which is suggested by previous publication. [113, 114] Besides this approach, the critical temperature can also be obtained from a three-term Wegner expansion based on the condensed phase and vapor phase density in the following form: [115, 116, 117]

$$\rho_{LV} = \rho_c + C_2(1 - \frac{T}{T_c}) \pm [B_0(1 - \frac{T}{T_c})^\beta + B_1(1 - \frac{T}{T_c})^{\beta+\Delta}] \quad (2.20)$$

where ρ_{LV} is the density from liquid phase or from vapor phase, ρ_c is the critical density, T_c is the critical temperature, C_2 , B_0 and B_1 are variable constants that can be obtained from the fit, β and Δ are the universal critical parameters from the renormalization group theory. β is taken to be 0.325 [118] and Δ is taken to be 0.5[115]. Using the optimization algorithm from Nelder and Mead, [119] the fit parameters can be determined. We note that according to our previous study, [120] critical temperature based on a Wegner fit usually results in a lower estimate value than that from the surface tension fitting approach from Equation 2.19.

2.5 Vapor Pressure

In the system of liquid-vapor interface, vapor pressure is another important property. Depending on which direction in the simulation box was defined as the normal vector, the vapor pressure P_{vap} equals to P_{normal} . [121, 122] If z direction is

considered as the direction normal to the surface, then $P_{vap} = P_{normal} = P_{zz}$. Similar to the surface tension, the vapor pressure is also dependent on the temperature. Based on the Antoine's law, [123] vapor pressure can be related to the temperature from the following equation:

$$\ln(P_{vap}) = A + \frac{B}{T + C} \quad (2.21)$$

where A, B, C are adjustable parameters can be determined from nonlinear fitting. Based on this equation, one can compute the critical pressure P_c from the critical temperature T_c .

2.6 Dipole Moment Profile

In a molecule, due to the differences in electronegativity of various of atoms and the geometry of the molecule, there exist a separation of positive and negative charges in the molecule. In this case, dipole moment μ is define as the product between charge q and separation of the positive charge and negative charge d. Here, q is a scalar and d is a vector pointing from negative charge to positive charge. Therefore, μ is also a vector. Dipole moment is an important measurement of the electrostatic and geometric property of the molecules. In the liquid-vapor interface system, the dipole moment profile of the solvent molecules along the normal vector is usually considered. The magnitude of the dipole moment of one molecule can be calculated by the following equation:

$$\mu = \sqrt{\left(\sum_i q_i x_i\right)^2 + \left(\sum_i q_i y_i\right)^2 + \left(\sum_i q_i z_i\right)^2} \quad (2.22)$$

where i denotes an atomic site in the molecule. The direction of dipole moment can also be useful as a measurement of the orientation of the molecule. In a liquid-vapor interface system with z direction as the normal vector, usually the orientation of water molecule can be estimated from θ , which is defined as the intersect angle between the z direction and the dipole moment vector in space. Furthermore, order parameter can be defined based on θ as $P_1 = \langle \cos \theta \rangle$ and $P_2 = \frac{1}{2} \langle (3\cos^2\theta - 1) \rangle$. P_2 value close

to zero indicates a random orientation with less order for the water molecules in the system.

2.7 Interfacial Potential

Interfacial potential is another important property for the liquid-vapor interface system as a reflection of a combination of orientation and electrostatic state. According to previous study, [124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 52] interfacial potential can be determined by the double integration of the z component of charge density along the surface normal as follows:

$$\Delta\Phi(z) = \Phi(z) - \Phi(z_0) = -\frac{1}{\epsilon_0} \int_{z_0}^z dz' \int_{z_0}^{z'} dz'' \rho(z'') \quad (2.23)$$

where z_0 is the center of mass of the bulk region and $\rho(z'')$ is the z component of charge density. The obtained interfacial potential can be further decomposed into two parts: contributions from molecular dipole moment and from molecular quadrupole moment. [134, 135, 136] The dipole moment density is defined as:

$$P_z(z) = \langle \sum_m \delta(z - z_m) (\sum_i q_{im} z_{im}) \rangle \quad (2.24)$$

where the indices m and i denote a molecule and an atom site within the molecule. The dipole moment contribution then is computed via the integration of the dipole moment density $P_z(z)$ over z direction:

$$\Delta\Phi_M = -\frac{1}{\epsilon_0} \int_{z_0}^{\infty} dz P_z(z) \quad (2.25)$$

Further, the quadrupole moment density can be expressed as:

$$Q_{zz}(z) = \langle \sum_m \delta(z - z_m) (\frac{1}{2} \sum_i q_{im} z_{im}^2) \rangle \quad (2.26)$$

In both the calculation of dipole moment density and quadrupole moment density, oxygen atom in the water molecule was taken to be the molecular specific center z_{im} . The quadrupole contribution to the interfacial potential is calculated from the difference of quadrupole density $Q_{zz}(\infty)$ and a reference value $Q_{zz}(0)$.

$$\Delta\Phi_Q = -\frac{1}{\epsilon_0} |Q_{zz}(\infty) - Q_{zz}(0)| \quad (2.27)$$

2.8 Free Energy and Potential of Mean Force

Free energy is considered to be a critical quantity in thermodynamics since it determines the equilibrium state of a system. For two defined states in the system, we can consider one is the initial state and the other one is the final state. What we care about is the free energy difference since it determines whether the change between these two states is thermodynamically favorable or not. Based on the statistical mechanics, the free energy difference between these two states can be expressed in the following:

$$\Delta A = A_{\text{final}} - A_{\text{initial}} = -\frac{1}{\beta} \ln \frac{Q_{\text{final}}}{Q_{\text{initial}}} \quad (2.28)$$

Q_{final} and Q_{initial} represent the partition function of final and initial state respectively. This equation can be further simplified as:

$$\Delta A = A_{\text{final}} - A_{\text{initial}} = -\frac{1}{\beta} \ln \frac{P_{\text{final}}}{P_{\text{initial}}} \quad (2.29)$$

P_{final} and P_{initial} represent the probability in the final and initial state. The logarithmic relationship implies that probabilities of finding the system in different states may have a tremendous difference due to the variation of free energy. Due to the limited time scale in the computational simulation, sometimes it is rather difficult to sample the system in some high free energy states. Several methods have been developed to overcome this sampling issue and we will discuss them in the later. In molecular dynamics simulation, sometimes we would like to have a detailed track of the free energy change along a path from the initial state to the final state. In this case, a collective variable of the system can be defined and varies along the path. The free energy involves in this type of collective variable change along a certain path is called potential of mean force (PMF). If we consider two particles are brought together from an infinite-separation, dissociated state to the associated, contact state, PMF is associated with the reversible work in this process. The connection between the PMF and free energy can be found in early work by Kirkwood[137] as expressed in Equation 2.30. Recently, a refined expression has been discussed by Wong et al as shown in Equation 2.31: [138]

$$\frac{dA(\xi_0)}{d\xi_0} = -\langle F_{\xi_0} \rangle_{\xi_0} \quad (2.30)$$

$$\begin{aligned}
\frac{dA(\xi_0)}{d\xi_0} = & \left\langle \left(\frac{\partial V(x)}{\partial q_\xi} \right)_{\{q_{m \neq \xi}\}^{N-1}} \right\rangle_{\xi_0} - \left\langle \frac{1}{\beta} \left(\frac{\partial \ln |J|}{\partial \xi_0} \right)_{\{q_{m \neq \xi}\}^{N-1}} \right\rangle_{\xi_0} \\
& - \frac{1}{\beta} \left\langle \sum_{m \neq \xi}^{N-1} \left[\delta(q_m - l_{U_m}) \frac{dl_{U_m}(q_m)}{dq_\xi} - \delta(q_m - l_{L_m}) \frac{dl_{L_m}(q_m)}{dq_\xi} \right] \right\rangle_{\xi_0}
\end{aligned} \tag{2.31}$$

where V is the potential energy and q_ξ is the generalized coordinate so the first term represents the negative value of mean force exerted on the collective variable of interest and integration of the mean force along the domain of the collective variable produce the PMF. The second term is due to the volume scaling of the transformation from Cartesian to generalized coordinates. The last term is considered as Leibnitzian contribution, which accounts the interchange of integral and differential operators during Jacobian transformation. In our current study, since the Cartesian representation of collective variable is retained, the Jacobian and Leibniz terms vanish. For this case, Equation 2.31 returns to Equation 2.30, so the thermodynamic free energy equals to the PMF. Further, some sampling methods are discussed in order to overcome the free energy barrier. We will discuss these methods in the following parts.

2.8.1 Adaptive Biasing Force

Adaptive Biasing Force (ABF) method is a way to enhance the sampling of collective variable in the high free energy region in order to obtain the converged PMF profile. [139, 140, 141, 142] In the ABF method, a biasing force opposing the actual force arising from system components is periodically applied to the collective variable to generate what is effectively a random walk along the collective variable (purely diffusive dynamics). The ABF free energy gradient is estimated from the force (F_{ξ_0}), which is accumulated in small finite bins of width $\delta\xi_0$. [143, 144] The applied biasing force, which is along the collective variable ξ_0 to overcome free energy barriers, is calculated as:

$$F^{ABF} = \nabla_x \tilde{A}(\xi_0) = - \langle F_{\xi_0} \rangle_{\xi_0} \tag{2.32}$$

$\langle F_{\xi_0} \rangle_{\xi_0}$ denotes the current average of F_{ξ_0} along the collective variable. As the estimate of the free energy derivative $\nabla_x \tilde{A}$, is refined with more sampling over the course of the simulation, the biasing force F^{ABF} applied will compensate the system force. As a result, no net force will act along the collective variable ξ_0 over time allowing the whole system dynamics to be diffusive. Since we need a mean value of a property (force) that depends on a continuous variable, we must integrate over the probability density distribution function of the collective variable ($\langle F_{\xi}(\xi) \rangle = \int P(\xi) F_{\xi}(\xi) d\xi$). This distribution is represented by the aggregate of configurations generated from the MD simulation. To enhance sampling of the distribution of configurations where the collective variable holds a particular value, the collective variable is restrained within a certain narrow range (instead of its entire span). At the boundaries of the narrow range of interest, relevant restraint potentials are introduced on the collective variable in order to prevent it from moving outside of the desired range.

2.8.2 Umbrella Sampling

Umbrella sampling is another technique to improve the sampling of collective variable in configuration space. The basic idea underlying this method is to modify the potential function so that it can ensure the adequate sampling through the whole configuration space as shown in the following equation:

$$U_{\text{total}}(r) = U_{\text{unbiased}}(r) + U_{\text{biased}}(r) \quad (2.33)$$

where $U_{\text{unbiased}}(r)$ is the unbiased potential, $U_{\text{total}}(r)$ is the total potential after modification and $U_{\text{biased}}(r)$ is the biased potential. Usually, the biased potential is in the following form:

$$U_{\text{biased}}(r) = k(r - r_0)^2 \quad (2.34)$$

r_0 means the equilibrium state and k is the force constant. The choice of this force constant should be appropriate such that it can overcome the free energy barrier and at the same time the neighboring windows can have enough overlap of along the reaction

coordinate. From this protocol, the distribution of sampling configurations is non-Boltzmann. Weighted histogram analysis method then is applied to obtain the unbiased free energy profile. [145, 146]

2.9 Instantaneous Protein Interface and Interface Fluctuations

We discuss the protocol to construct liquid-vapor interface and protein-solvent interfaces. It has been previously explored by Willard and Chandler [147] that one could construct a coarse-grained solvent density field from the atomic coordinate in individual snapshot. Then the interface related to the solvent is defined as a constant density surface for the coarse-grained field in space. Specifically, in this work, we are interested in the water-vapor interface and water-protein interface. Therefore, water oxygen density field is constructed as follows: we set up a series of spatial grid points and compute the corresponding coarse-grained densities at space-time point \mathbf{r} , t , represented as $\bar{\rho}(\mathbf{r}, t)$ by Equation 2.35.

$$\bar{\rho}(\mathbf{r}, t) = \sum_i \Phi(|\mathbf{r} - \mathbf{r}_i(t)|; \xi) \quad (2.35)$$

where $\mathbf{r}_i(t)$ is the i th water oxygen atom's position in space and summation of each water molecule's density contribution in the whole space to this point yields the coarse-grained density of the particular grid point. Each water molecule's density contribution is modeled as a Gaussian function in Equation 2.36.

$$\Phi(\mathbf{r}; \xi) = (2\pi\xi^2)^{-d/2} \exp(-r^2/2\xi^2) \quad (2.36)$$

where r is the magnitude of \mathbf{r} , ξ is taken as 3.0 Å, and d stands for dimensionality (3 in this case). The final d dimensional density field will be constructed by acquiring each grid point's density. Then the interface is determined as the $(d - 1)$ -dimensional manifold with a constant value c . In practice, some differences arise to construct the liquid-vapor interface and liquid-protein interface in this work considering the shape of the liquid-vapor interface is flatter while protein-water interface possesses some curvature. Therefore, we select Cartesian coordinate system to construct the liquid-vapor

interface and spherical coordinate system for protein-water interface. For the liquid-vapor interface, coordinate (x,y,z) for each grid points in space is set up and the surface is obtained as the manifold by setting $\rho(x, y, z) = \rho_{bulk}/2$. That is, for a specific (x, y) coordinate set in 3 dimensional space, it defines a line which is parallel to the z axis. Along this line, if water density of one point satisfies condition $\rho(x, y, z) = \rho_{bulk}/2$, then this point is assigned to the interface. This instantaneous surface is denoted as $(h_t(x, y),$ at time $t)$. We can average these instantaneous surfaces to obtain the mean surface $\langle h(x, y) \rangle$ and furthermore, subtracting the mean values from the $h_t(x, y)$, we obtain $\delta h_t(x, y)$ as surface height and the height fluctuations $\langle \delta h^2(x, y) \rangle$. For protein-water interface, grid points in space are defined by (r, θ, ϕ) and for a specific (θ, ϕ) coordinate set in the spherical system, it defines a radial vector. r is the radial distance of end point of the radius vector from the origin $(0,0,0)$; θ is polar angle, which is defined as intersection angle between the radius vector and the positive z vector; and ϕ is azimuthal angle defined by the positive x vector and orthogonal projection of the radius vector on XY plane. The spherical coordinates (r, θ, ϕ) of a point could be derived from its Cartesian coordinates (x,y,z) by the following formulas: $r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$, $\theta = \arccos(\frac{z}{r})$ and $\phi = \arctan(\frac{y}{x})$. Points are defined to belong to the interface if $\rho(r_0, \theta, \phi) = 0.6\rho_{bulk}$. We use a different constant value c here compared with liquid-vapor interface case because this choice will result in a more unambiguous construction of protein-solvent interface. We note that other parameters, ξ and d remain the same as in the case of the liquid-vapor interface. Correspondingly, instantaneous protein interface can be expressed as $(h_t(\theta, \phi))$, mean surface as $\langle h(\theta, \phi) \rangle$, surface height as $\delta h_t(\theta, \phi)$ and height fluctuation as $\langle \delta h^2(\theta, \phi) \rangle$.

Chapter 3

EVALUATION OF EFFECTIVE HYDROPHOBICITY AROUND PROTEIN SURFACE AND IMPLICATIONS FOR BINDING PATCH THROUGH HYDROPHOBIC EFFECTS

Reproduced with permission from

Di Cui, Shuching Ou, Sandeep Patel. "Protein-Spanning Water Networks and Implications for Prediction of Protein-Protein Interactions Mediated through Hydrophobic Effects." *Proteins: Structure, Function, and Bioinformatics*. **2014**,82 (12), 3312-3326. Copyright © 2014, Wiley Periodicals Inc.

3.1 Introduction

It is generally accepted that water molecules play a crucial role for the stability of structures, dynamics and functions of biomolecules, such as proteins. [148, 149, 150, 151] One key feature of water molecules surrounding biomolecules is the tendency to connect and cluster with each other, forming branched networks arising from short-ranged and directional hydrogen bonding interactions; the broad description of these structures has enjoyed a rich history in the context of percolation theories in reduced dimensions [152, 153, 154, 155, 156]. It's been suggested that the continuous (dynamic) formation, dissolution, and rearrangement of water networks around protein surfaces is responsible for conformational transitions of the biomolecule. [157] The structures of water networks are essentially determined by water-water interactions and water-protein interactions. [154, 158] Due to the strong heterogeneity of protein surfaces in terms of both topography (local, as well as global, geometry and shape) and chemical composition, [15] the first hydration shell water molecules network arranges around the protein in a non-uniform manner. Water network structure is accommodated near some

regions while disfavored near other regions. Therefore, distribution of water molecules is dependent on local environment defined by the combination of water and protein residue density at a point in space in the vicinity of the protein/solvent interface. The local environment is thus defined by a collection of protein residues. For instance, a canonically-defined hydrophobic residue may represent a more or less hydrophobic environment right in the vicinity of its spatial location due to perturbations from its neighboring residues as discussed by Giombavtista et al[26] and others [27, 159]. These arguments and observations lead to the perspective of effective hydrophobicity of a protein surface residue, which warrants the view that the hydrophobicity of a group is context-dependent and thus a reflection of multiple effects of surrounding moieties on protein surfaces. [28]

Evaluating the effective hydrophobicity of protein surface regions is important since various observations indicate a correlation between effective hydrophobic regions, defined by groups of residues with high effective hydrophobicity, and binding patches of proteins that are predominantly mediated by hydrophobic effects for the association. [160, 161] Since a hydrophobic solute (generally of dimensions up to about 1 nanometer in diameter) is often considered (due to a variety of physical rationalizations) incapable of hydrogen bonding with water, water-water hydrogen bonds are reoriented along the surface of hydrophobe in order to minimize disruption of the three-dimensional hydrogen bonded network of water molecules. This leads to a structured water cage around the hydrophobe in the spirit of the classical Frank and Evans model.[8] The propensity of water molecules to predominantly adopt a subset of configurations to maximize interaction leads to significant loss of configurational entropy of water molecules. Such unfavorable effects can be minimized if hydrophobe molecules aggregate. Upon aggregation, water molecules surrounding the hydrophobic aggregate experience a loss in hydrogen-bonding network interactions (increased orientational and translational freedom); concomitantly, the solvent-exposed surface area of such aggregates is smaller than the sum of surface areas of individual solutes. This makes the entropic contribution less unfavorable and, hence, makes the free energy more favorable (though still not

necessarily dictating a state with minimum free energy under the appropriate external constraints). The hydrophobic effect is generally considered to be one of the driving forces for protein associations [162, 163, 91] as well as hydrophobic self-assembly of micelles, lipids, and lipid bilayers. In such association, one can consider regions of enhanced effective hydrophobicity associate under a mechanistic process described above. From the perspective of individual protein residues, those located within regions of higher effective hydrophobicity bear higher propensity to belong to binding interface(s) of the protein, associating with effectively hydrophobic residues from other proteins to avoid direct exposure in the aqueous medium. This is not to conclude that protein-protein association is predicated solely on interactions between effectively hydrophobic regions. The complexity of protein surfaces leads to multiple types of interactions. These include hydrogen bonding, ion pairing, interactions mediated by aromatic rings and the existence of these interactions may balance the ratio between hydrophilic residues and hydrophobic residues on protein interfaces, [161, 164] especially for those weak-binding proteins.

The idea of effective hydrophobicity offers a further way to view association of proteins in the weak binding limit. The description of weak-binding proteins is usually based on protein-protein interaction strength with an equilibrium dissociation constant K_d larger than 1 μM . The low-affinity nature of these proteins make them rapidly assemble and disassemble within a protein network which is important for mediating many cellular events. [165] It has been shown that the lower binding affinities for these proteins correspond to some structural characteristics involving smaller and less hydrophobic protein-protein interfaces. [166] These structural characteristics lessen the hydrophobic effect so that unlike obligate proteins, which exist in the form of complexes to avoid exposure of hydrophobic interfaces to solvent, single weak-binding proteins would be stable on their own in vivo. A consequent question involves distinctions of effective hydrophobic and hydrophilic regions on these protein interfaces and further evaluation of hydrophobic effects in mediating weak protein-protein interactions. In the present study, we will specifically focus on identification of effective hydrophobic

regions for small, weak-binding proteins.

A straightforward approach to identifying effective hydrophobic protein surface regions might exploit behaviors of first solvation shell water networks. Due to the hydrophobic nature, water networks should display a more significant propensity to form around the effective hydrophilic regions, while showing virtually no preference for the effective hydrophobic regions. Godawat et al [36] monitored the water density near the surfaces of fully solvated self-assembled monolayers (SAMs) with different functional groups exposed in aqueous solution and found that water density shows a poor distinction around hydrophobic head groups ($-CF_3$, $-CH_3$) and hydrophilic head groups ($-OH$, $-CONH_2$). Alternatively, it has been suggested that differences arise when considering the fluctuations of water density near the two regions. Enhanced fluctuations, reflected by the broad probability distributions of water number density are observed around effective hydrophobic surfaces compared with the bulk solution and effective hydrophilic surfaces. [28, 34] Moreover, the enhanced density fluctuations around hydrophobic surfaces could further be characterized by more compressible hydration shells and increased cavity formation, [167, 168] indicating that the nature of hydration shells around hydrophobic surfaces is softer and more flickering than that of hydrophilic ones. This approach has been further explored recently by Patel et al [35], who presented a more efficient method to estimate the cavity formation free energy to characterize effective hydrophobicity of the protein hydrophobin-II (HFBII). Furthermore, instead of considering water density fluctuation, Cui et al [169] distinguished the effective hydrophobicity of three different regions of HFBII based on the protein-solvent interface height fluctuations [169]. Although conceptually different from water density fluctuations, these studies reflect the malleable nature of hydration water structure around effective hydrophobic protein surfaces.

Another approach to map the effective hydrophobicity on protein surfaces is taken by Acharya et al, [27] who performed simulations of HFBII protein in aqueous solution with limited amounts of probe hydrophobic solutes and considered the local number density of small probe solutes in the vicinity of different regions of the protein.

Inspired by this, we propose a somewhat complementary, or “in like spirit”, protocol to identify ostensibly hydrophobic interaction regions of a protein with known three-dimensional structure, using local water number around solutes to scale the effective hydrophobicity. Instead of using actual, physical hydrophobic entities, we use the amount of water as probe, considering the distribution of water molecules around the hydrated protein surface. In theory, there should be an existence of a critical hydration level at which water network coverage will manifest a distinct variance around effective hydrophobic groups and effective hydrophilic ones. In this sense, we could determine the effective hydrophobicity around protein surfaces at this critical hydration level. Thus, one objective of this work is to study a single hydrated protein with different numbers of hydration water by computational simulation method so as to locate a proper hydration level that would identify the effective hydrophobic region.

This study is organized as follows. In Section 3.2 we present the computational details and selected proteins in this study. Our results are organized in Section 3.3, starting with probing the effective hydrophobic regions of two representative proteins: ubiquitin and HFBII and manifesting our whole approach; assessment of the protocol with other proteins and comparison of our identified effective hydrophobic regions with experimentally determined results are presented next. We finish with our conclusions and general discussion in Section 3.4 for this chapter.

3.2 Materials and Method

3.2.1 Selection of Proteins

Several types of protein-protein interactions fall into the category of weak-binding. Interactions between ubiquitin and ubiquitin-binding domains (UBDs), which are critical downstream of the ubiquitylation event, are typically weak with K_d around 10-500 μ M. It has been extensively characterized so far that the interaction surfaces on ubiquitin with UBDs involve a hydrophobic patch (residues L8-I44-H68-V70). Mutation of these key residues on ubiquitin [170] has suggested that these regions are involved intimately in the binding interface. In light of this, we select this protein as a

candidate to establish a protocol to identify this widely accepted effective hydrophobic binding region. This protocol would further apply to its binding partners, UBDs with diverse structures, including CUE domain, [171] UBA of DSK2, [172] GGA3 GAT domain, [173] and UBA of Human BMSC-Ubp [174]. Moreover, some proteins bearing similarities in structure and sequence to ubiquitin, defined as ubiquitin-like proteins, are also considered here. These proteins include Ubl-domain of HHR23A, [175] Ubl-domain of HHR23B [176] and NEDD8 [177]. Another kind of weak-binding protein is Src-homology-3 domain(SH3), which is found in the context of proteins involved in signaling pathways regulating several biological functions. [178] Several types of SH3 domain protein along with their binding partners are investigated in this study, including Sla1 SH3-3 domain, [179] Nck-2 SH3 domain and its binding partner Pinch-1 LIM4 domain, [180] CIN85 SH3-3 domain, [181] Abl SH3 domain and its binding partner Crk SH2 domain. [182] Histidine phosphocarrier protein (HPr) is a small protein playing a vital role in the process of phosphoryl group transfer in phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS). This function is largely determined by the fact HPr can form weak complexes with several enzymes in this process. Therefore, HPr protein is also one of the candidates in the study. Finally, as a control study, we consider to evaluate the effective hydrophobic region of an obligate protein HFBII that is present as a dimer in solution. HFBII is a protein expressed by filamentous fungi and is known for its ability to form a hydrophobic coating on the surface of an object and it can self-assemble into a monolayer on hydrophobic/hydrophilic interfaces. [183] These functions are mainly determined by the amphiphilic structural characterization. On one side of the protein, there is a large hydrophobic patch that consists of residues L7, V18, L19, L21, I22, V24, V54, V57, A58, A61, L62 and L63; other regions of the protein surface are generally hydrophilic. These distinctive regions offering different levels of effective hydrophobicity make HFBII an ideal protein for further testing of our protocol.

3.2.2 Simulation Details

Molecular dynamics simulations were performed with NAMD, version 2.9b3,[184, 185] using the CHARMM 22 all-atom force field (Chemistry at Harvard molecular mechanics) [186] with CMAP backbone torsion correction term.[187] Canonical ensemble (NVT) simulations were performed using a cubic cell with a box size $100 \text{ \AA} \times 100 \text{ \AA} \times 100 \text{ \AA}$, and periodic boundary conditions were applied. Since we are interested in small, weak-binding proteins (typically with numbers of amino acid residues less than 150), even with the consideration of first two hydration layers, the diameters of the hydrated proteins are no larger than 60 \AA . Therefore, the box size we applied is sufficient to avoid van der Waals (VDW) interactions from images. The initial structures of the proteins were obtained from Protein Data Bank [188] and prepared using CHARMM-GUI. [189] A single protein was placed in the center of the box and totally fixed during the simulation, surrounded by desired amounts of TIP3P model[190] water molecules. The number of water molecules was chosen to cover the percolation thresholds (further discussion below) of the proteins. Based on previous work from Brovchenko et al [156] on the determination of the water percolation threshold around a small globular protein, lysozyme, we test a range of solvating water numbers from $N_w = 200$ to 1000 to span across the percolation threshold of the proteins under study. If the proteins are net charged, opposite charges are uniformly distributed among all the atoms of proteins in order to make the system neutral during the simulation (for example, the total charge on protein Cue2 was $-5e$, so we add a charge of $+5e/780 \approx 0.0064e$ to each atom of the Cue2 molecule); this is an adaptation of the protocol of Brovchenko et al [156]. Constant temperature was maintained by Langevin bath at 300K, and the pressure was kept constant by Langevin pressure control at 1 atm. A switching distance of 8 \AA , non-bonded real-space cutoff of 9 \AA and pairlist generation distance of 10 \AA were used for the van der Waals interactions. The particle mesh Ewald (PME) method was employed for the calculation of conditionally-convergent electrostatic interactions.[191] The number of grid points of PME in x dimension is 100, in y dimension is 100, and in z dimension is 100 (as close to a 1 \AA grid point separation as

possible). The SHAKE algorithm [192] was used to constrain bond lengths involving hydrogen atoms and an integration time step of 2 fs was used. The trajectories were saved every 10 ps and the first 5ns was allowed for equilibration before a total of 50ns production data were generated for proteins with each hydration level. One snapshot of the simulation system in equilibrated state is shown in Figure 3.1A and B.

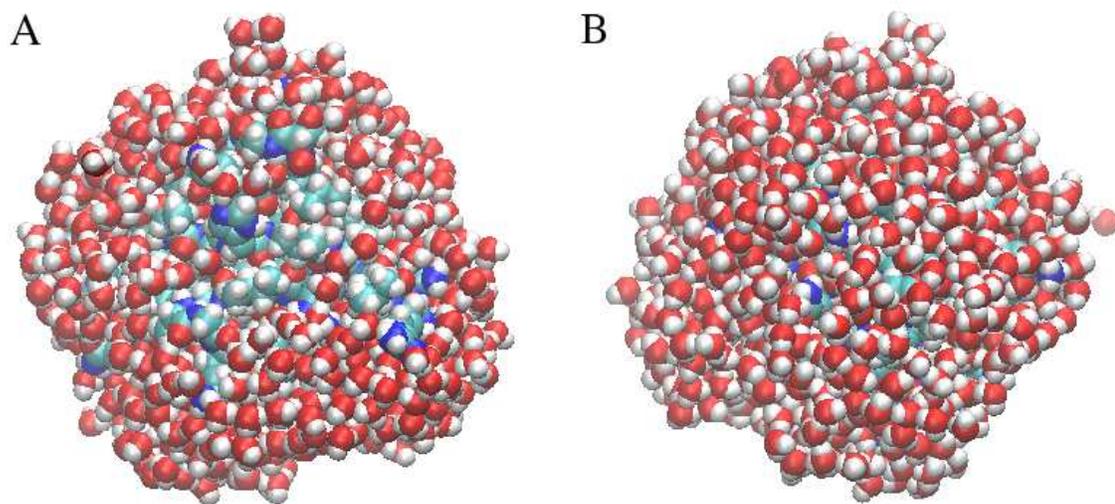


Figure 3.1: (A) Representative snapshot of ubiquitin solvated by 600 water molecules
(B) Representative snapshot of ubiquitin solvated by 600 water molecules
(rotated by 180° of Panel A).

3.3 Results

3.3.1 Water Networks around Ubiquitin

Previous reports indicate the existence of a minimum amount of hydration water for proteins to perform their biological functions [193], and this minimum amount of water usually correlates with the hydration level for the formation of the spanning water networks around protein surface. Inspired by this, we anticipate that the critical hydration level around a single protein surface that best distinguishes the effective hydrophobic and hydrophilic regions will also be related to the appearance of spanning water networks. That is, the hydration level (number of waters) that gives rise to an initial single, self-connected network of water molecules on the protein surface could be used as a filter for separating the most effectively hydrophobic and hydrophilic regions of the protein surface. That is to say, at the percolation threshold, where a single large networked water cluster is formed on the protein surface, the network will avoid effectively hydrophobic regions and cover effectively hydrophilic regions. Here we are probing this percolation transition by increasing the number of hydration water molecules; we are forcing a wetting of the protein surface. In a recent study, Patel et al [194] explore hydrophobic interfaces of proteins in the context of their dewetting behaviors upon external perturbing potentials. The authors suggest that effectively hydrophobic regions of a protein, when involved in interactions with other partners, will undergo dewetting. This is very similar to the percolation network avoiding the effectively hydrophobic regions of a protein as we consider in this work. This spanning water network could be further evaluated as the largest water cluster in the biological system as shown in previous publications. [195, 196, 156] Here we consider the largest water cluster within the first hydration shell of ubiquitin at different hydration levels including $N_w = 400, 500, 600, 700, 800$ and 900 . A water molecule belongs to the first hydration shell if the distance between its oxygen and the nearest heavy atom of the protein is $< 5.0 \text{ \AA}$, which is based on the minima in the pair correlation functions between the water oxygen and the heavy atoms of the protein as shown in Figure 3.2. A water cluster is defined by a continuous connection of water molecules by

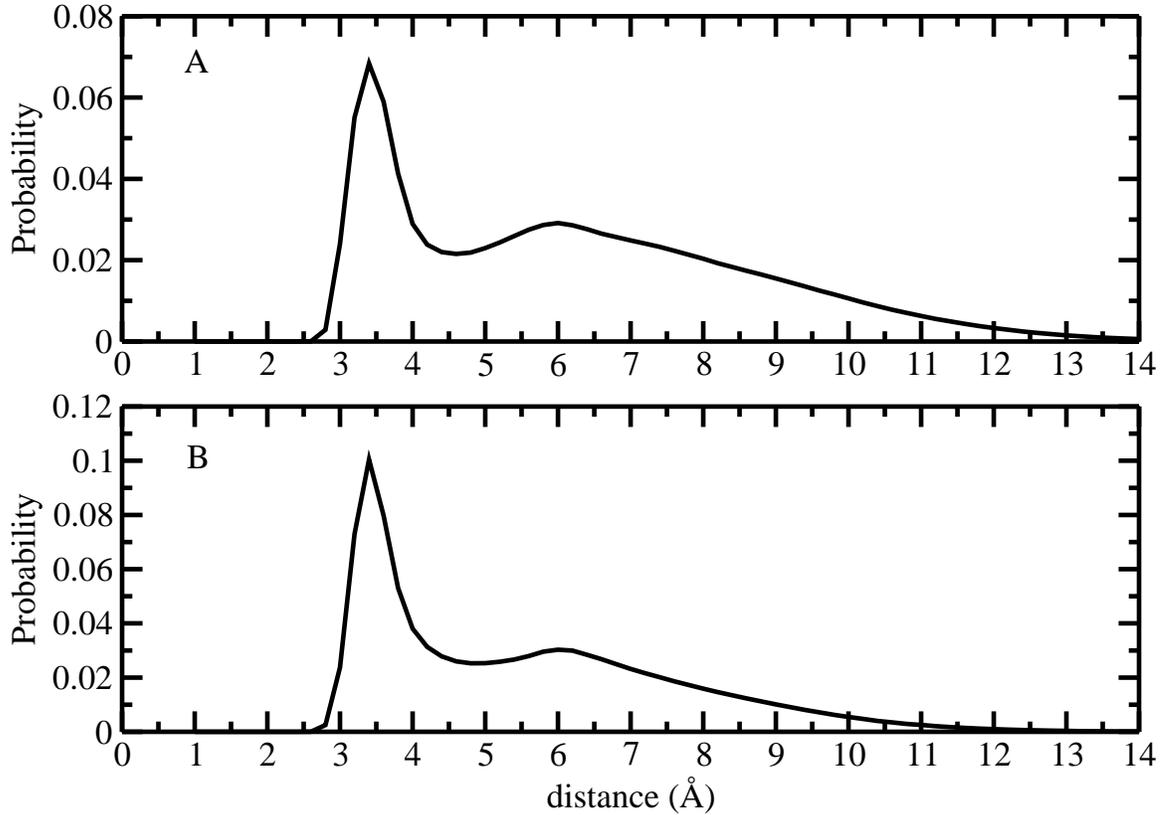


Figure 3.2: (A) Distribution of distance between water oxygen and the nearest heavy atom of protein ubiquitin (B) Distribution of distance between water oxygen and the nearest heavy atom of protein HFBII.

hydrogen-bonds. Two water molecules are defined as hydrogen-bonded by a combined distance-energy criterion such that the distance between the oxygen atoms is $< 3.5 \text{ \AA}$ and the water-water interaction energy is $< -2.7 \text{ kcal/mol}$. [197] The size of the cluster n , is considered to be the number of water molecules forming the cluster. [152] The methodology employed here is based on the literature of percolation theory as applied to aqueous networks [152].

An arrangement of water molecules belonging to the largest water cluster around a single ubiquitin with $N_w = 600$ is shown in Figure 3.3A. We consider the probability distribution $P(S_{max})$ of the size of the largest water cluster S_{max} in the system which can be obtained from the evolution of S_{max} with simulation time as shown in Figure

3.3B. $P(S_{max})$ of single ubiquitin at various hydration levels is shown in Figure 3.3C-H. Spanning and nonspanning largest clusters can be distinguished by considering that the left-hand peak of the distribution profile (Panel C) corresponds to the largest nonspanning water cluster, while the right-hand peak corresponds to the largest spanning water cluster (Panel G and H). In the middle range of hydration levels, the distribution profiles are bimodal; with increasing hydration levels from $N_w = 500$ to $N_w = 700$, the height of the right peak outweighs that of the left, indicating a transition between nonspanning to spanning water cluster, which is termed as the percolation transition. [198, 195] With $N_w = 500$, the probability of the spanning water cluster reaches about 50%, which is considered as the lower boundary of the percolation threshold; with $N_w = 700$, the probability of the spanning water cluster becomes dominant with a marginal peak for nonspanning cluster and a more pronounced peak for the spanning cluster, which is considered as the upper boundary of the percolation threshold. To further explore which hydration level is the critical one, we next consider the water density distribution around single ubiquitin with hydration level $N_w = 500$, $N_w = 600$ and $N_w = 700$ respectively. Our goal is to compare the hydration for the known effective hydrophobic region of this protein (reported as the hydrophobic binding patch involving L8, I44, G47, H68 and V70) at these three hydration levels in order to connect the $P(S_{max})$ distribution profile with an optimized hydration level at which to evaluate the effective hydrophobic region based on the water density distribution in the vicinity of a particular region.

3.3.2 Surface Water Density around Ubiquitin

Having determined the percolation threshold, we next discuss our approach to assess the water density in the local vicinity of a protein surface residue. We adopt a residue-based approach to characterize the water density distribution around single protein. First hydration shell water molecules are assigned to belong to the nearest amino acid residue around them; the distance between a water molecule and one amino acid is defined as the shortest separation between any of the heavy atoms of the residue

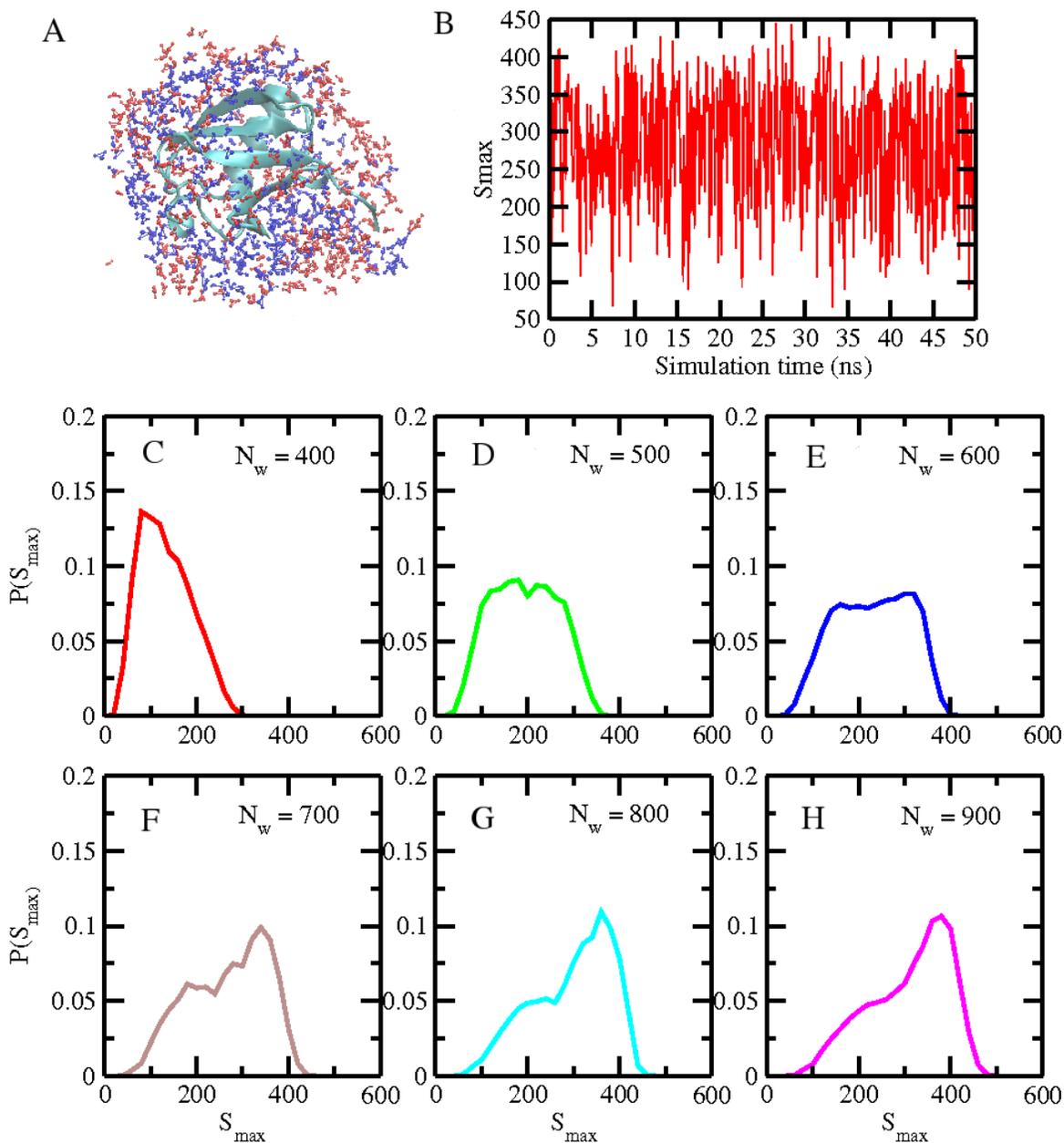


Figure 3.3: (A) Arrangement of water molecules on the surface of ubiquitin solvated by 600 water molecules. The water molecules that belong to the largest cluster are colored in blue; those of all other water molecules in red (B) Evolution of the largest water cluster size (C)-(H) Probability distribution $P(S_{max})$ for the largest cluster size S_{max} of the water molecules around ubiquitin protein surface at various hydration levels from $N_w = 400$ to $N_w = 900$.

and the oxygen atom of the water. Despite this approach for assignment, the presence of water molecules in a specific location in space is not solely affected by their nearest amino acid residue. As we have mentioned in the Introduction, this is actually context dependent and an effect resulting from all the possible interactions between a water molecule and residues, including the nearby residues and neighboring water molecules. The reason we adopt this residue-based assignment is that it allows us to compare our identified effective hydrophobic region with the literature reported hydrophobic binding patches of the proteins, which is usually based on the unit of residue. For each amino acid residue on protein surface, the surface area exposed to the solvent is dissimilar, leading to a natural bias in the number of hydration waters. To account for this, we define a surface water density to describe the hydration level around individual amino acid residues using the following relation:

$$N_{den} = \frac{\langle N_w \rangle}{S_{sasa}} \quad (3.1)$$

where N_{den} is the surface water density, $\langle N_w \rangle$ is the average number of water molecules around each residue and S_{sasa} is the solvent accessible surface area (SASA) of the residue, which is defined by Lee and Richards to measure the surface area of protein residues accessible by the solvent molecules within a specific radius of a probe molecule [199, 200]. Using CHARMM, the SASA for each residue was obtained by analyzing the initial structure of protein molecule using the “COOR SURFACE” module with a probe radius of 1.4 Å. Surface water density around each amino acid residue of ubiquitin is shown as a bar graph in Figure 3.4. Since we are interested in the location of effective hydrophobic region exposed on protein surface, we do not consider those residues buried inside of the protein with little to no SASA. In this sense, we only consider the residues with SASA larger than 20 Å². We note such a criterion could essentially give a similar selection of the exposed residues based on other online servers. [201] We note that this approach will lead to variations of the computed SASA for the same residue depending on the local environment of the residue; in our opinion, this is necessary as well for incorporating local effects. For these residues, a more straightforward way to

display the surface water density is to map this information around the three dimensional structure of the protein. Figure 3.5 qualitatively displays a colored map of the surface water density around different residues of the protein ubiquitin embedded in an aqueous medium with $N_w = 600$. Red colour depicts regions with low-hydration levels; while blue colour represents regions with high-hydration levels. Obviously, near the C-terminal there is a patch that is composed of low-hydrated residues, which is the effective hydrophobic region we seek; except for this area, other regions around the protein surface manifest a middle to high level of hydration. Considering the structural continuity of patches on the protein surface, a mathematically rigorous way to outline this region should be based on the cluster analysis of low-hydrated residues, defined by the residues with a surface water density less than $50 \times 10^{-3} \text{ \AA}^{-2}$. Based on the positions of centers of mass of these selected residues on protein surface, single-linkage clustering [202] is applied to identify their clustering (grouping) information. We turn to the clustering analysis next. Regarding the outlined protocol, we note that the choice of density is a free parameter, and can be selected on a by-protein basis. In this work, however, we use the same value of the threshold water density throughout. Furthermore, this value may depend on the molecular mechanics force fields used to describe both water and protein, and this should be kept in mind upon application of such a protocol.

3.3.3 Cluster Analysis to Identify the Effective Hydrophobic Interface

In this section, we discuss our approach for clustering residues with similar associated water densities, with the clustering being determined by a chosen level (distance) of spatial proximity (separation of residues). We use single linkage clustering[202], a simple, direct method, but appropriate for our purposes. For single-linkage clustering, each element (each residue associated with the threshold water density) to be assigned to a cluster is initially considered as a cluster of its own. Thus initially, the distance between two clusters is defined as the minimum separation between two elements (one in each cluster) in each amino acid residue with selected threshold water density. The

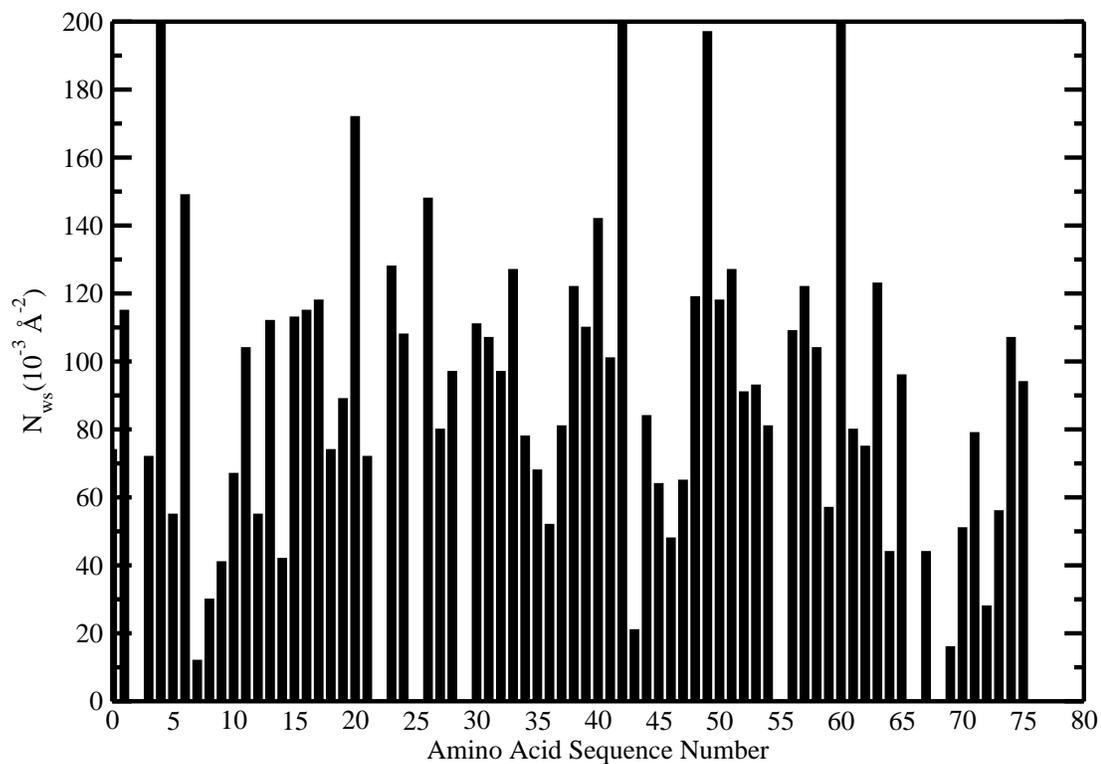


Figure 3.4: Surface area water density around each amino acid residue of ubiquitin protein at hydration level of $N_w = 600$.

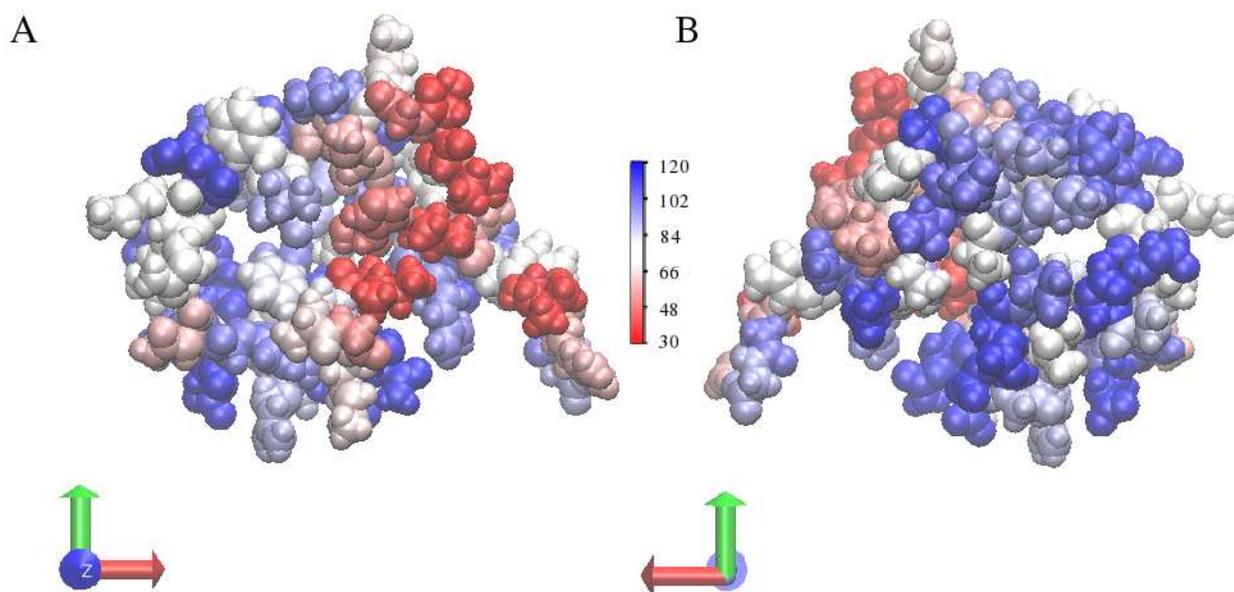


Figure 3.5: Water density map around ubiquitin with $N_w = 600$. (A) and (B) represent two side of the protein respectively with a rotation of 180° . Red colour indicates a lower water density; blue colour indicates a higher water density.

grouping of amino acid residues into clusters proceeds by defining a separation cutoff (minimum separation between elements of each cluster determined up to this point). With increasing separation cutoff, two existing clusters would merge into a larger one if the distance between these two clusters is less than the cutoff. Finally, all elements are classified into one single cluster as the cutoff becomes sufficiently large to encompass all elements. The clustering result can be visualized as a dendrogram shown in Figure 3.6. The X axis displays the sequence of amino acid residues and the Y axis depicts the cutoff separation distances used to assign elements into a single cluster. With a specific separation cutoff, the corresponding point on Y axis is located. A line that passes this point and is parallel to the X axis can be drawn and the number of intersection points between this line and the dendrogram corresponds to the number of clusters under this cutoff. The amino acid residues in the cluster correspond to those whose vertical line projections intersect the horizontal line representing the separation cutoff. As mentioned before, we would like to compare the identified effective hydrophobic region of this protein at different hydration levels around the percolation threshold including $N_w = 500$, $N_w = 600$ and $N_w = 700$. Figures 3.6A, B and C depict the clustering of low-hydrated residues of ubiquitin under these three hydration levels respectively. Based on the previous study, a cutoff distance of 8.0 Å was applied to define elements in space as a cluster.[30] At the hydration level of $N_w = 500$ (Figure 3.6A), residues F4, L73, V70, L71, G47, I44, H68, L8, K6, T9 and G10 can be considered to form a large patch, which is further represented in Figure 3.7A; at $N_w = 600$ (Figure 3.6B), residues G47, I44, H68, V70, L8, T9 and G10 constitute the hydrophobic cluster shown in Figure 3.7B; at $N_w = 700$ (Figure 3.6C), the cluster reduces its size to only contain three elements: L8, T9 and V70, as in Figure 3.7C. Figure 3.7D shows the literature reported hydrophobic interaction interface of ubiquitin including L8, I44, G47, H68 and V70. With $N_w = 600$, the identified effective hydrophobic region matches best with these residues. Therefore, $N_w = 600$ is the critical hydration level to distinguish the effective hydrophobic and hydrophilic region. At $N_w = 500$, low hydration level will limit the formation of a spanning water network that essentially could cover all

the hydrophilic regions around protein surface, so some of the hydrophilic residues that are close to the effective hydrophobic region are exposed on protein surface with lower water density leading to a false positive in the estimation of the effective hydrophobic region; at $N_w = 700$, water molecules will cover all the hydrophilic region, with the extra being retained around the effective hydrophobic region, leading to a false negative in the estimation of effective hydrophobic region. We notice that $N_w = 600$ corresponds to the hydration level in the middle of the percolation transition, where the probability of the largest spanning water cluster just outweighs that of the non-spanning cluster, as shown in Figure 3.3E. Besides using the probability distributions $P(S_{max})$ for the largest cluster size S_{max} around specific protein, another way to locate the percolation threshold is based on the probability distribution n_S of clusters with various size S . It has been reported at the percolation threshold, n_S and size S obey a power law $n_S \sim S^{-\tau}$, with the exponent termed as Fisher exponent. At percolation threshold, the Fisher exponent corresponds to a specific value for an infinite system in theory. Depending on the dimensionality, this value is 187/91 for a 2d infinite system and 2.18 for a 3d infinite system. [198] Around the real biomolecules, the H-bond based formation of the water cluster is finite. This value is sensitive to the parameters to define the water clusters. In this case, it is the H-bond criteria with distance criterion R_w representing the oxygen-oxygen distance and energy criterion U_w representing the water-water pair interaction energy. [195] Therefore, the specific Fisher exponent value to judge the percolation transition is ambiguous. In another approach, we judge the threshold based on the convergence of the Fisher exponent at various hydration levels. The identical representative protein ubiquitin was applied here. In the Figure 3.8, it displays the probability distribution n_S of clusters with size S around the protein at hydration levels ranging from $N_w = 400$ to $N_w = 800$ (Panel A to Panel E), with both S and n_S using log scale. The Fisher exponent at each hydration level can be obtained by fitting the distribution plot with S ranging from 0.5 to 1.5, corresponding to the linear region of the distribution profile. The fitting Fisher exponent of various hydration level is shown in Figure 3.8F. Apparently, a plateau region occurs starting from

hydration level $N_w = 600$, corresponding to the percolation threshold. This judgment of location of percolation threshold is consistent with the approach from distribution of largest water clusters in the system. Furthermore, the sampling interval between two successive hydration levels around a single protein surface is 100 water molecules in our current approach. To verify whether this resolution is sufficient to capture suitable hydration levels to distinguish the effective hydrophobic regions, we test the clustering of low-hydrated residues with more closely-spaced hydration levels between $N_w = 600$ and $N_w = 700$, including $N_w = 620$, $N_w = 640$, $N_w = 660$ and $N_w = 680$ in Figure 3.9. The outcomes suggest that the identified low-hydrated patches on protein surface at $N_w = 620$ and $N_w = 640$ are identical to the case of $N_w = 600$; while at $N_w = 660$ and $N_w = 680$, it is close to the result from $N_w = 700$. Therefore, under detailed hydration conditions, the identified effective hydrophobic clusters remain the same as the ones from our initial choice of hydration level for this test system.

3.3.4 Another Example: HFBII

As a further verification of the approach to identify effective hydrophobic interfaces originally developed from ubiquitin, we consider another example, the protein HFBII, which has served as a typical protein to characterize the effective hydrophobicity[27, 35]. Considering first the analysis of water networks, Figure 3.10 displays the probability distributions of the largest water cluster around HFBII at various hydration levels ranging from $N_w = 400$ to $N_w = 900$. The hydration level of $N_w = 500$, exhibits the dominance of the spanning cluster in the system. Therefore, we consider $N_w = 500$ as the critical hydration level to locate the effective hydrophobic patch for HFBII. Under this hydration level, the dendrogram of low-hydrated residues and surface water density map around the protein surface are shown in Figure 3.11 and 3.12. In the independent work of Patel et al. the authors show the hydrophobicity map of the identical protein HFBII in Figure 4 of that work [35]. Interestingly, both results indicate the existence of two effective hydrophobic regions around protein surface. In the present case, the larger one consists of residues L7, P8, T16, V18, L19, L21, I22, V24, A61, L62 and

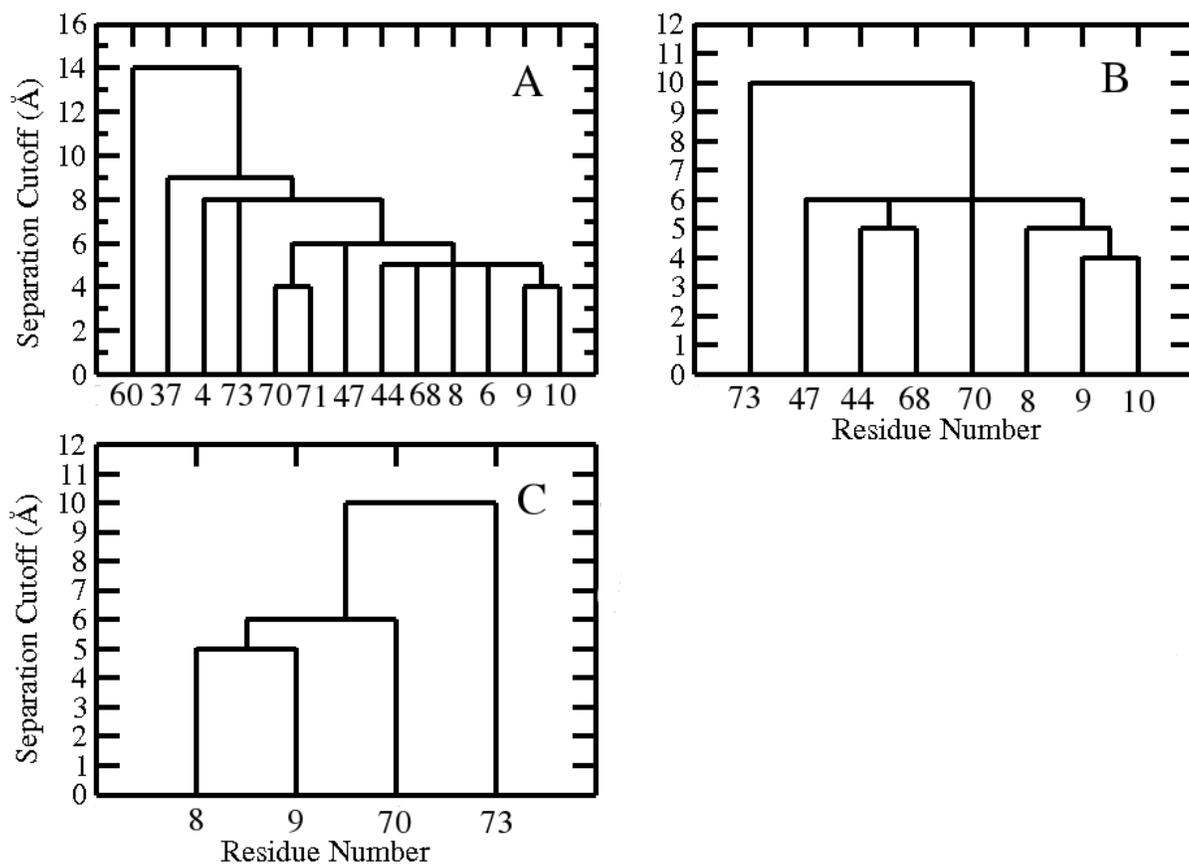


Figure 3.6: (A) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 500$ (B) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 600$ (C) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 700$.

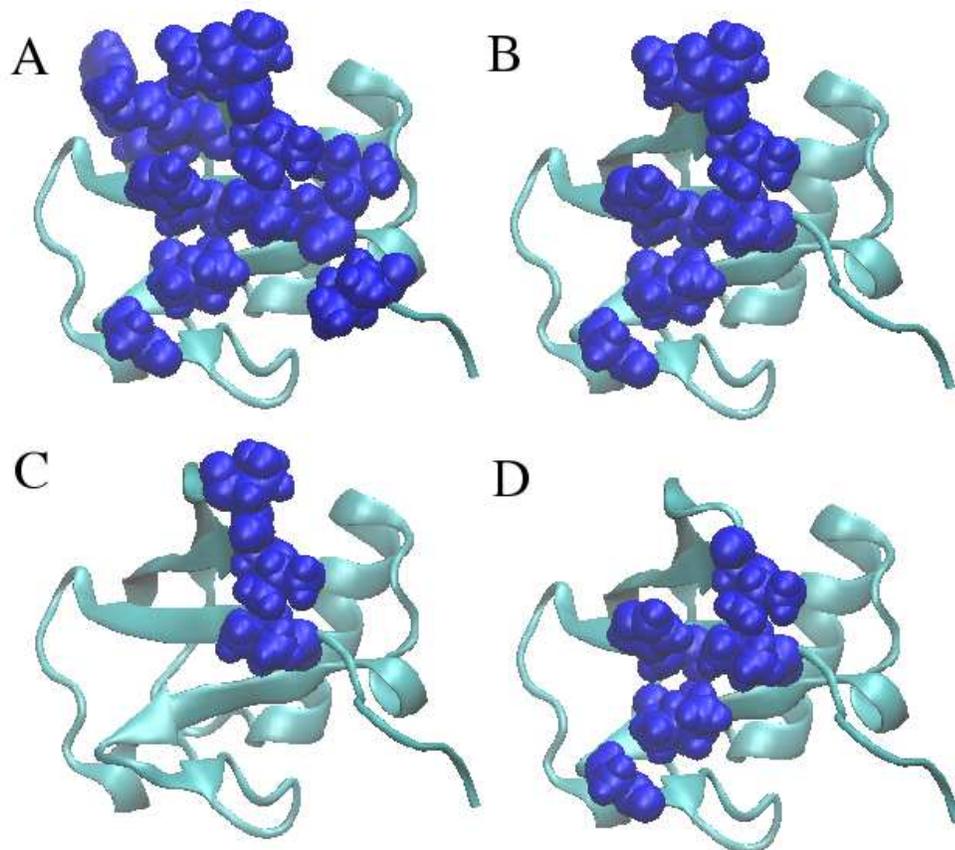


Figure 3.7: (A) Representative of the cluster of low-hydrated residues of ubiquitin with $N_w = 500$ (B) Representative of the cluster of low-hydrated residues of ubiquitin with $N_w = 600$ (C) Representative of the cluster of low-hydrated residues of ubiquitin with $N_w = 700$ (D) Representative of the reported binding patch for ubiquitin. In all cases, the residues involved are shown in blue colour.

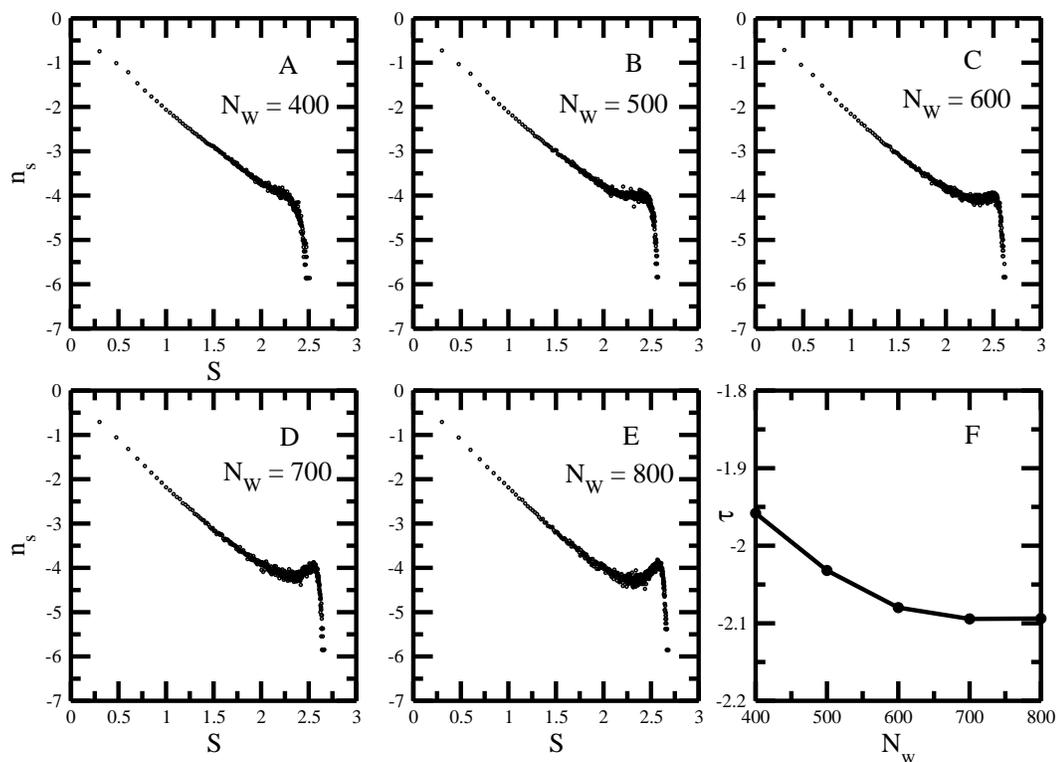


Figure 3.8: (A)-(E) Probability distribution n_S of water clusters with size S around ubiquitin at various levels from $N_w = 400$ to $N_w = 800$ (F) Fisher exponent (from fit of the cluster size distribution graph) τ of ubiquitin at various hydration levels.

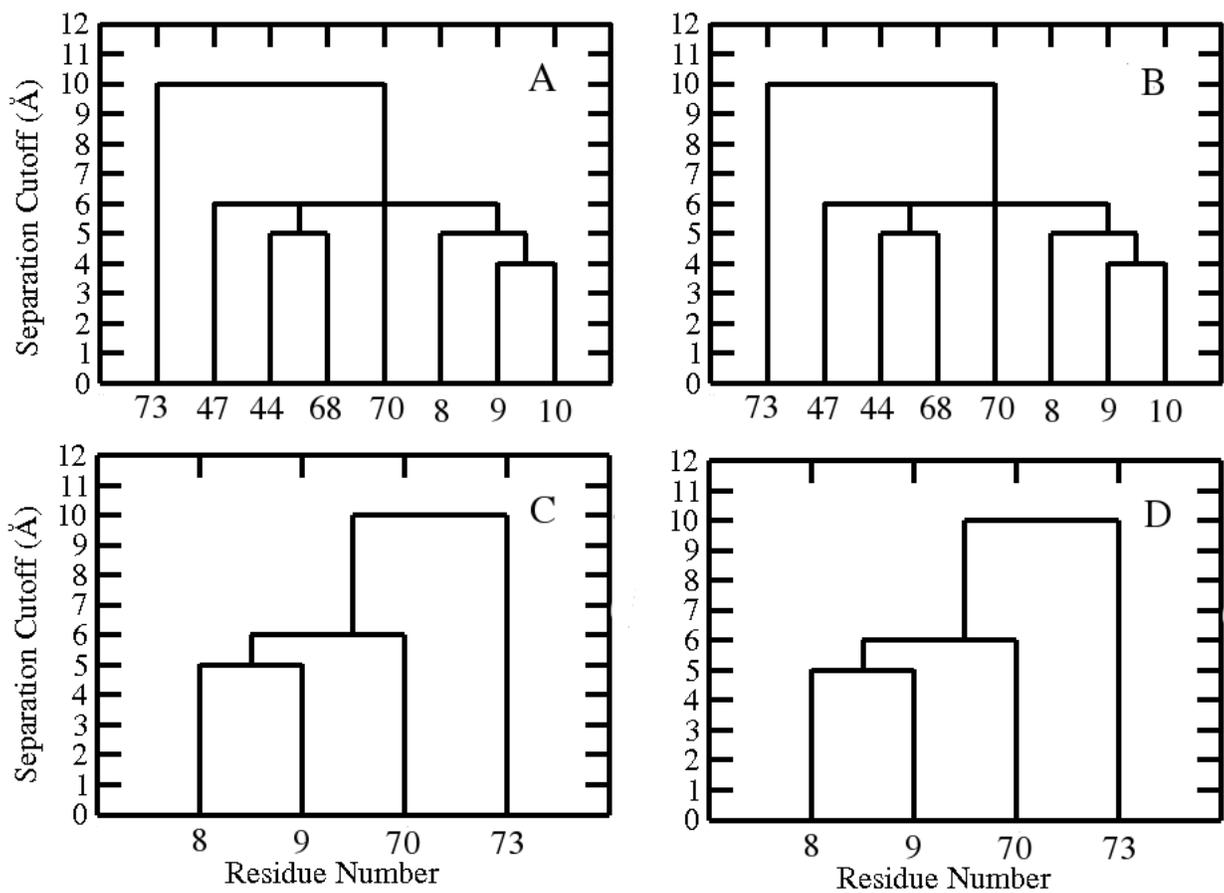


Figure 3.9: (A) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 620$ (B) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 640$ (C) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 660$ (D) Cluster analysis of low-hydrated residues of ubiquitin with $N_w = 680$.

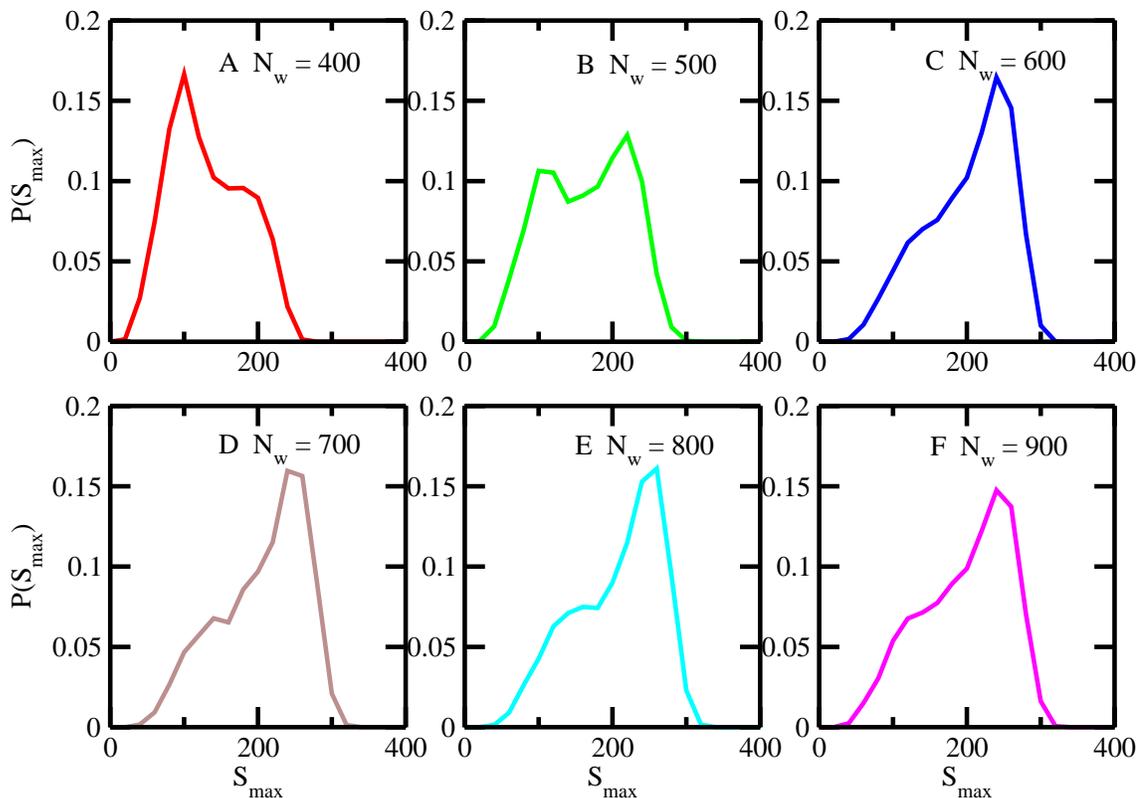


Figure 3.10: Probability distribution $P(S_{max})$ for the largest cluster size S_{max} of the water molecules around HFBI protein surface at various hydration levels. (A) $N_w = 400$ (B) $N_w = 500$ (C) $N_w = 600$ (D) $N_w = 700$ (E) $N_w = 800$ (F) $N_w = 900$.

L63 (red in Figure 3.12A), corresponding to the binding residues of this protein when it binds to other hydrophobic surfaces; the smaller one consists of residues T35, A37, I38, A41, A44 and S45, located at the other side of the protein (red in Figure 3.12B).

We pause here to further address minor aspects related to our algorithm here. To distinguish effective hydrophobic regions from hydrophilic ones in a most straightforward way, we choose the surface water density. In a previous study, it has been shown that around a fully hydrated surface, the water density itself will give a poor distinction. Instead, the fluctuation of the density could be a more relevant property. However, we are dealing here with a partially solvated surface with limited amounts of water in the system. Unlike previous studies with fully hydrated systems, water

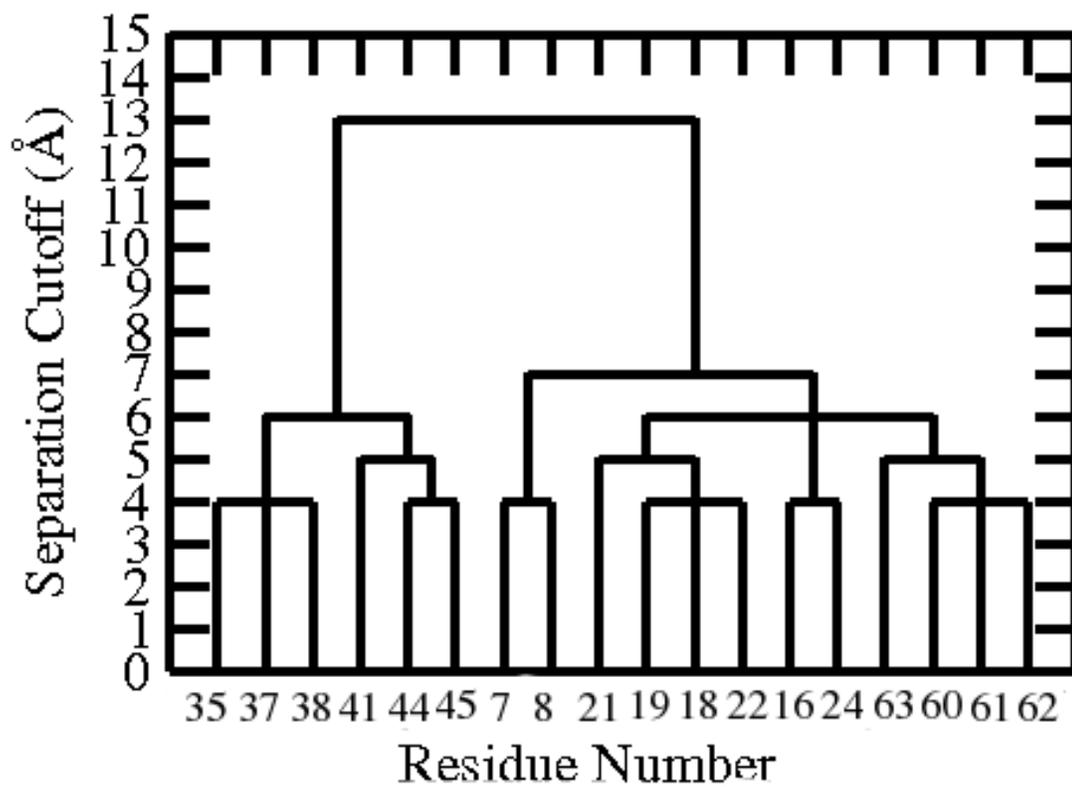


Figure 3.11: Cluster analysis of low-hydrated residues of HFBII at critical hydration level.

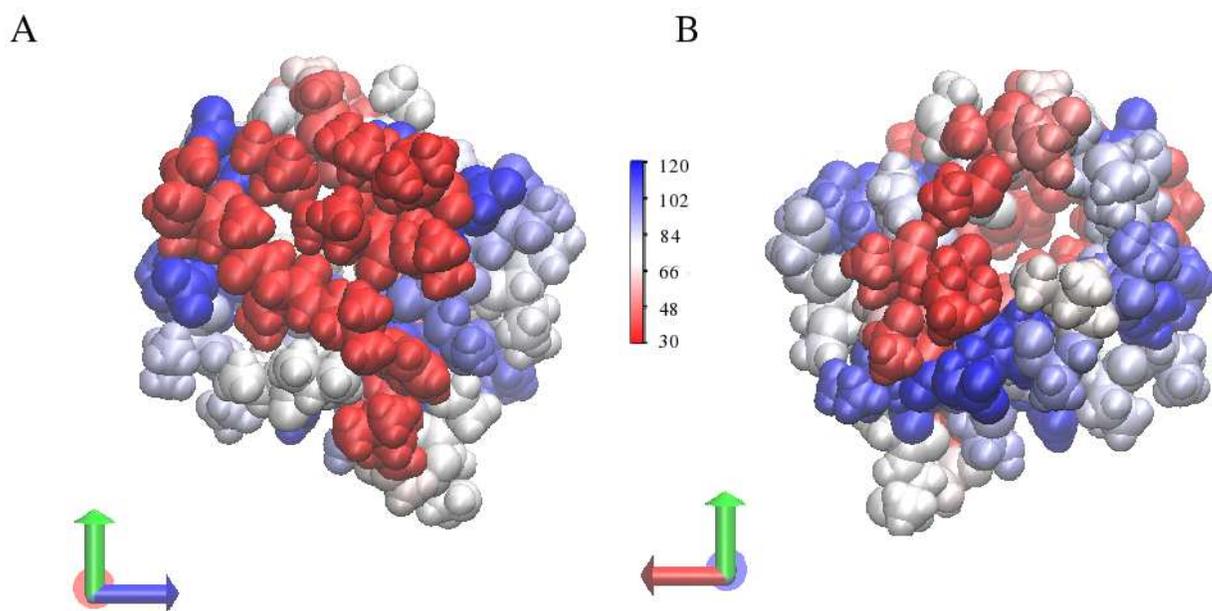


Figure 3.12: Water density map around HFBII with $N_w = 500$. (A) and (B) represent two side of the protein respectively with a rotation of 180° . Red colour indicates a lower water density; blue colour indicates a higher water density.

in our system should be able to arrange between hydrophobic and hydrophilic protein surface regions; there is no constraint that water density be equivalent at all positions along the protein surface. Based on this, we simply utilize the surface water density to map effective hydrophobicity. For a further connection with previous work, we also consider the fluctuation of water number around each residue on protein surface which is defined in the following equation.

$$N_{fl} = \frac{\langle N_w^2 \rangle - \langle N_w \rangle^2}{\langle N_w \rangle^2} \quad (3.2)$$

where N_{fl} is the normalized fluctuation of number of water around certain residues; $\langle N_w \rangle$ is the average number of water molecules as we previous defined. A map reflecting the fluctuation of number of water around ubiquitin surface at critical hydration level $N_w = 600$ is shown in Figure 3.13 with regions showing higher fluctuation colored as red and regions showing lower fluctuation colored as blue. Qualitatively, the hydrophobicity maps from our two approaches matched with each other, with lower surface water density region manifesting larger density fluctuation, which could be considered as effective hydrophobic region. For ubiquitin, this region includes L8, T9, G10, I44, G47, H68 and V70. A detailed comparison of the prediction results between these two approaches for more proteins can be found in Table 3.1. Due to the consistency of the results, in the following section, where we compare our predicted effective hydrophobic residues with the experimentally determined binding patches, only results based on the surface water density are listed.

3.3.5 Applications to Other Proteins

Next, we consider other proteins by identifying the effective hydrophobic patches and evaluating the hydrophobic effect in mediating their associations with other proteins. The proteins we probe in this paper are listed in the Selection of Proteins section. We note that such selection has considered some of the structural characteristics of these proteins. These proteins are all globular proteins small in size, with numbers of amino acid residues ranging from 60 to 150 and with nominal radius of gyration around

Protein	Based on surface water density	Based on fluctuation
Ubiquitin	L8 , T9, G10, I44 , G47 , H68 , V70	L8 , T9, G10, I44 , G47 H68 , V70 , L71
HFBII	L7,P8,T16, V18 , L19 , L21 , I22 , V24 ,Q60, A61 , L62 , L63	G6, L7, P8, T16, V18 , L21 , I22 , V24 ,Q60, A61 , L62
CUE domain	I15 , M19 , P21 , L41	I15 , M19 , N37, L39 , L41, L47
UBA of DSK2	M342 , G343, F344 , F345, Q362 , L365 , L369 ,N370,G371	M342 , G343, F344 , F345, L365 , L369 , G371
GGA3 GAT domain	F263 ,S267, L276 , L280 ,Q281	F263 , T269, L276 , G277, L280
UBA of Human BMSC-Ubp	M76 ,G77, I78 ,Q79,Q96, L99 ,F103, A104,P108	M76 , G77, I78 , Q96, L99 , L101, F103, A104
Ubl-domain of HHR23A	L10 , Q11 , Q12 , I49 ,G52, V73 , M75	I49 , Y50, G52, V73 , M75
Ubl-domain of HHR23B	L8 ,Q10, I47 ,A49,G50,K51, K67,N68,F69, V71 , M73	I47 , A49, G50, N68, F69, V71 , M73
NEDD8	L8 , T9, I44 , G47, V70 , L73	L8 , I44 , G47, H68
Pinch-1 LIM4 domain	I192 , R197 , P199	I192 , P199
Sla1 SH3-3 domain	Y362 , F364 , P406 ,Q408, F409	Y362 , F364 , L404, P406 , F409
Nck-2 SH3 domain	L203 ,Y204, V253	L203 , Y204, V253
CIN85 SH3-3 domain	I275, F276 ,Y278,I302,V304, P319 ,K324	I275, F276 , Y278, P319
Crk SH2 domain	P67 , P69 ,P70, V71 , P72 , P73, P75 ,A76,Q77,P78, P79,P80,G81,V82	P67 , P69 , P70, P72 , P73, P75 ,A76, Q77, P78, P79, P80, G81
Abl SH3 domain	V67, L69, Y70 , Y115	V67, L69, Y70 , V119
HPR	F48 , Q51 , T52	L47, F48

Table 3.1: Comparison of identified effective hydrophobic regions based on surface water density and based on water number fluctuation.

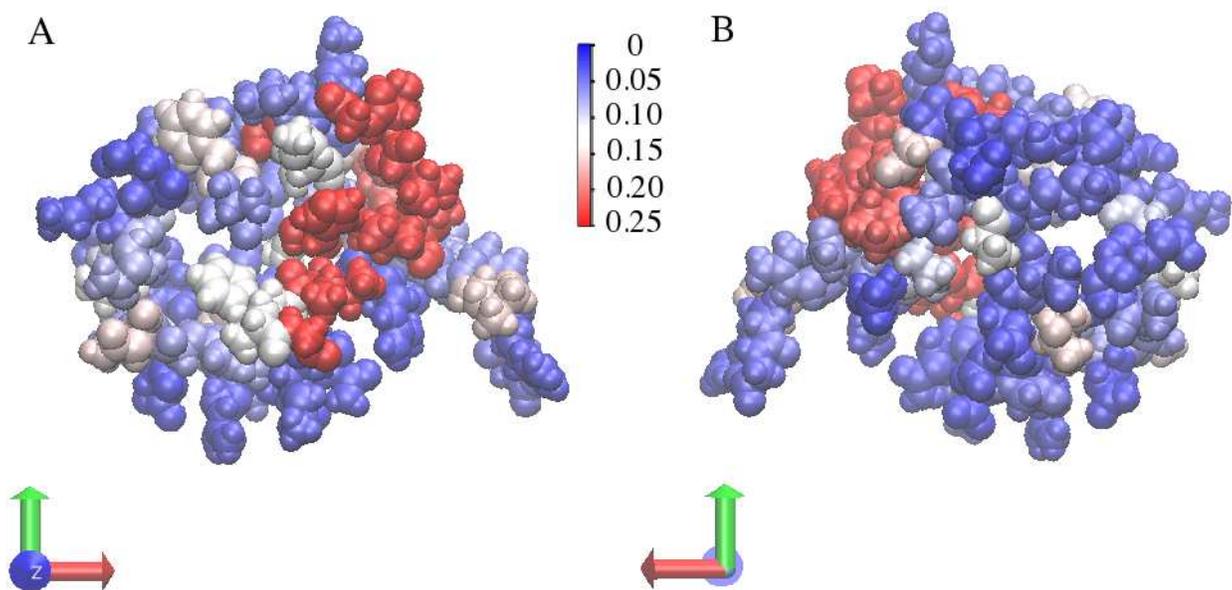


Figure 3.13: Map based on the fluctuation of number of water around ubiquitin surface at critical hydration level.

10 Å to 15 Å. These considerations in shape and size are necessary since it has been pointed out that there is an increase of hydration water density around small solutes compared with larger ones. [203] In light of this, the proteins we investigated are all comparable in size with ubiquitin to allow us to use the same criteria to define the low-hydrated residues. For some of the proteins we studied, the main globular structures have elongated tails at terminal regions, such as Ubl-domain of HHR23B, which has a long tail at C-terminal end that is composed of ten residues. The wire-like shape leads to poor solvation of the tail. These regions should not be considered as effective hydrophobic interface for the protein. A further test of this involves the remove of the tail for the initial structure to identify effective hydrophobic region. In Figure 3.14, we compare the water density map for the protein at the identical hydration level $N_w = 600$ with and without the long tails. In both cases, the same effective hydrophobic patch was identified.

Following our protocol, we first consider the largest water cluster distribution at various hydration levels for each protein, selecting the hydration levels corresponding to the case of just formation of the spanning cluster in the system. In Table 3.2, it displays the critical hydration levels to locate effective hydrophobic interfaces. Under these hydration levels, the effective hydrophobic regions were identified based on selection of low-hydrated and surface-exposed residues and further clustering them in space. The corresponding dendrogram for each protein is shown in Figure A.1-A.14 in the Appendix A. The summarized results for the elements of effective hydrophobic patch for each protein along with the literature reported residues that are responsible for the hydrophobic association of the protein are listed in Table 3.3. Here, we emphasize that hydrophobic interaction may serve as a key contribution for these proteins to bind with others, but it is not the sole contributor. Therefore, the listed reference residues involved in hydrophobic association in the third column may not cover all binding sites for the proteins. For example, for the interactions between ubiquitin and ubiquitin

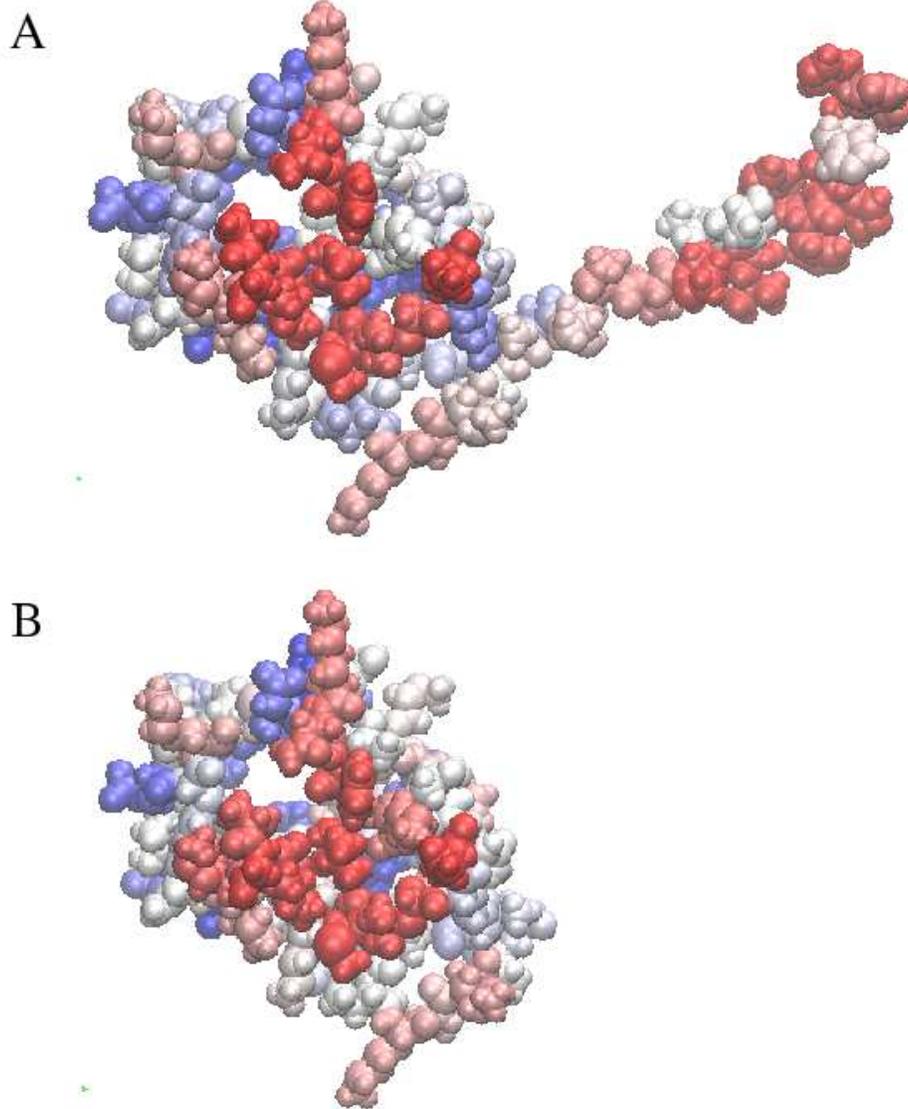


Figure 3.14: (A) Water density map around protein Ubl-domain of HHR23B at critical hydration level $N_w = 600$ with long tail around C-terminal (B) Water density map around protein Ubl-domain of HHR23B at critical hydration level $N_w = 600$ without long tail around C-terminal.

Protein	Critical hydration levels
Ubiquitin	600
HFBII	500
CUE domain	500
UBA of DSK2	400
GGA3 GAT domain	700
UBA of Human BMSC-Ubp	300
Ubl-domain of HHR23A	600
Ubl-domain of HHR23B	600
NEDD8	600
Pinch-1 LIM4 domain	500
Sla1 SH3-3 domain	500
Nck-2 SH3 domain	600
CIN85 SH3-3 domain	500
Crk SH2 domain	600
Abl SH3 domain	400
HPR	500

Table 3.2: Critical hydration level to identify effective hydrophobic regions for each protein.

interaction motifs (UIM), besides the hydrophobic interactions mediated by L8-I44-H68-V70 patches, residues ARG42, ARG72 and ARG74 involve extensive hydrogen-bonded interactions with some of the GLU residues of UIM. [170, 204] Also, residues GLU233 and ASN250 from Nck-2 SH3 domain are responsible for the hydrophilic interactions with Pinch-1 LIM4 domain. [180] These residues were not included in the reference column for the comparison of effective hydrophobic regions. Besides, it has been previously discussed by Winget et al [205] using ubiquitin as an example that protein-protein recognition is a complicated issue involving specific interactions with different binding partners. However, it has also been noted that although there exist distinct binding sites, a single protein actually possesses some conserved binding site motifs that are repeatedly used with different binding partners. [206] The reference binding locations we present in the table are based on these common motifs, considered as key residues.

Protein	Identified patches	Reported patches
Ubiquitin	L8 ,T9,G10, I44 , G47 , H68 , V70	L8,I44,G47,H68,V70[170]
CUE domain	I15 , M19 , P21 ,L41	I15,M19,P21,L39,I43,L47[171]
UBA of DSK2	M342 ,G343, F344 ,F345, Q362 , L365 , L369 ,N370,G371	D341,M342,F344,V361,Q362, L365,L369[172]
GGA3 GAT domain	F263 ,S267, L276 , L280 ,Q281	F263,A266,L276,L280[173]
UBA of Human BMSC-Ubp	M76 ,G77, I78 ,Q79,Q96, L99 ,F103, A104,P108	M76,I78,L99[174]
Ubl-domain of HHR23A	L10 , Q11 , Q12 , I49 ,G52, V73 , M75	L10,Q11,Q12,K47,I49,A51,K53, I54,V73,M75,T77,K78[175]
Ubl-domain of HHR23B	L8 ,Q10, I47 ,A49,G50,K51,K67, N68,F69, V71 , M73	L8,I47,V71,M73[176]
NEDD8	L8 ,T9, I44 ,G47, V70 ,L73	L8,I44,V70[177]
Pinch-1 LIM4 domain	I192 , R197 , P199	I192,R197,R198,P199[180]
Sla1 SH3-3 domain	Y362 , F364 , P406 ,Q408, F409	Y362,F364,W391,P406,F409[179]
Nck-2 SH3 domain	L203 ,Y204, V253	L203,V253,V254[180]
CIN85 SH3-3 domain	I275, F276 ,Y278,I302,V304, P319 ,K324	F276,W306,P319,F322[181]
Crk SH2 domain	P67 , P69 ,P70, V71 , P72 ,P73, P75 , A76,Q77,P78,P79,P80,G81,V82	P67,P69,V71,P72,P75[182]
Abl SH3 domain	V67, L69, Y70 , Y115	Y70,F72,W99,W110,Y115[182]
HPR	F48 , Q51 ,T52	T16,R17,L47,F48,Q51[207]
HFBII	L7,P8,T16, V18 , L19 , L21 , I22 , V24 ,Q60, A61 , L62 , L63	V18,L19,L21,I22,V24,A61,L62, L63[183]

Table 3.3: Identified effective hydrophobic patches for various kinds of proteins.

For some of the proteins, more than one hydrophobic patch around single protein surface were identified, such as in the case of HFBII as previously discussed, UBA of Human BMSC-Ubp (Figure A.4 in the Appendix A) and Crk SH2 domain (Figure A.12 in the Appendix A). Considering the sizes of the proteins we investigated, usually there is only a single hydrophobic interfacial region involved in protein-protein association. Therefore, only the components from the largest hydrophobic cluster are listed here to compare with experimental results. Those identified residues that match the ones belonging to the hydrophobic binding patch are presented in bold-face type in the Table 3.3. Taking into account the types of residues that were identified as elements of effective hydrophobic region, most of the residues belong to the strong hydrophobic residues (LEU, ILE, VAL, ALA) defined by Kyte-Doolittle hydrophathy values. However, a small portion of residues that are traditionally classified as hydrophilic residues based on Kyte-Doolittle hydrophathy scale, such as R197, P199 in Pinch-1 LIM4 domain and Q11, Q12 in Ubl-domain of HHR23A, are detected as effective hydrophobic residues, which reflects the key point about the importance of considering the local context of a residue into account in evaluating its hydrophobicity.

For further assessment of the identification results, we consider the accuracy and coverage of our predictions. We use metrics presented by Zhou et al[1].

$$Cov = \frac{TP}{RI} \quad (3.3)$$

$$Acc = \frac{TP}{TP + FP} \quad (3.4)$$

where Cov is the coverage; Acc is the accuracy; TP is the number of true predictions based on our approach; RI is the number of real residues involved in the hydrophobic association according to the literature, which is the reference; FP is the number of false positives in the prediction. Table 3.4 presents these results for various proteins. Overall, our identified effective hydrophobic regions match with the key reported binding residues of these proteins involved in hydrophobically-mediated associations, with an average coverage around 75%. For the generation of the false positives, like residues

K6, T9 and G10 in ubiquitin and residues P8, T16 and Q60 in HFBII, we attribute this to neighbor effects. These residues usually appear around the key hydrophobic binding patches so the water densities around them are largely affected by the hydrophobic local environment nearby. As a result, they were considered as a part of effective hydrophobic patches. A possible solution to reduce these false positive predictions may introduce a polarizable force field to better distinguish water densities around these residues and the effective hydrophobic residues in future work. Less false negatives are generated in our prediction as indicated by the high coverage, which may be related to the fact that we have already reduced our target area to the effective hydrophobic binding regions.

Another comparison may involve our binding prediction and prediction from servers online, which are usually based on bioinformatics information like protein sequence conservation, secondary structures, solvent accessibility and so on. [1] Here, we acknowledge the power of these well developed servers providing information of all the possible binding sites for the proteins without any biases in emphasizing the key spots, which displays a high coverage and low accuracy in prediction of the residues involved in the hydrophobic association as shown in Figure 3.15. In contrast, our approach possesses the advantage to pick out the conserved effective hydrophobic binding sites which is essential in mediating protein-protein association driven by hydrophobic effects. The increase in coverage is not at the expense of large sacrifice of accuracy as shown in Figure 3.15. Compared with other web servers, our method is sufficiently robust to select the possible key hydrophobic binding sites with little interference from the false positives, which suggests that the piece of effective hydrophobicity information of residues around protein surface may need to be accounted for in the future development of more advanced binding interface prediction algorithms.

3.4 Summary

We presented a method that exploits water network percolation behavior in the first solvation shell of small proteins in order to predict clusters of residues potentially

Proteins	Coverage(%)	Accuracy(%)
Ubiquitin	100	72
CUE domain	50	75
UBA of DSK2	72	56
GGA3 GAT domain	75	60
UBA of Human BMSC-Ubp	100	33
Ubl-domain of HHR23A	50	86
Ubl-domain of HHR23B	100	36
NEDD8	100	50
Pinch-1 LIM4 domain	75	100
Sla1 SH3-3 domain	80	80
Nck-2 SH3 domain	67	67
CIN85 SH3-3 domain	50	29
Crk SH2 domain	100	36
Abl SH3 domain	40	50
HPR	40	67
HFBII	100	67

Table 3.4: Coverage and accuracy of various approaches of predictions. Based on the reference, [1] the coverage is defined as $Cov = TP/RI$ and the accuracy is defined as $Acc = TP/(TP+FP)$. TP is the number of true prediction; RI is the number of real residues involving in the hydrophobic association according to the literature; FP is the number of false positives in the prediction.

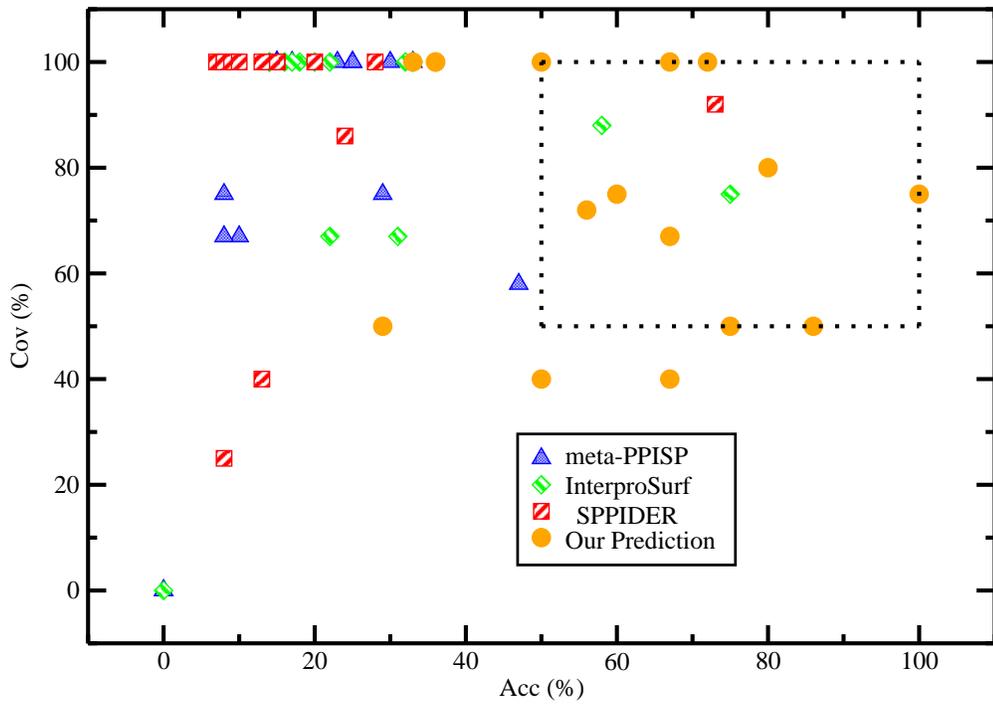


Figure 3.15: Prediction coverage vs prediction accuracy of different approaches.

involved in binding interactions with other proteins. The binding interfaces we focus on rely in some part to hydrophobic characteristics of the residues involved. The novel idea here is that we consider hydrophobicity of a particular residue not solely based on its lone properties, but rather based on an analysis that includes the local chemical context of the residue. That is, we consider an effective hydrophobicity of a residue that is dictated by the character of neighboring residues as well as local water. Though simplistic in principle and spirit, this method is able to predict with significant accuracy and coverage the binding interaction residues for a series of small proteins. The results of our work are consistent with previous studies that consider water density fluctuation based approaches for characterizing local hydrophobicity of protein surface regions. Furthermore, a central component of the algorithm presented is a critical percolation threshold of solvent of the first hydration shell of the model proteins selected for this work. This phenomenon connects with recent molecular simulations suggesting that biological molecules and molecularly hydrophobic interfaces exist in thermodynamic states on the border of dewetting transitions. These states are sensitive to perturbations (chemical, environmental) which can modulate and/or fine tune the nature of interactions of these interfaces with other interfaces or molecules in order to effect or inhibit biological function [194]. Taken together, the composite picture is one suggesting the importance of accounting for local chemical environment when characterizing the hydrophobicity of residues in conjunction with solvent density fluctuations in the vicinity of hydrophobic regions giving rise to tunable propensities for wetting and dewetting these critical biochemical interfaces.

Using ubiquitin as an example, we developed a protocol to identify the effective hydrophobic interface by first determining the critical hydration level at which percolation transition of the water network occurs. The approach is adapted from percolation theory, and we have isolated a protocol which determines the percolation threshold by finding a hydration level where a unimodal distribution of the probability of largest cluster transitions (biased to small cluster sizes) to a bimodal distribution. This point,

the percolation threshold, gives rise to a networked, spanning water cluster which effectively seeks less hydrophobic regions of the protein surface in order to maximize favorable interactions. The differences in water density around particular residues can then be used to cluster the residues based on a cutoff value of density in order to cluster residues which are putatively hydrophobic. Using this approach we have studied the proteins UBDs, ubiquitin-like proteins, SH3 domain. We have also compared our predictions of binding patch residues to those from automated servers (SPPIDER, InterproSurf and meta-PPISP). We find that the current method is competitive in terms of the average accuracy 60% and the average coverage 75% across the series of proteins studied.

Chapter 4

ION-SPECIFICITY AROUND EFFECTIVE HYDROPHOBIC REGIONS OF PROTEIN SURFACE

Reproduced with permission from

Di Cui, Shuching Ou, Eric Peters, Sandeep Patel. "Ion-Specific Induced Fluctuations and Free Energetics of Aqueous Protein Hydrophobic Interfaces: Towards Connecting to Specific-Ion Behaviors at Aqueous Liquid-Vapor Interfaces." *Journal of Physical Chemistry B*. **2014**, *118* (17), 4490-4504. Copyright © 2014, American Chemical Society

4.1 Introduction

The fundamental nature of interactions between ions, co-solutes, and proteins in aqueous solutions continues to garner attention [37, 38, 39] due to its importance in understanding protein denaturation, folding, protein-protein interactions to name a few examples. In the context of protein denaturation, Hofmeister effects or ion-specific effects, related to the modulation of surface tension and protein solubility by additive salts that influence the strength of direct and water-mediated interactions in solution have been intensely explored with the ultimate aim of extracting basic physical insights into the above mentioned processes[40, 41, 42, 43]. At the heart of specific-ion effects as related to protein denaturation is the molecularly-resolved interface between protein and aqueous solution; moreover, the nature of the differences in behavior of cations/anions at such interfaces (including both liquid-vapor interfaces and liquid-solute interfaces) weighs heavily on the interpretation and definition of these processes. Now amassed is a vast literature that discusses specific-ion effects as embodied in differential stabilities of halide ions at liquid-vapor interfaces [45, 46, 47, 48, 49, 50, 51, 52].

It has been widely shown that larger halide ions such as I^- and Br^- tend to bind to liquid-vapor interfaces more strongly and with lower transfer free energies than smaller, more charge-dense, and more strongly-hydrated Cl^- and F^- anions. The microscopic origins and molecular mechanisms of these behaviors are concerned with several factors ranging from ion size, ion polarizability and ion hydration properties to solvent polarizability.[53] Recent studies [55, 56, 57] have begun to consider differential perturbations of liquid-vapor interface fluctuations by different anions. Ou et al studied ion-specific effects at the aqueous liquid-vapor interface by exploring ion-induced interfacial fluctuations in the case of two chemically distinct anions Cl^- and I^- , which represent the neutral and chaotropic positions in the Hofmeister series, on distant liquid-vapor interface. [56, 57] They observed that the more surface stable I^- anion (as observed elsewhere [208, 51, 50, 55]) induces larger interfacial fluctuations compared to the non-surface active species Cl^- , thus demonstrating a strong correlation with induced interfacial fluctuations and anion surface stability as observed from molecular simulations. The authors trace these differences in induced interfacial fluctuations by Cl^- and I^- to the nature of the hydration environment around the anions; water molecules in the hydration shells of I^- are shown to be more dynamic and less persistent compared to those in proximity to Cl^- . When approaching the liquid-vapor interface, coupling of local solvent around anions with solvent further away and near an interface leads to different perturbations of the interface by the two anions, and thus different contributions to interface height fluctuations, and ultimately surface stability via contributions from interfacial entropy arising from surface fluctuations correlations[55, 57, 56].

This ion-specific effect is not necessarily restricted to the liquid-vapor interface; one might consider how the perturbation-inducing properties of the two anions may play out generally in the vicinity of hydrophobic interfaces. Heyda et al.[59] examined systems of N-methylacetamide (NMA) in the presence of monovalent cations and anions in water. The larger anions, I^- and Br^- , demonstrated preferential spatial correlation with the hydrophobic methyl group, which supports earlier experiments addressing the

importance of the nonpolar methyl groups for the halide ion-NMA interactions. [209] Horinek et al. investigated the potential of mean force (PMF) for Na^+ , Cl^- , Br^- and I^- to transfer from bulk aqueous solution to a hydrophobic self-assembled monolayer-water interface in an infinite dilution. [210] Similarly, soft polarizable monovalent anions (I^- and Br^-) prefer to accumulate around the hydrophobic interface. In another contribution, Lund et al. probed the distribution of F^- and I^- around a spherical macromolecule. [60] They found that when the nanosphere is uncharged and considered as a hydrophobic particle, F^- ions are repelled while I^- ions are weakly attracted to it. In a recent molecular simulation study, Friedman et al [211] analyzed extensive molecular dynamics simulations of three proteins in aqueous salt solutions. The authors concluded that binding of cations and anions to protein surfaces is heterogeneous, with the same amino residue demonstrating a wide range of binding probability to a particular ion. This heterogeneity stems from the heterogeneous environments found on protein surfaces. As pointed out by Giovambattista et al [26] and others [27, 159], the local environment of any given amino acid residue is largely perturbed and defined by its neighboring residues. Jungwirth and coworkers have further provided volumes of data on the nature of differential, or ion-specific, binding of cations and anions to protein surfaces. [61, 62, 63] Specifically, using lysozyme as an example, they indicate that in the mixed aqueous solution of KCl and KI, I^- is preferential to be in close vicinity of the hydrophobic groups. [212, 213] Furthermore, this specific-ion effect may play a crucial role in modulating protein-protein interaction in solution. [66]

Since there is implied a connection of the behaviors of ions at aqueous liquid-vapor interfaces to those of possibly biochemical relevance (protein-water, bilayer-water, etc)[58], we seek to begin to address connections with particular focus on hydrophobic regions of proteins (to use a model system that is a natural extension of the ideally hydrophobic aqueous liquid-vapor interface). We propose to consider how anions, in particular Cl^- and I^- , induce fluctuations at the interface around hydrophobic patch of a rigid protein in aqueous environment. We also seek to make connection of observed induced interfacial fluctuations to the free energetics (probabilities) of the

two types of anions near the hydrophobic protein region. We anticipate that similar qualitative trends and behaviors should arise in the biomolecular context as observed for aqueous liquid-vapor interfaces. We note that unlike the liquid-vapor interface, the protein-water interfaces are more complicated because of their inherent chemical and topographical heterogeneity. The heterogeneities account for different effective hydrophobicity around protein surfaces, influencing the behavior of hydration water significantly. [26] With molecular dynamics simulations, Godawat et al [36] found that water density near the surfaces of self-assembled monolayers (SAMs) with hydrophobic head groups ($-CF_3$, $-CH_3$) shows a poor distinction from that of SAMs with hydrophilic head groups ($-OH$, $-CONH_2$). However, differences arise when considering the fluctuations of water density near the two regions. Enhanced fluctuations, reflected by the broad probability distributions of water number density are observed around hydrophobic surfaces compared with the bulk solution and hydrophilic surfaces. [28, 34] Moreover, the enhanced density fluctuations around hydrophobic surfaces could further be characterized by more compressible hydration shells and increased cavity formation, [167, 168] indicating that the nature of hydration shells around hydrophobic surfaces are softer and more flickering than that of hydrophilic ones. Since the long-ranged ion-induced perturbations of aqueous protein interfaces involve the coupling of local hydration shells of the ions with distant hydration shells around protein surfaces, the nature of both would affect the extent of induced interfacial fluctuations. It would be interesting to compare the interface height fluctuations as Cl^-/I^- approaching the hydrophobic/hydrophilic protein regions. We note that the interface height fluctuations we are pursuing here are conceptually different from the density fluctuations, while both of them reflect the nature of hydration water around protein surfaces. Additionally, it has been reported that the ion-specific effects are dissimilar around hydrophobic and hydrophilic surfaces, with large I^- showing a stronger affinity than the smaller halide ions to the hydrophobic surfaces while the reverse trends of size-dependence of halide ions are realized at the hydrophilic surfaces. [64, 65, 66, 60] We would like to further connect this affinity (probabilities) differences of Cl^-/I^- around protein patches

with different hydrophobicity to their induced aqueous protein interfacial fluctuations correspondingly.

The particular protein we focus on in this study is hydrophobin-II (HFBII), which is a small protein with 71 amino acid residues expressed by filamentous fungi. The protein is known for its ability to form a hydrophobic coating on the surface of an object and it can self-assemble into a monolayer on hydrophobic/hydrophilic interfaces such as a water/air interface. [183] These functions are mainly determined by the amphiphilic structural characterization. Acharya et al. [27] mapped the effective hydrophobic regions and effective hydrophilic regions of HFBII by considering the density of small probe hydrophobic solutes around each region of the protein. Moreover, they selected three regions with different hydrophobicity based on this and further monitor the density fluctuations in their vicinity. The calculations showed that around most hydrophobic region, they observe the largest density fluctuations whereas the least density fluctuations were detected around most hydrophilic region. Considering of this, this protein is an ideal candidate to compare the characters between hydrophobic and hydrophilic interfaces.

The chapter is organized as follows. In Section 4.2 we discuss the simulation protocols and computational details of liquid-vapor interface and aqueous protein interfaces. Our results are presented in Section 4.3 and are organized into four topics. We start the discussion by investigating the PMFs and interfacial fluctuations as single Cl^-/I^- translocate across the aqueous liquid-vapor interface. We consider Cl^-/I^- density distributions around aqueous HFBII hydrophobic interface in 1.0 molal solutions in the second part. We further investigate the PMFs and interfacial fluctuations as single Cl^-/I^- approach the aqueous protein hydrophobic interface, demonstrating the similarity between liquid-vapor interface and hydrophobic protein interface in terms of ion specific induced perturbations of the interface. We finish this section by examining single Cl^-/I^- approaching another two regions with different hydrophobicity on the protein surface compared with the hydrophobic region we initially studied. We finish with our conclusions and general discussion in Section 4.4.

4.2 Method

4.2.1 Simulation Details

Molecular dynamics simulations performed in this study include: 1. umbrella sampling molecular dynamics simulations of translocation of a single Cl^-/I^- across the aqueous liquid-vapor interface; 2. molecular dynamics simulations of a single, fully rigid hydrophobin HFBII protein in 1.0 molal concentration of KCl/KI aqueous solutions; and 3. potential of mean force calculations using molecular dynamics simulation trajectories of single Cl^-/I^- approaching three different regions of the protein that are defined as hydrophobic, less hydrophobic and hydrophilic. Detailed simulation protocols are now discussed as follows.

4.2.1.1 Umbrella Sampling Potential of Mean Force Calculations: Ion Translocation Across Aqueous Liquid-vapor Interface

Molecular dynamics simulations were performed using the CHARMM package.[214, 186] Simulations of liquid-vapor interfaces were performed in the NVT ensemble. Temperature was maintained at $T = 300$ K using Nosé-Hoover thermostat.[215] The simulation cell was rectangular with dimensions $24 \text{ \AA} \times 24 \text{ \AA} \times 100 \text{ \AA}$, in which z is the direction normal to the liquid-vapor interface. A bulk slab consisting of 988 water molecules (represented by the nonpolarizable TIP3P model[190]) and a single anion (Cl^- , I^-) was positioned in the center of the simulation cell, resulting in two liquid-vapor interfaces. We note that Lennard-Jones parameters for ions that are suitable with TIP3P were taken from Cheatham et al[216]. The parameters are listed in Table 4.1. To verify if these parameters are suitable to TIP3P water model in CHARMM force field, we did some tests to compare the single water-ion binding distances and the single water-ion binding energies of the ions for the two cases. In these tests, one single ion and one water molecule were placed in a large enough simulation box and the non-bonded cutoff distances were also set as large as possible. Since our production simulations were performed with NAMD simulation package using CHARMM force field, ideally, we would like to also perform the test in the same way. Unfortunately, because

the default minimization algorithm in NAMD has a conflict with SHAKE algorithm which is used to constrain bond lengths involving hydrogen atoms, it is impossible to accurately minimize the structures with fixed water bond distances. Therefore, the energy minimization was performed with CHARMM package using CHARMM force field. With a stable structure, the binding energies can be recalculated using NAMD. The test results are shown in Table 4.1. For the cations (K^+ , Na^+ , Cs^+), test results match well with the original; while for the anions (Cl^- , I^-), we observe very small deviations for both single water-ion binding distances and single water-ion binding energies. We note that this is due to the fact that the TIP3P water model implemented in the CHARMM force field is slightly different from that of AMBER. In the CHARMM version of TIP3P, Lennard-Jones parameters on hydrogen is nonzero, whereas in the original version, Lennard-Jones interaction contributions from water hydrogen atoms not included. We verify this with a further test where we applied non-bonded fixed (NBFIX) strategy in CHARMM to ensure that there are no Lennard-Jones interactions involving hydrogen atoms from water; we obtained an exact matching result of binding distances and binding energies compared with original report. Despite this minor issue, we still transfer the ion parameters in TIP3P AMBER to the CHARMM force field in this study. We consider this as a valid combination because this empirical model could reproduce the most important characteristics that we would like to address between the two distinct anions: for Cl^- , it is small, fully hydrated with rigid hydration shell; while for I^- , it is large, soft, partial hydrated with malleable hydration shell. These characters can be proved by the ion-water RDF in Figure 4.1. We note that currently no ion parameters could be considered as absolutely correct in conjunction with proteins during the simulation, because essentially no ion are parameterized based on the interactions with proteins. Therefore, in discussion about ions' effects related to the proteins, as long as the model could reproduce currently accepted experiment observables, it could be considered as validation.

A rigid water geometry is enforced using SHAKE[192] constraints. Conditionally convergent long-range electrostatic interactions were treated using Particle Mesh Ewald

Non-Bonded Parameters	σ (Å)	ε (kcal/mol)
Cl ⁻	5.026	-0.0355910
I ⁻	5.720	-0.0536816
K ⁺	3.410	-0.1936829
Single Water-Ion Binding Distances	Cheatham (Å)	this research (Å)
K ⁺	2.66	2.66
Na ⁺	2.29	2.29
Cs ⁺	3.00	2.98
Cl ⁻	3.09	3.13
I ⁻	3.48	3.50
Single Water-Ion Binding Energies	Cheatham (kcal/mol)	this research (kcal/mol)
K ⁺	-18.51	-18.52
Na ⁺	-24.29	-24.30
Cs ⁺	-15.08	-15.10
Cl ⁻	-14.26	-14.15
I ⁻	-11.34	-11.37

Table 4.1: LJ parameters for ions applied in this work and verification. Note: for the case of single water-anion binding case, there are two geometries for the binding structure, one is C_S , another is C_{2v} ; the binding distances and binding energies for Cl⁻ and I⁻ shown here are from C_S geometry.

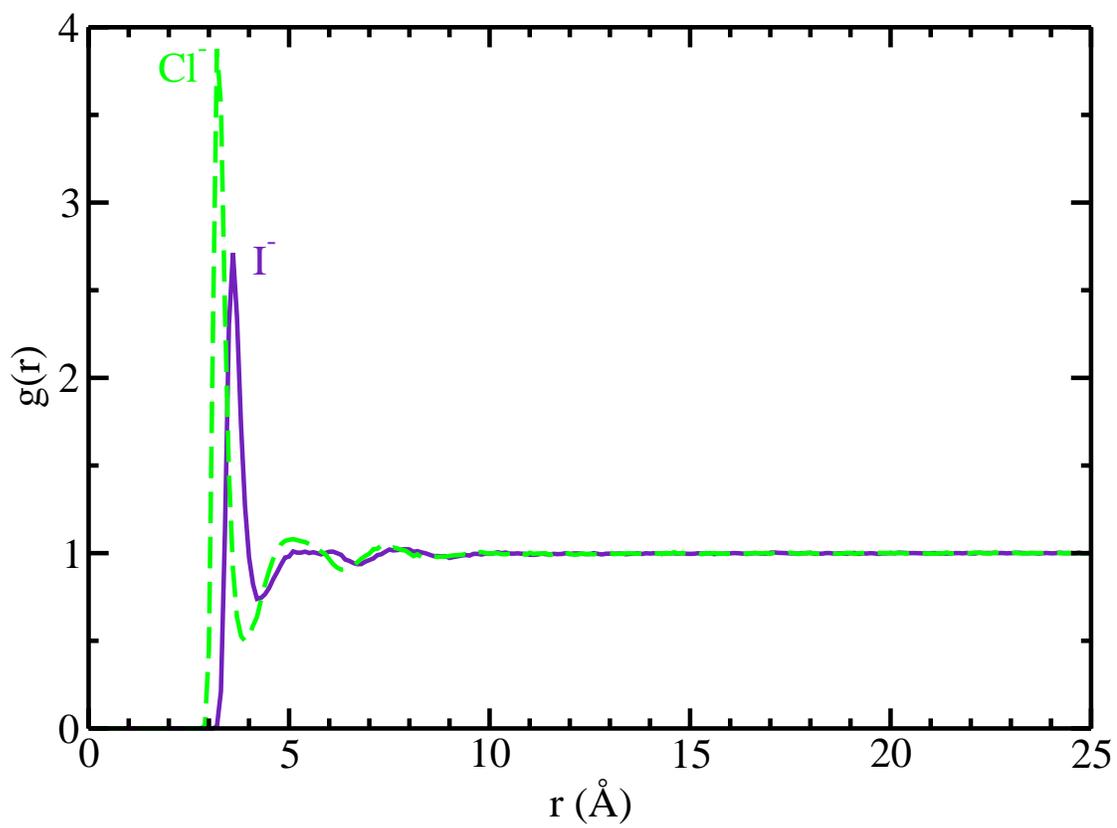


Figure 4.1: Ion-water radial distribution function (RDF) for anions in TIP3P.

(PME)[191] approach with a 30 x 30 x 128 point grid, 6th order interpolation, and $\kappa = 0.33$. Dynamics were propagated using a Verlet leap-frog integrator with a 1.0 fs timestep. Computational experiments measuring the reversible work for transferring single ions/molecules from bulk aqueous environment to the aqueous solution liquid-vapor interface have enjoyed a long history as a means to explore the origins of surface stability.[45, 217] In order to determine the PMF, a reaction coordinate defining this pseudo-chemical reaction must be defined. Our reaction coordinate for PMF is the Cartesian z -component of the separation between the water slab center of mass and ion position. Along the z axis, to enhance sampling of the distribution of configurations where the reaction coordinate holds a particular value, the reaction coordinate was restrained within a certain narrow range (instead of its entire span). In this case, we constructed 26 continuous “windows” with width 1.0 Å. In each window, single anion was restrained to z -positions from 10 Å to 35 Å relative to the water slab center of mass using a harmonic potential $U_{\text{restraint}}(z; z_{\text{relative,ref}}) = \frac{1}{2}k_{\text{restraint}}(z - z_{\text{relative,ref}})^2$ with the force constant of 4 (kcal/mol)/Å²; this encompassed a range approximately 15 Å below the GDS to approximately 10 Å above it at 300 K. Though one could probe separations further into the bulk (towards the center of the system) this distance is sufficient to probe the differences of interest in this study. Total sampling time for each window was 30 ns; properties were calculated from all but the initial 1.0 ns, which was treated as equilibration.

4.2.1.2 Protein in KCl/KI Aqueous Solution

Simulations of a single hydrophobin in 1.0 molal concentration of KCl/KI aqueous solution were performed with NAMD, version 2.9b3,[184, 185] using the CHARMM 22 all-atom force field with CMAP backbone torsion correction term.[187] Identical parameters for water (TIP3P) and ions (Cl^- , I^- and K^+) were applied as the ones from liquid-vapor interface simulation. Isothermal - isobaric ensemble (NPT) simulations were performed using a cubic cell with a box size 60 Å × 60 Å × 60 Å. NPT ensemble was used to eliminate the liquid-vapor interfaces, so only the protein-water interfaces

were considered in the system. The initial structure of the protein was constructed using CHARMM-GUI. [189] The protein structure was based on the ultra-high resolution structure at 0.75 Å of hydrophobin HFBII, with PDB code 2B97. [218] The original crystal structure was actually the dimerization complex of the protein. Only one monomer, composed of 70 residues, was modeled in this study. The protein was placed in the center of the box, with center of mass located at ($x = 0$ Å, $y = 0$ Å, $z = 0$ Å), the rest of the box was filled with 6481 water molecules, 116 K^+ and 116 Cl^-/I^- , resulting in a molal concentration of 1.0 m. Temperature was maintained by Langevin bath at 300K, and the pressure was kept constant by Langevin pressure control at 1 atm. A switching distance of 10 Å, non-bonded real-space cutoff of 12 Å and pairlist generation distance of 14 Å were used for the van der Waals interactions, and the particle mesh Ewald (PME) method was employed for the calculation of conditionally-convergent electrostatic interactions. The grid size of PME in x dimension is 60, in y dimension is 60, and in z dimension is 60 (as close to a 1Å grid point separation as possible). The SHAKE algorithm was used to constrain bond lengths involving hydrogen atoms and an integration time step of 1.0 fs was used. The protein was fixed during the simulation with other components could move randomly. We provide the NAMD input file for our simulations in Table 4.2. A total of six different replicates were used and the first 2.0 ns of each replicate was considered as equilibration. At least 10 ns of production run for each replicate was used to compute properties.

4.2.1.3 Aqueous Protein Interfaces

In order to illustrate the molecular detail and free energetics of Cl^-/I^- approaching the aqueous protein interfaces with different hydrophobicity, we further simulated systems with 6481 TIP3P water molecules and a single Cl^-/I^- , transferring from bulk to the protein interfaces. A representative snapshot of the simulation system can be found in Figure 4.2A. HFBII protein was arranged in a way that its largest hydrophobic patch, consisting of V18, L19, L21, I22, V24, V54, A61, L62 and L63 (shown in Figure 4.2B), was nearly perpendicular to the z direction (further quantitative information is

exclude	scaled1-4	langevinPistonPeriod	50
1-4scaling	1	langevinPistonDecay	25
COMmotion	no	langevinPistonTemp	300
zeroMomentum	no	useFlexibleCell	no
dielectric	1.0	useGroupPressure	yes
switching	on	cellBasisVector1	60.00 0.00 0.00
switchdist	10	cellBasisVector2	0.00 60.00 0.00
cutoff	12	cellBasisVector3	0.00 0.00 60.00
pairlistdist	14	cellOrigin	0.00 0.00 0.00
timestep	1.0	wrapAll	on
stepspercycle	20	PME	yes
nonbondedFreq	1	PMEGridSizeX	60
fullElectFrequency	2	PMEGridSizeY	60
rigidBonds	all	PMEGridSizeZ	60
langevin	on	constraints	on
langevinDamping	5	selectConstraints	on
langevinTemp	300	selectconstrX	on
langevinHydrogen	off	selectconstrY	on
langevinPiston	on	colvars	on
langevinPistonTarget	1.01325		

Table 4.2: NAMD input parameters for the simulations.

in Table 4.3) and the whole protein was fixed during the simulation with center of mass located at ($x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$, $z = 0 \text{ \AA}$). A single Cl^-/I^- was added in the solution with one counter ion, K^+ , to neutralize the negative charge of the monovalent anion. The K^+ was fixed at position ($x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$, $z = -15 \text{ \AA}$). Similar to the liquid-vapor interface situation, for calculation of PMF, we consider the Cartesian z component of the separation between the center of mass of protein and center of mass of the single Cl^-/I^- as the reaction coordinate for the present umbrella sampling molecular dynamics simulations. Single Cl^-/I^- was aligned along the z direction, approaching the specific spot on the patch with position $x = 0 \text{ \AA}$ and $y = 0 \text{ \AA}$ by freezing the orthogonal degrees of freedom along x axis and y axis via the use of strong restraining potentials. We center on one specific region of the patch, acknowledging that the heterogeneity of the protein surface necessitates some care in interpreting the results, which we will address further below. For a meaningful discussion and interpretation of ion-induced fluctuation (interface fluctuation in addition to the level present in pure water) as the ion approaches the hydrophobic interface, one reference location with fixed position has to be defined. Using NAMD’s “selectConstraints” infrastructure, x component of the ion was restrained at $x = 0 \text{ \AA}$ and y component was restrained at $y = 0 \text{ \AA}$ with a force constant of $1000 \text{ (kcal/mol)/\AA}^2$ respectively. Along the z axis, we constructed 46 continuous umbrella sampling “windows” with width 0.2 \AA along the positive z -direction ranging from area around protein-solvent interface to bulk water region. The spans of the windows going from interfacial region to bulk region (in \AA) were: [16.0:16.2], [16.2:16.4], [16.4:16.6] [24.4:24.6], [24.6:24.8], [24.8:25.0]. This range of ion position (from 16 \AA to 25 \AA) is sufficient to probe the differences of single ion around interface and that in bulk water region, while minimizing the number of windows that is required to construct. In each window, a harmonic restraint potential with force constant of $10 \text{ (kcal/mol)/\AA}^2$ was applied on Cartesian z component of the ion. Other simulation conditions remain the same as that of the 1 m concentration of KCl/KI aqueous solution. The first 2ns was allowed for equilibration before a total of 20ns production data was generated for each window.

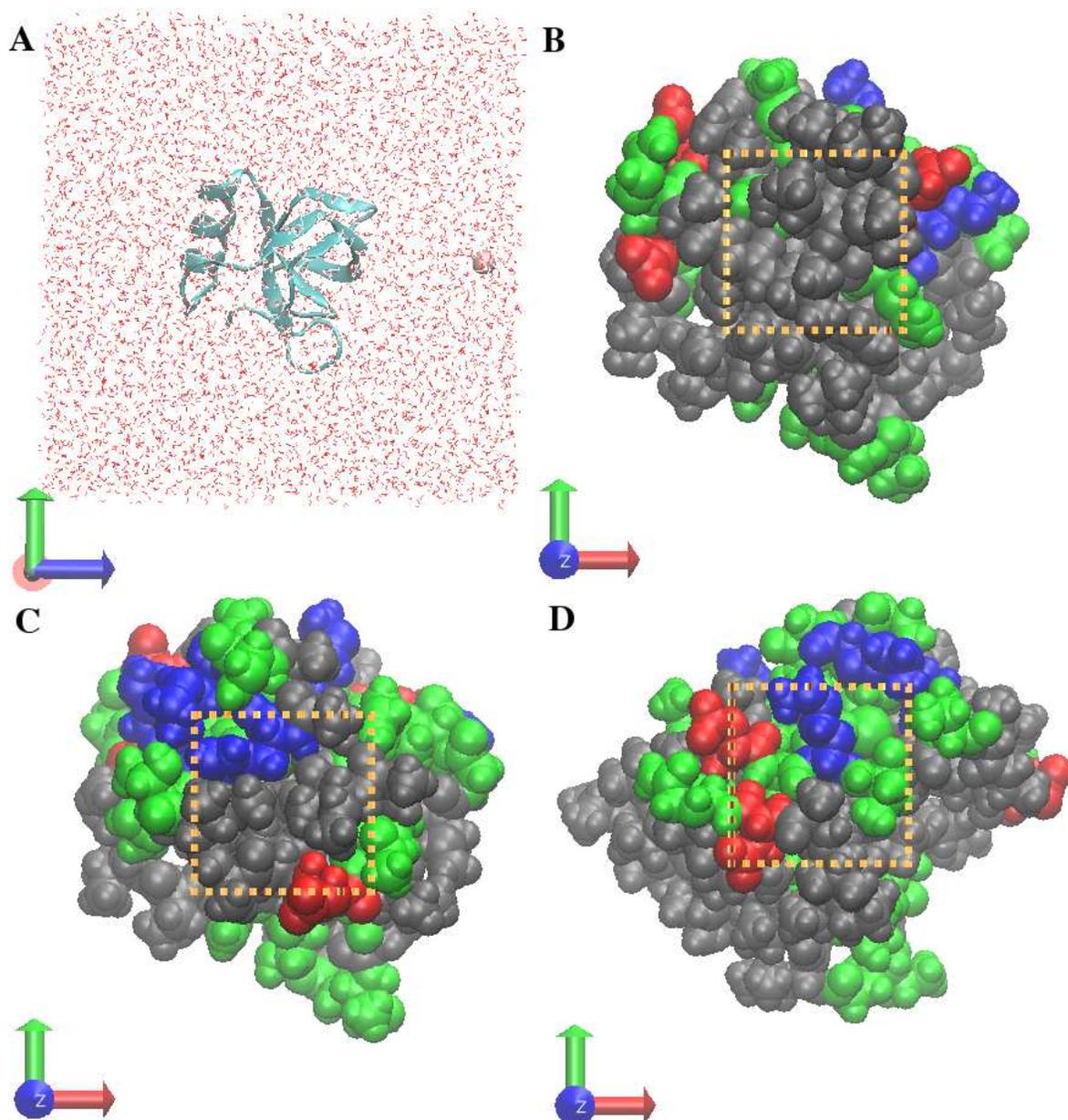


Figure 4.2: (A) Representative snapshots of the system used in the study (B) representation of the hydrophobic interface defined in this study. Including residues L7, V18, L19, L21, I22, V24, V54, A61, L62, L63. Dash orange line roughly select the region of interest (C) representation of the less hydrophobic interface defined in this study. Including residues I31, A32, D34, I38, A41, H42, S45. (D) representation of the hydrophilic interface defined in this study. Including residues D25, C26, K27, T28, A58, D59, Q60.

<i>Residue Name</i>	<i>Atom Type</i>	<i>Angle(°)</i>
L21	<i>N</i>	113
L21	<i>CA</i>	110
L21	<i>CB</i>	91
L21	<i>CG</i>	86
L21	<i>CD1</i>	72
L21	<i>CD2</i>	87
L22	<i>CD</i>	75
L63	<i>CD1</i>	76
L63	<i>CD2</i>	90

Table 4.3: Angle between positive z vector and the line connecting central position of the patch (0,0,12) with each of the heavy atom position on the patch.

For comparison, we performed another set of simulations to compute the PMF of the anion approaching the hydrophobic patch using average force integration; in these simulations, both anion and protein are held fixed so as to realize a series of center of mass separation distances; the potential of mean force is obtained by integration of the average force along the reaction coordinate obtained from simulation trajectory analysis. Furthermore, to attempt to consider effects of protein motion on ion-induced interface fluctuations, we performed simulations with protein under restraint conditions instead of totally fixed. HFBII was placed in the center of box with exactly the same starting structure as in the fixed(rigid) protein case. During the simulation, the protein was strongly restrained to remain in a single orientation and its center of mass at a specific position, chosen as ($x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$, $z = 0 \text{ \AA}$) via the use of strong restraining potentials. Using NAMD’s collective variable infrastructure, HFBII’s center of mass was restrained at ($x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$, $z = 0 \text{ \AA}$) using a force constant of $5000 \text{ (kcal/mol)/\AA}^2$, and its orientation was restrained about the crystal based orientation using a harmonic restraint potential with force constant of $5000 \text{ (kcal/mol)/\AA}^2$. Single Cl^-/I^- was fixed along positive z axis, starting from $z = 16 \text{ \AA}$ to $z = 25 \text{ \AA}$, a total of ten continuous windows with width 1.0 \AA . We note that PMF calculations will not be concerned in the restrained protein case since it requires extensive simulation time for a well-converged PMF with flexible protein. Instead, we

are only interested in the comparison of ion-induced interfacial fluctuations of total fixed protein and restrained moving protein as Cl^-/I^- locates at particular separations along the reaction coordinate. Besides the simulations of single Cl^-/I^- approaching to the most hydrophobic region of the protein, we considered two other scenarios in which single anions approach protein regions with different hydrophobicity. Depending on the nature of residues that are exposed, we defined one patch as less hydrophobic interface and the other as a hydrophilic interface in order to distinguish them from the hydrophobic interface we previously described. For these additional two cases, the simulation conditions remained identical to those in the hydrophobic patch calculations, except that the protein was oriented in a different way in the simulation cell. For the simulations in which the anions approach the less hydrophobic region, the interface is composed of residues I31, A32, D34, I38, A41, H42 and S45, arranged perpendicular to the z direction (shown in Figure 4.2C). Forty-nine (49) continuous windows with width 0.2 Å along the positive z-direction, starting with [15.4:15.6], [15.6:15.8], [15.8:16.0] to [24.4:24.6], [24.6:24.8], [24.8:25.0] are constructed. For the hydrophilic interface case, the interface we centered on consists of residues D25, C26, K27, T28, A58, D59, Q60 (shown in Figure 4.2D). Similarly, this interface was oriented in a way that is perpendicular to the z direction. The window setup ranged from [14.0:14.2], [14.2:14.4], [14.4:14.6] to [24.4:24.6], [24.6:24.8], [24.8:25.0], a total of 56 windows.

Finally, we address the protocol for simulations where the PMF is computed by an average force integration method. The PMF of single Cl^-/I^- approaching protein interface can be calculated by integration of the average forces acting on the anion as shown in Equation 4.1.

$$W(\xi_0) = - \int \langle F(\xi_0) \rangle d\xi_0 \quad (4.1)$$

where ξ_0 is the reaction coordinate taken as the separation distance between the Cl^-/I^- and the center of mass of the protein; $\langle F(\xi_0) \rangle$ denotes the average forces acting on the anion at each separation along the reaction coordinate. Uncertainties in PMF are

determined as: [219]

$$\text{var}[G(\xi_N)] \approx \sum_{i=1}^N \text{var}[K \Delta \xi \bar{z}_i] \quad (4.2)$$

where $\text{var}[G(\xi_N)]$ is the variance, \bar{z}_i is the mean position of z in the i_{th} window, which can be obtained from block averages.[220] The standard deviation $\sigma[G(\xi_N)]$ is then the square root of $\text{var}[G(\xi_N)]$.

4.3 Result and Discussion

4.3.1 Liquid-Vapor Interface

We start to look at the free energetics of single Cl^-/I^- across the liquid-vapor interface. Results of PMF for Cl^- and I^- are shown in Figure 4.3A. For clarity, we added a vertical offset of 0.1 kcal/mol for the Cl^- case. To better compare the interface stability between the two ions, in the large graph of Panel A, we emphasized the PMFs around the interfacial region while the whole PMFs along the reaction coordinate can be found in the small inset. The PMF is defined to be zero in the bulk (which is determined by window $z = 10 \text{ \AA}$). I^- has a slight PMF minimum ($\approx 0.05 \text{ kcal/mol}$) prior to the GDS (Gibbs dividing surface is around $z = \pm 25.5 \text{ \AA}$ in this case). Due to the uncertainty reported, whether I^- shows surface stability is ambiguous. However, we notice that there is a barrier (around $z = 19 \text{ \AA}$) prior to the PMF minimum, which is also observed in other studies;[45, 56] as a result, although being less explicit than the interfacial stability reported in experiments and other force fields,[208, 51, 50, 55, 57, 56] qualitatively we consider that I^- exhibits a surface-stable state in the current simulations. In contrast, Cl^- is repelled from the L-V interface in the current and other force fields.[56] In the Drude polarizable force field, Cl^- shows similar behavior as I^- , having a marginal stabilizing/negative free energy minimum state followed by a barrier (from bulk to vapor phase). Unlike the nonpolarizable force fields, the Drude force fields encounter the issues of overpolarization,[221] which leads to differences in describing the presence of Cl^- at the interface using Drude and non-polarizable and other polarizable force fields[222]. Consequently, we do not consider Cl^- to be

interface stable, and I^- as having liquid-vapor interface stability with the current force field, consistent with previous studies. In this work, we stress that we are not focusing on the exact values of free energetics of single Cl^-/I^- adsorption at the liquid-vapor interface, but rather we want to emphasize the interfacial stability difference between Cl^- and I^- and related physical and structural properties. More importantly, we would like to connect these ion-specific behaviors at aqueous liquid-vapor interfaces to those of more general aqueous protein hydrophobic interfaces. The nonpolarizable water model and nonpolarizable protein parameters combination would clarify these issues with the benefit of saving computational resources compared with the polarizable force field. In light of this, we argue that the current force field we are applying is sufficiently robust and appropriate.

Our previous studies have demonstrated a connection between interfacial stability of Cl^-/I^- around liquid-vapor interface and the magnitude of their induced fluctuations of the interface in SPC/E, TIP4P-FQ, SWM4-NDP and TIP4P-QDP water models. [56, 57] It is found that the species demonstrating an interfacial stability appear to enhance liquid-vapor interfacial fluctuations significantly, while those that show no interfacial stability induce no further fluctuation (or may even suppress levels of fluctuations). Here we explore the differences in interfacial fluctuations for the two anions discussed in the current simulations. The fluctuations were computed with the protocol as we state in Chapter Two. From our previous work, [56] the geometry of the fluctuation surface $\langle \delta h^2(x, y) \rangle$ is radically symmetric, with the largest value at the center $x = 0, y = 0$ (right towards the ion). For convenience, we use $\langle \delta h^2(x = 0, y = 0) \rangle$ to compare the magnitude of interface fluctuations when Cl^-/I^- are restrained at different z-positions, with the result shown in Figure 4.3B. The fluctuation profile is normalized by the fluctuation value in pure water (i.e., the system in the absence of the ion, which has a value of 0.77 \AA^2). Normalization in this manner somewhat accounts for neglecting effects of larger wavelength undulations of the interface and affords a way to compare systems of different lateral dimensions if so needed. In this convention of normalized surface fluctuation ($\langle \delta h_{\text{L}}^2 \rangle$) we extract the ion-induced contribution from each species

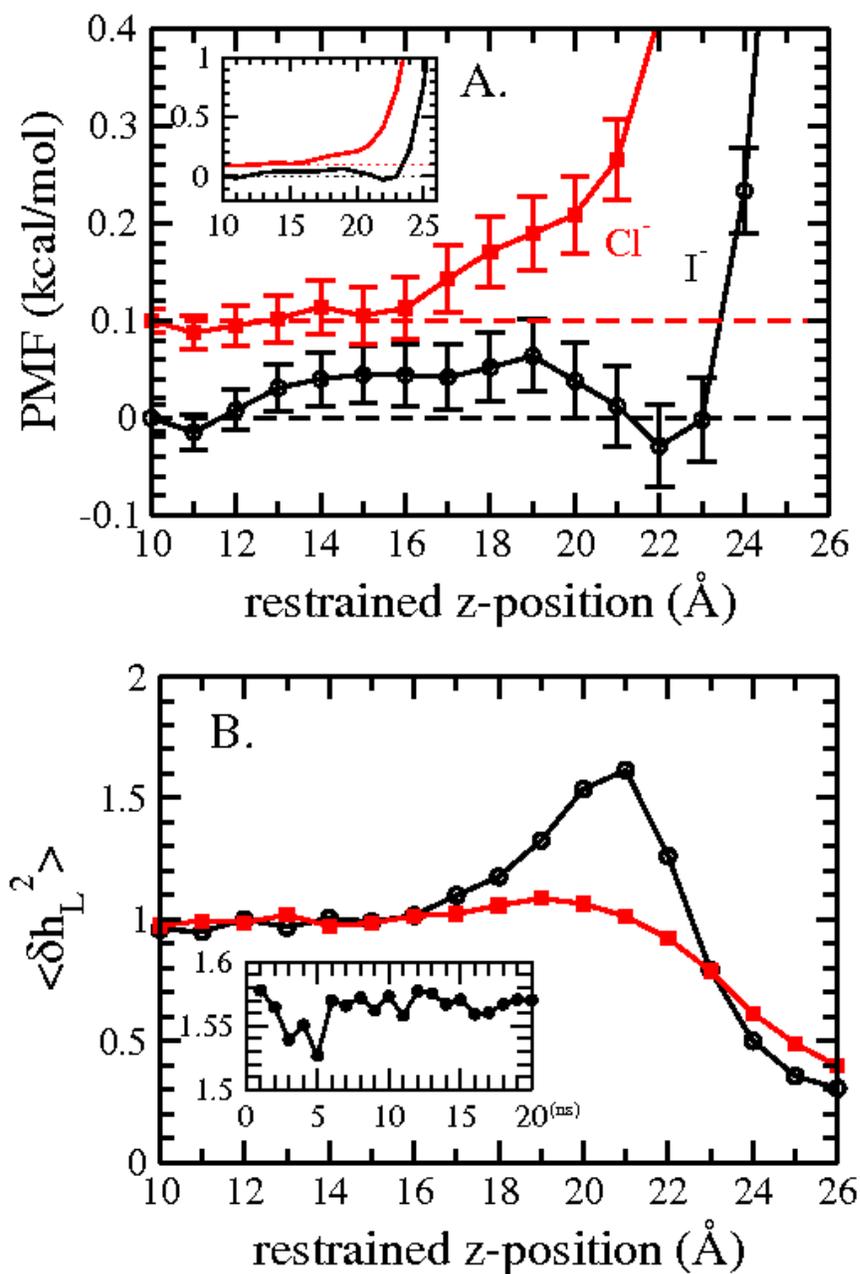


Figure 4.3: (A) PMF of single Cl⁻/I⁻ approaching the liquid-vapor interface in TIP3P water (B) Normalized liquid-vapor interface fluctuation at (x = 0, y = 0) as a function of anion restrained position for Cl⁻ and I⁻.

at different z -position. When $\langle \delta h_L^2 \rangle$ equals 1, the effect of ion is zero; when $\langle \delta h_L^2 \rangle > 1$, the surface height fluctuation is enhanced relative to pure water with the presence of ion; when $\langle \delta h_L^2 \rangle < 1$, the surface height fluctuation is suppressed. No obvious enhancement of surface fluctuations is associated with Cl^- ; on the other hand, I^- induces larger fluctuation, with the maximum normalized fluctuation value around 1.5 at the location of $z = 21 \text{ \AA}$, which is before the position of the free energy minimum. Also presented in the inset is the time profile of $\langle \delta h_L^2(x = 0, y = 0) \rangle$ for I^- at the window $z = 21 \text{ \AA}$ (which possesses the largest surface fluctuation) to show the convergence of the fluctuation. Previously, by studying a wide variety of force fields (polarizable and nonpolarizable), our results [56] suggest a threshold value of the maximum normalized interfacial fluctuations about 1.5 dividing those ions that are interfacially stable and those that are not. The largest normalized fluctuation and ΔG for I^- are 1.55 (unitless) and -0.03 kcal/mol , just barely placing it on the critical/transitional position in Figure 4 of Reference. [56] For Cl^- , we found the maximum fluctuation is 1.1 and the corresponding $\Delta G = 0.52 \text{ kcal/mol}$, which falls in the quadrant for non-surface stable species. It indicates that in terms of the surface stability, the current force fields for anionic behavior is consistent with other force fields. The differential behavior of the two ions at the pure aqueous liquid-vapor interface, consistent with previous studies, thus provides the control needed to interpret the simulations in a protein context.

We note that the differences in induced interfacial fluctuations by Cl^- and I^- may attribute to these two types of ions presenting distinct hydration shell environments. The first solvation shell of Cl^- is more rigid and less malleable than that of I^- . The nature of the solvent structure around I^- determines that it is more amenable to inducing fluctuations of the interface as a consequence of a greater disruption of the solvent structure on approach to the interface. This solvation shell property difference between Cl^- and I^- in polarizable water has been discussed previously [223]. To corroborate that these characteristics are similar when using the current nonpolarizable force field, we show the radial distribution functions (RDFs) based on water oxygen - single Cl^- and water oxygen - single I^- in Figure 4.1. Cl^- shows a predominant

first solvation peak, and an oscillatory probability function, signifying a substantially structured hydration environment; in contrast, the I^- RDF exhibits a modest peak, and markedly less oscillations, which is consistent with the results previously we have obtained for RDFs in different water models (SPC/E, TIP4P-FQ, SWM4-NDP and TIP4P-QDP). Overall, with the current force field, we observed ion-specific interfacial behaviors between Cl^- and I^- and also their distinct ability to induce long-ranged perturbations of the aqueous liquid-vapor interface as we have previously discovered in other water models. A further step in this work is that we attempt to extend this investigation from the ideally hydrophobic aqueous liquid-vapor interface to a more somewhat more realistic, and certainly more complex, aqueous protein hydrophobic interface.

4.3.2 Ion Distributions Around Protein in 1 Molal Aqueous Environment

Here we consider the protein in 1.0 m KCl/KI aqueous solutions, seeking a general overview of the relative stability of Cl^- and I^- around the hydrophobic interface of the protein; superficially, we compare the relative probability of finding an anion of each type in the vicinity of the protein interface. Figure 4.4 shows spatial distribution of number density of Cl^-/I^- around the hydrophobic interface of HFBII in 1.0 m KCl/KI aqueous solution. The composition of the hydrophobic patch has been discussed in the Method Section and roughly the position of the patch is within the range of ($-10 \text{ \AA} < x < 10 \text{ \AA}$, $-10 \text{ \AA} < y < 10 \text{ \AA}$, $6 \text{ \AA} < z < 13 \text{ \AA}$), so we consider anion density distribution only around this region. The x-axis represents the lateral distance $r = \sqrt{x^2 + y^2}$ (the sign of r depends on that of x component and y component; if they are the same, $r > 0$; if they are different, $r < 0$), and the y-axis is the z distance from the center of mass of protein located at (0,0,0). Comparison of Panels A (Cl^- density distribution) and B (I^- density distribution) indicates that I^- has a higher propensity for the hydrophobic protein interface. For a more quantitative comparison, in Figure 4.5 we show the number of bins (i.e., effective volume) with Cl^-/I^- densities above certain threshold values around the hydrophobic patch. The bins were constructed in three dimensional

space with size 1 Å x 1 Å x 1 Å, and the ion densities in each bin were computed as normalized values by dividing the numbers of Cl⁻/I⁻ in the bin in the presence of the protein with the number in the absence of protein. Therefore, a normalized density value that is larger than one implies the existence of protein enhances the anion density in the particular site of interest. We consider scenarios with normalized anion densities greater than 3, 4, 5, 6, 7 and 8 for Cl⁻ and I⁻, shown in different panels in the figure. We find that, consistently, at different radii close to the hydrophobic patch and above various thresholds, there is greater enhancement of I⁻. Our observation agrees with those of Lung et al. [212] in their simulation study on lysozyme in a mixed aqueous solution of KCl and KI. They observed a specific ion effect around the protein showing that Cl⁻ has virtually no preference for nonpolar regions, but positively charged residues, whereas I⁻ accumulates in the vicinity of hydrophobic groups. They explain the behavior of Cl⁻ as a direct ion pairing interaction, involving small, fully hydrated Cl⁻ with cationic groups, and I⁻'s behavior as solvent-assisted attraction of large, soft, and partially-hydrated I⁻ to nonpolar protein surface patch. This view of the differences in ion behavior suggests an underlying ligand-substitution theme as well. Chloride must substitute a rigid, strongly held solvation shell with another ligand (this terminology is intentionally used broadly and non-specifically in this situation); this ligand is a polar or charged entity. The iodide, due to its low charge-density arising from the classical representation of this entity, can accommodate loss of its rather loose, less well-defined solvation shell. For a further atomic level understanding of this solvent-assisted mechanism and a quantitative comparison of the stability of Cl⁻ and I⁻ around particular region of HFBII, in the next subsection, we consider the potential of mean force to as a single Cl⁻/I⁻ approaches, from the bulk, a specific point on the hydrophobic interface of HFBII.

4.3.3 Potential of Mean Force

The umbrella sampling molecular dynamics PMF for both anions approaching the hydrophobic interface are shown in Figure 4.6A; large values of the x-axis represent

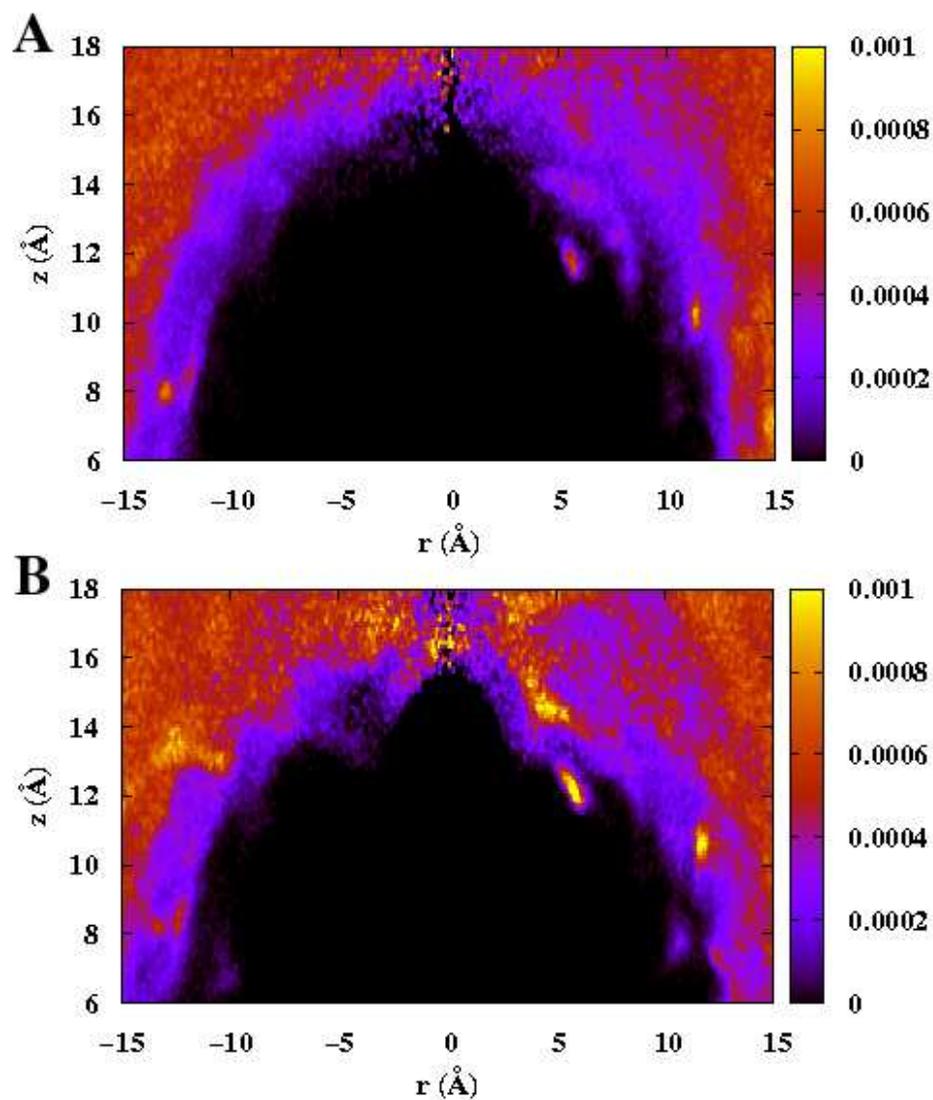


Figure 4.4: Number density distribution of Cl^-/I^- around the hydrophobic interface of HFBII in 1.0 m KCl/KI aqueous solution. (A) Cl^- density distribution (B) I^- density distribution. X axis represents the lateral distance $r = \sqrt{x^2 + y^2}$. We define $r > 0$ means the signs of x component and y component are the same; while $r < 0$ means the signs of x component and y component are different.

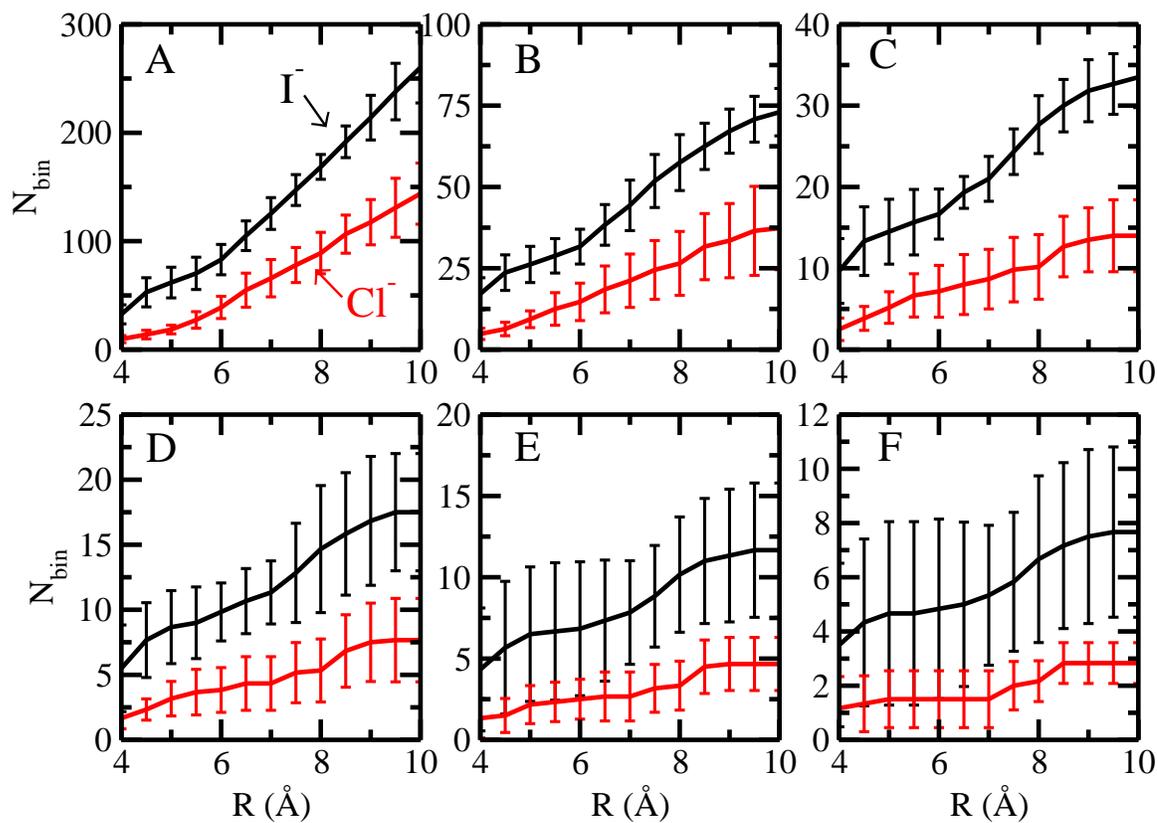


Figure 4.5: Number of bins that display Cl^-/I^- densities above certain threshold values around the hydrophobic patch of HFBII in 1.0 m KCl/KI aqueous solution. (A) above threshold value 3 (B) above threshold value 4 (C) above threshold value 5 (D) above threshold value 6 (E) above threshold value 7 (F) above threshold value 8.

large separation of anion and protein center of mass, and the PMF's are zeroed at large separation. To assess the convergence of the potential of mean force, we show the time evolution of the minimum of the PMF in Figure 4.7. Also, the PMF from this restrained anion protocol is shown to be consistent with the fixed anion approach (average force integration), a comparison of which is shown in Figure 4.8. For Cl^- , there is a small barrier around $z = 19.5 \text{ \AA}$, followed by a shallow minimum around $z = 18.5 \text{ \AA}$; a similar trend is seen for I^- , with a small barrier around $z = 20.5 \text{ \AA}$ and a minimum afterward. For Cl^- , the PMF minimum is $-0.06 \pm 0.05 \text{ kcal/mol}$; for I^- , it is $-0.08 \pm 0.04 \text{ kcal/mol}$. In light of the uncertainty estimates, both Cl^- and I^- exhibit little stabilization at the hydrophobic protein interface. However, as the single Cl^-/I^- draws near the interface, significant differences arise. The Cl^- PMF starts to increase monotonically; the I^- PMF shows a slightly more complex trend. Unlike the situation of Cl^- , the PMF profile of I^- shows a second minimum, which is a little higher ($0.20 \pm 0.04 \text{ kcal/mol}$) than the first one. At this second minimum position, the free energetic difference between Cl^- and I^- is about $0.78 \pm 0.09 \text{ kcal/mol}$, even with the consideration of the uncertainty. This implies that close to the hydrophobic protein interface, I^- tends to be more interface stable than Cl^- , although compared with bulk, neither of them displays the stabilization effect around the interface within the context of the specific force field we have chosen to use in this study. We note that the dramatic increase of PMF for both Cl^- and I^- starting around $z = 18.5 \text{ \AA}$ may be related to the change of number of coordinate water in the first hydration shell around the ion, as it has been shown in Figure 4.6B. When the ions are close enough to the interface, there will be the decrease of hydration water. Consequently, the favorable interaction between single anion and water will be lost, entailing the increase of free energy. Since the two anions display distinct free energy profiles nearing the interface, we next consider the induced fluctuations associated with the approach of these ions in the spirit of earlier studies. [55, 56, 57]

The aqueous protein interface was constructed based on the protocol mentioned in Chapter Two. Figure 4.9 displays the mean protein-solvent interface along with

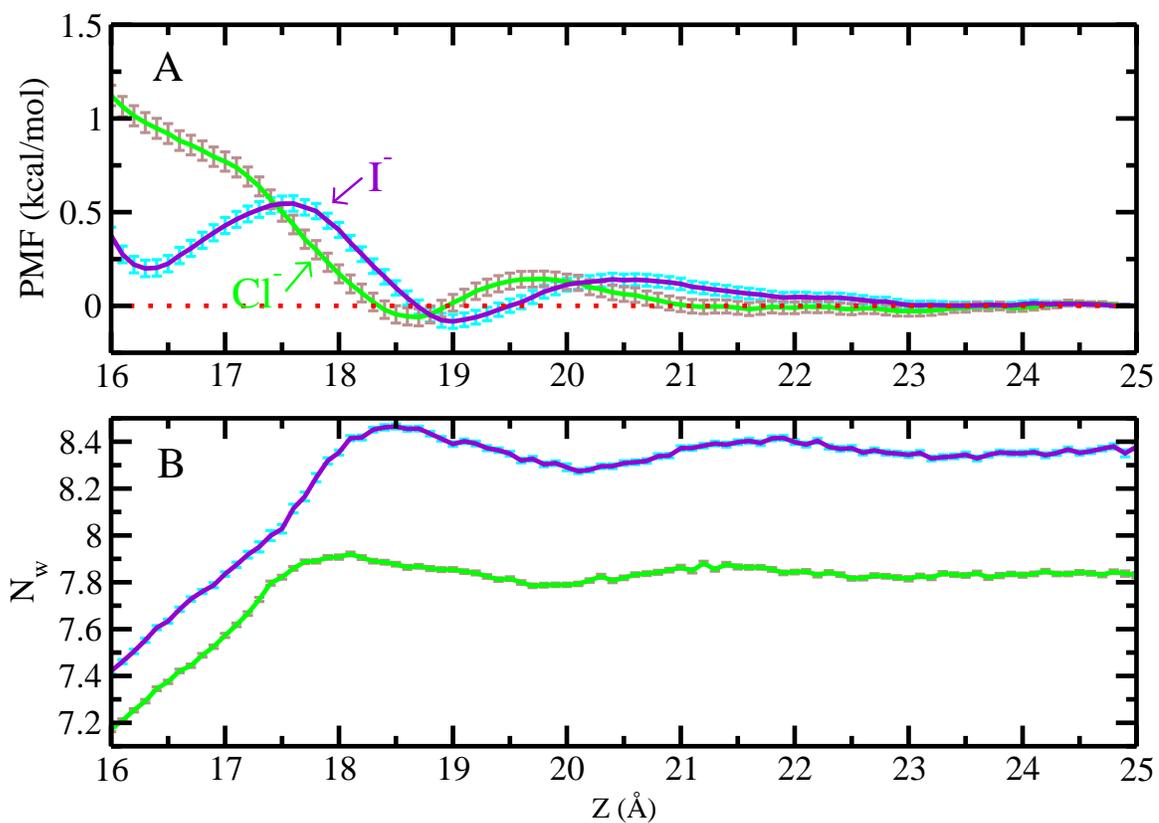


Figure 4.6: (A) PMF for single Cl^-/I^- approaching the hydrophobic protein-solvent interface (B) Coordinate water numbers around single Cl^-/I^- as a function of the reaction coordinate.

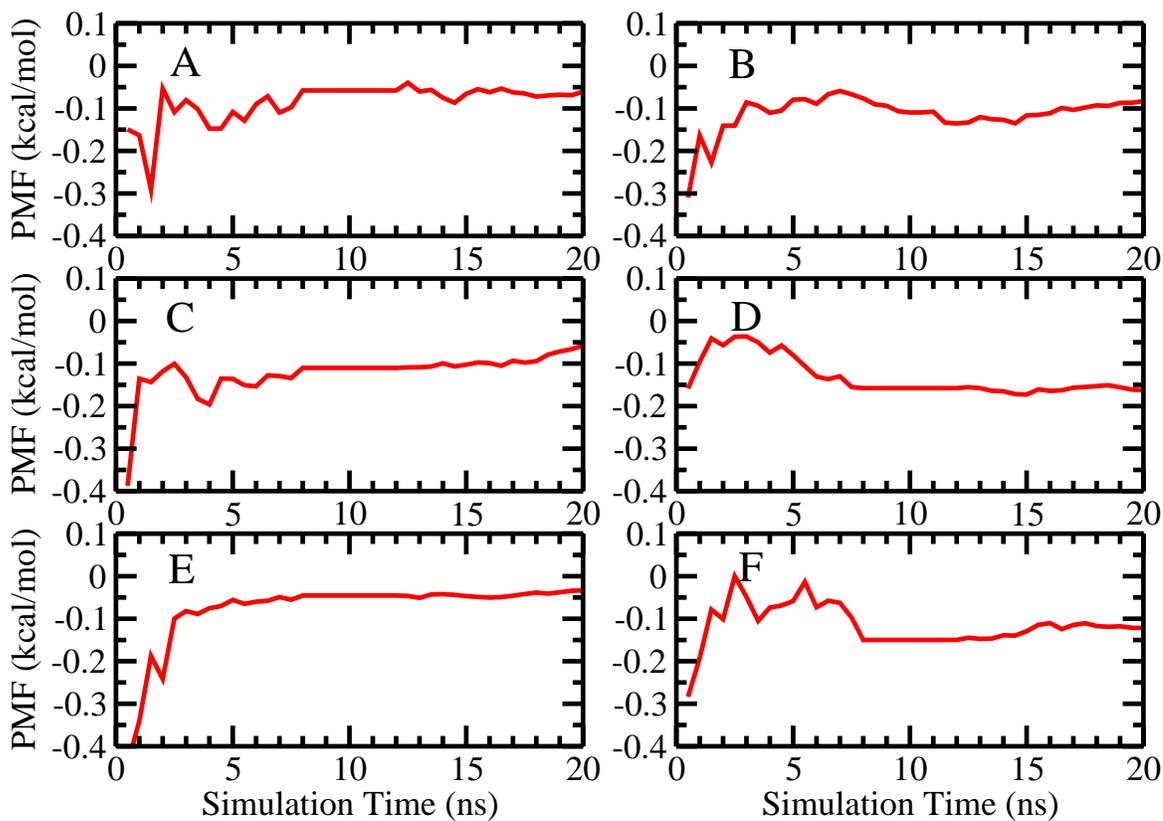


Figure 4.7: (A) Evolution of PMF for Cl⁻ approaching hydrophobic interface (B) Evolution of PMF for I⁻ approaching hydrophobic interface (C) Evolution of PMF for Cl⁻ approaching less hydrophobic interface (D) Evolution of PMF for I⁻ approaching less hydrophobic interface (E) Evolution of PMF for Cl⁻ approaching hydrophilic interface (F) Evolution of PMF for I⁻ approaching hydrophilic interface.

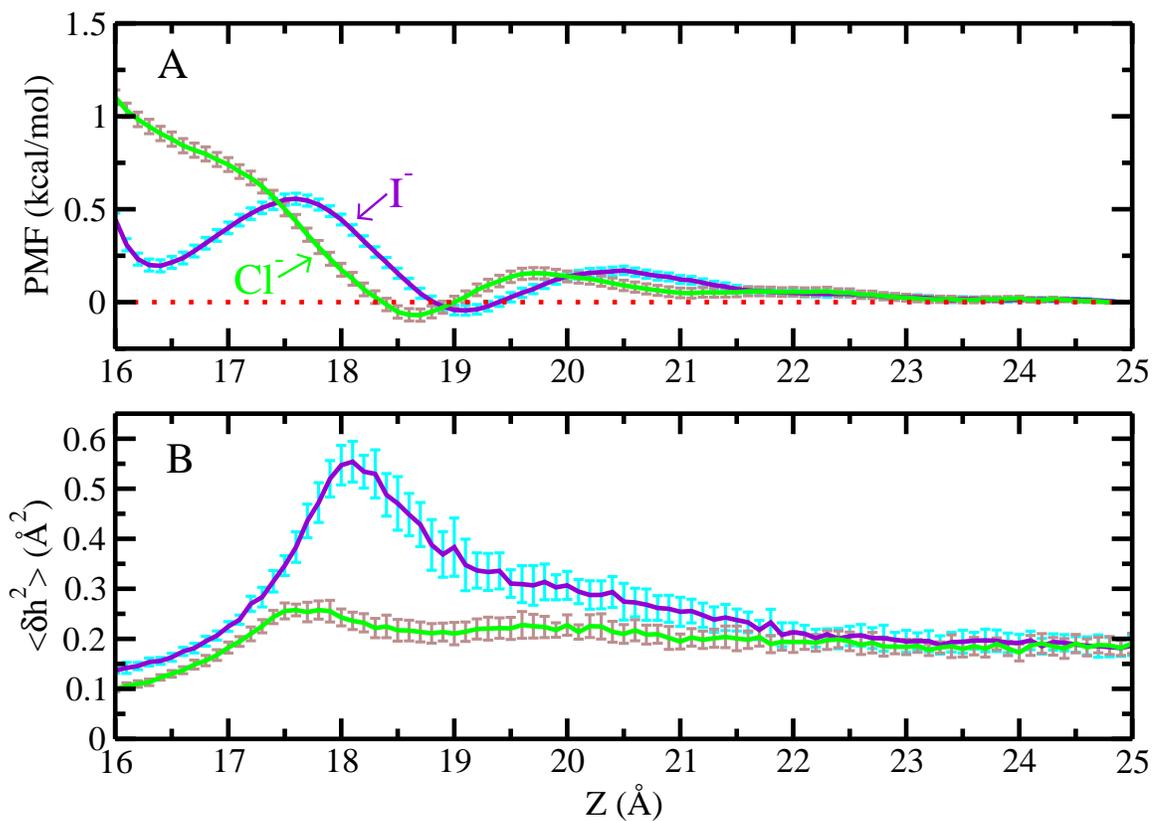


Figure 4.8: (A) PMF for single Cl^-/I^- approaching the hydrophobic protein-solvent interface with fixed protein and fixed anion (B) Hydrophobic protein-solvent interface fluctuation at $(x = 0, y = 0)$ as a function of anion restrained position for Cl^- and I^- with fixed protein and fixed anion.

the interface fluctuation. From the color scale, one can judge the magnitude of the fluctuation at each position around the whole protein. Panel A and B represent the situation that single Cl^-/I^- resides at $z = 24 \text{ \AA}$, in which case anions are far away from the protein interface and there will be no induced interface fluctuation. These are the inherent fluctuations of the interface, which are completely determined by the structural character of the protein itself. The Figure shows that one region manifests larger inherent fluctuation in Panel A and B. This region is in fact part of the largest hydrophobic patch of the protein. We will compare and discuss more about the inherent interface fluctuation among different regions of the protein, including hydrophobic, less hydrophobic and hydrophilic patches in the next subsection. As single Cl^-/I^- approach the hydrophobic interface, ion-induced perturbations of the aqueous interface around protein surface are more pronounced as reflected in Figure 4.9C and Figure 4.9D. These two figures depict the protein interface fluctuation when single Cl^-/I^- resides at $z = 18 \text{ \AA}$. Right above the position ($x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$) where single anion approaches the interface, we notice that fluctuations induced by I^- are much larger than those induced by Cl^- . As single anions move closer to the interface ($z = 16 \text{ \AA}$), this large difference of fluctuation between Cl^- and I^- lessens as shown in Figure 4.9E and Figure 4.9F. Due to the heterogeneous features of the protein surface, the extent of induced fluctuation is not perfectly symmetric about ($x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$). However, judging from Figure 4.10, we could find that ($x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$) is a feature point displaying largest induced fluctuations compared with other regions on protein surface as anions reside at various separations. To better illustrate the change in interface fluctuation magnitude as single anions move toward the point ($x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$), we plot $\langle \delta h^2(x = 0, y = 0) \rangle$ along the reaction coordinate in Figure 4.12A. We stress our intent to discuss the behavior of interfacial fluctuations as the anions move toward the patch; we are not interested solely in the nature of fluctuations when the anions reside at the interface. From the total 20ns production data, we obtained the fluctuations at this point by using every one nanosecond of data; the values shown here are the average of each one-nanosecond data

block and correspondingly, uncertainties were obtained based on the standard deviations. In the bulk region (z ranges from 24 to 25 Å), $\langle \delta h^2(x=0, y=0) \rangle$ is around 0.2² for both Cl⁻ and I⁻, which corresponds to the protein interface inherent fluctuation in the pure water due to the thermal fluctuations. For purpose of demonstrating and comparing the fluctuations induced from single Cl⁻/I⁻, we defined $\langle \delta h_L^2(x=0, y=0) \rangle$ as the normalized fluctuation value which is obtained via dividing $\langle \delta h^2(x=0, y=0) \rangle$ by the inherent fluctuation value, shown in Figure 4.12B. For the single Cl⁻ case, fluctuations almost remain the same as in the bulk. At $z = 17.5$ Å, slight enhancement of fluctuation was observed, with a normalized value of 1.36. In stark contrast, for the case of I⁻, onset of enhanced fluctuation relative to the bulk occurs at $z = 22$ Å. As I⁻ moves closer to the hydrophobic patch, induced fluctuations continue increasing and this enhancement reaches a maximum with a normalized value around 3.0 while I⁻ is located at $z = 18$ Å. Finally, the fluctuation is lower compared to the bulk when the anion is close to the interface. Comparing the trends of surface fluctuation as single Cl⁻/I⁻ move toward the hydrophobic protein interface and liquid-vapor interface, we find that in both cases the fluctuation is enhanced with presence of I⁻; however, there is only marginal perturbation of the interface by Cl⁻. We stress that this enhancement of interfacial fluctuations occurs as the ions approach the interface, not while they directly reside there.

Again, this originates, we claim, from the fact that Cl⁻ presents a more rigid hydration environment due to the more effective hydrogen bonding of water, thus decreasing the efficacy of promoting interfacial fluctuations. To visualize these different manners in which the hydration shells of Cl⁻ and I⁻ couple with the solvation structure at the hydrophobic protein interface, Figure 4.11A and B present the 180° angle-averaged radial water density around Cl⁻ and I⁻ as they reside at $z = 18$ Å, the position of maximum $\langle \delta h_L^2(x=0, y=0) \rangle$ for the anions. In this map, we only consider the water density distribution along positive z side, since single anion approaches the protein interface from this side. For the Cl⁻, the first hydration shell remains in its entirety as shown in the bright yellow ring. This implies that the hydration shell

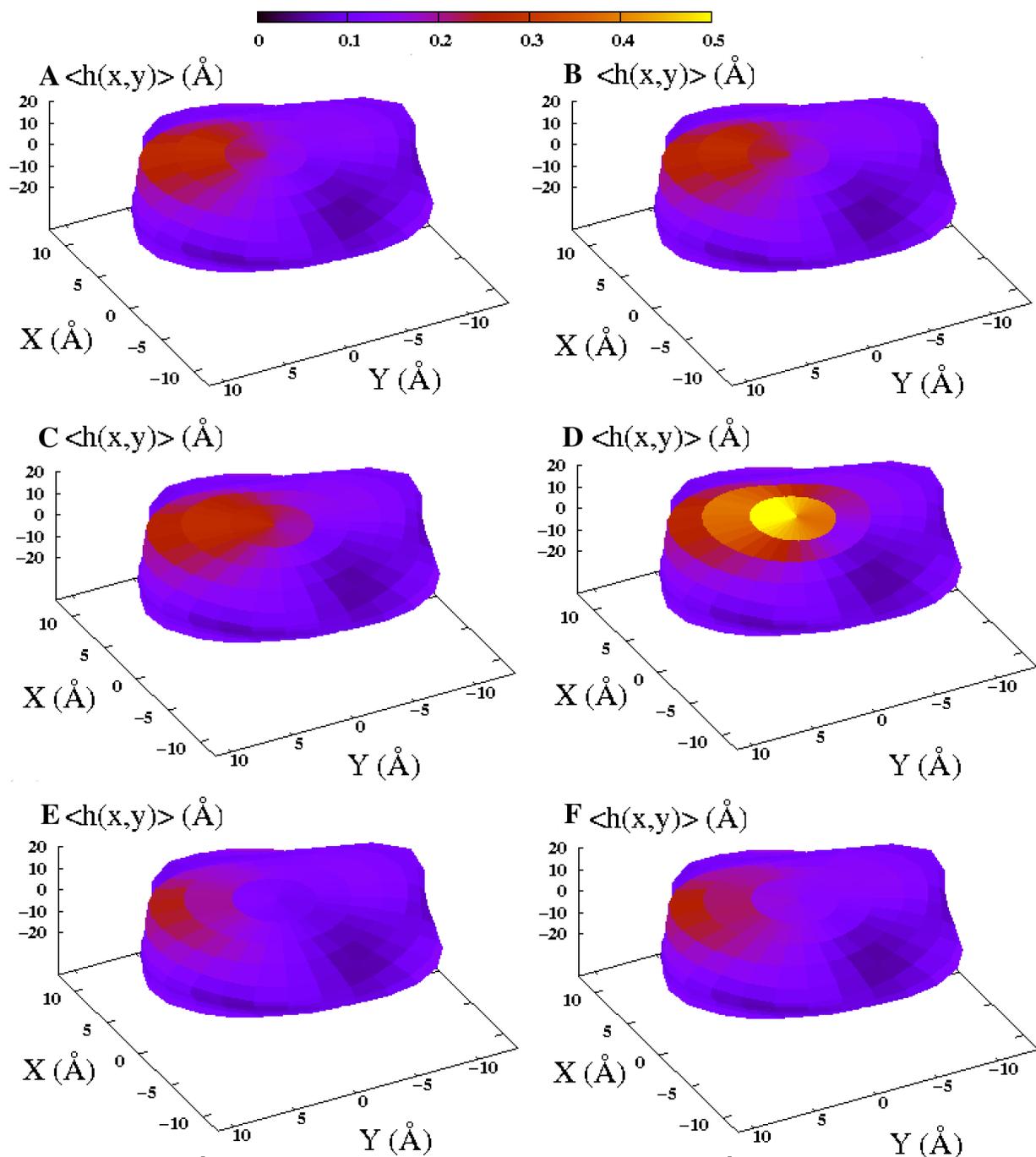


Figure 4.9: Protein-solvent mean interface $\langle h(x,y) \rangle$ and interface fluctuations $\langle \delta h^2(x,y) \rangle$ in single Cl⁻/I⁻ solution. The color scale represents the interface fluctuations (A) Cl⁻ residues at $z = 24$ Å (B) I⁻ residues at $z = 24$ Å (C) Cl⁻ residues at $z = 18$ Å (D) I⁻ residues at $z = 18$ Å (E) Cl⁻ residues at $z = 16$ Å (F) I⁻ residues at $z = 16$ Å.

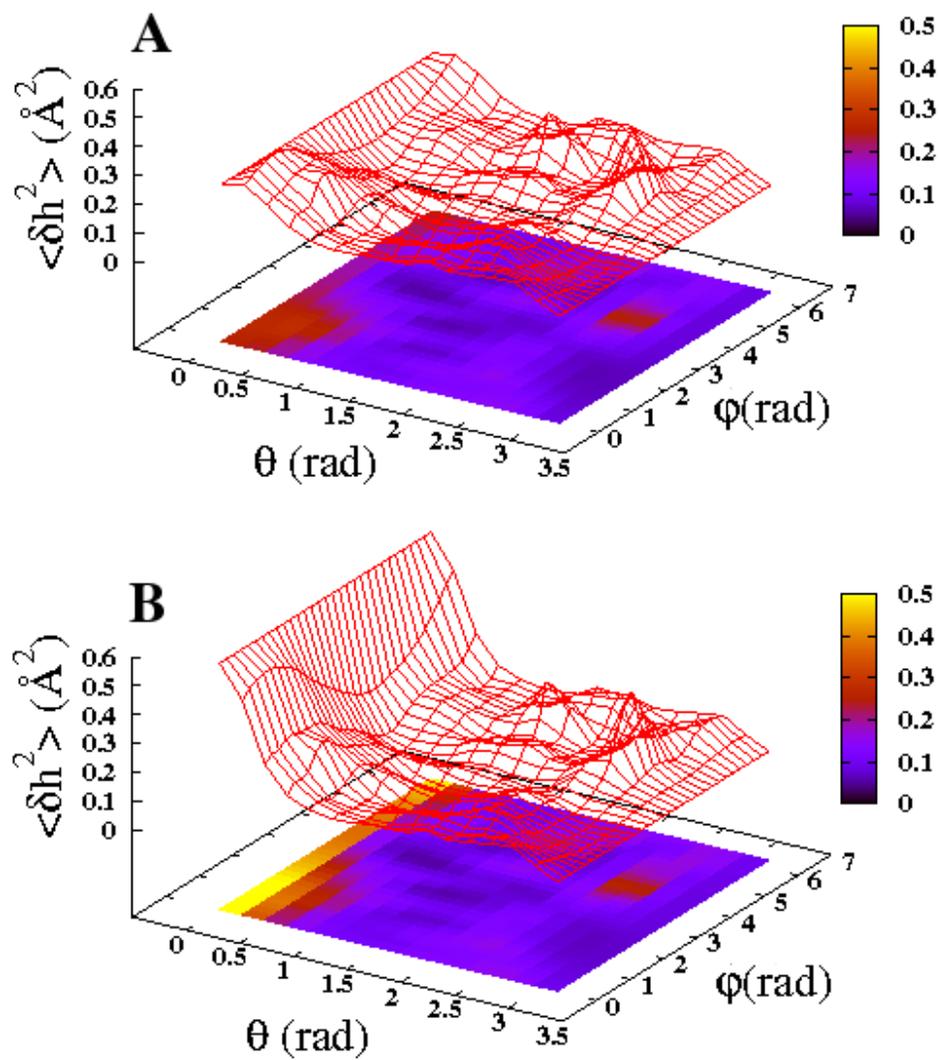


Figure 4.10: (A) Average aqueous protein surface fluctuation $\langle \delta h^2 \rangle$ as a function of θ and ϕ (A) Cl^- locates at $z = 18 \text{\AA}$ (B) I^- locates at $z = 18 \text{\AA}$.

environment for Cl^- is still quite rigid, well-ordered and tightly bound to the central anion, which will not cause an increased dynamical perturbation of local solvent (the Cl^- will not give up local solvation water unless there is a sufficiently acceptable ligand to substitute in water’s place); I^- , in contrast, possesses the first hydration shell that is weakly-bound and less-ordered, so that it has more tendency to break, as shown in Panel B, the bright yellow ring was broken at some region. This malleable hydration layer accommodates greater coupling with the solvation shell of the protein interface, consequently, inducing a larger interface fluctuation. For a comparison, we also shown the density map at $z = 19 \text{ \AA}$ in Figure 4.11C and D, a little ahead of the position of largest fluctuation. In our recent studies, we have demonstrated a connection between L-V interfacial stability of chemical species and the extent to which the presence of these molecular species approaching the interface induces collective fluctuations of the interface in addition to the level inherent in pure water due to thermal motion. Next, we would like to also discuss the induced protein interface fluctuation difference for Cl^- and I^- as a further contribution in explaining their differences in free energy profiles approaching the hydrophobic patch; the contribution arises in the context of a mechanistic view of how the system ultimately finds stability with I^- near the interface. We observe that the iodide anion induces larger fluctuations on approach to the interface; this increases interface entropy (based on References [56, 55]). This increased interface entropy may contribute to differentially stabilizing microstates where the iodide is closer to the interface compared to chloride. Based on the potentials of mean force of Figure 4.6, the highest induced fluctuations correspond to barrier states. The fluctuations induced by iodide, being larger than for chloride, may tend to lower the barrier required for the iodide to move to the interface. Thus, the fluctuations provide a mechanism for iodide ultimately presenting at the interface.

We pause here to address potential artifacts in our algorithm for computing interfacial fluctuations. One may ask whether the instantaneous coarse-grained interface we construct can artificially pass “through” the ion, thus giving rise to artificially large fluctuations. To explore this, we plot the difference in the z -position of the ion center

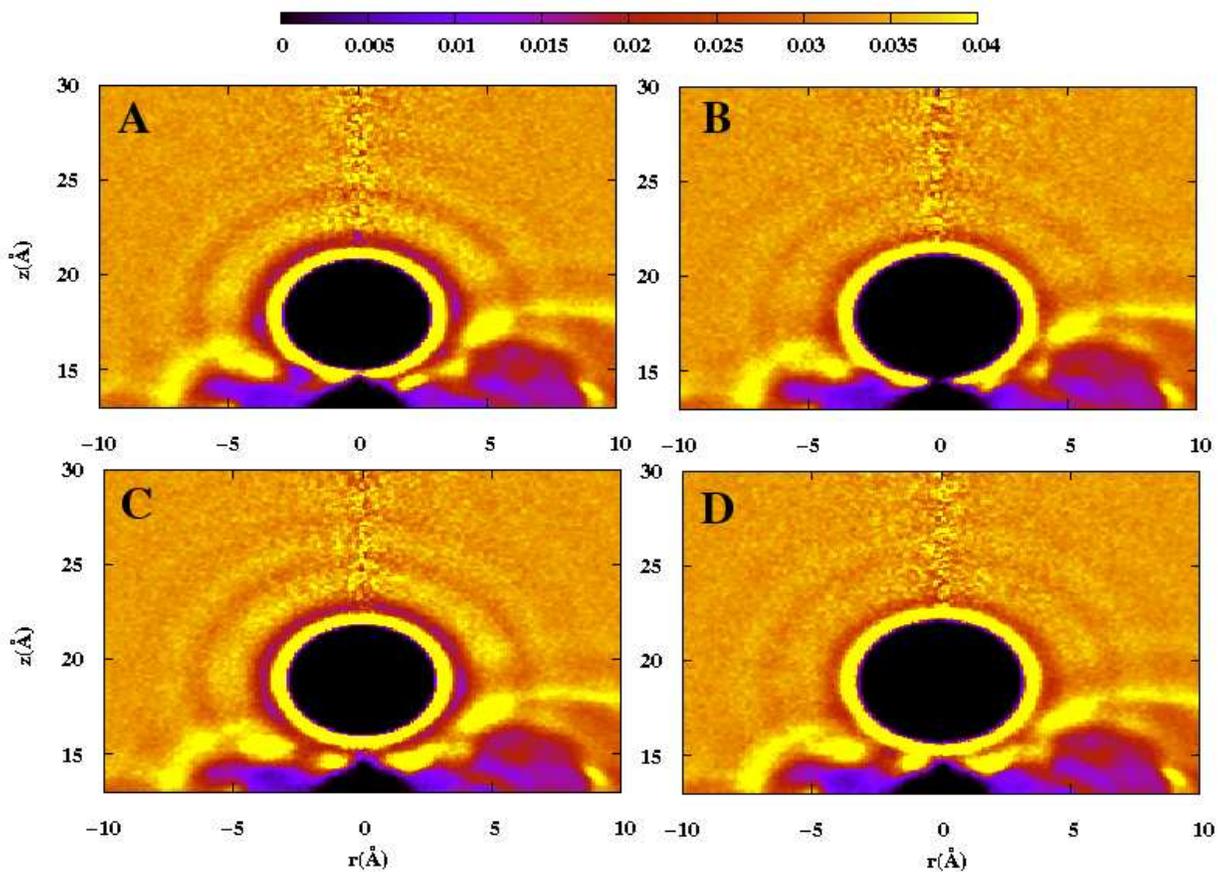


Figure 4.11: Average water oxygen density around (A) Cl^- at position $z = 18 \text{\AA}$ (B) I^- at position $z = 18 \text{\AA}$ (C) Cl^- at position $z = 19 \text{\AA}$ (D) I^- at position $z = 19 \text{\AA}$. X axis represents the lateral distance $r = \sqrt{x^2 + y^2}$ and Y axis represents the distance from positive z direction.

(z_{ion}) and the z-position of the interface ($z_{interface}$) as the ion moves toward the protein along the axis passing through the z-axis; that is, we plot the difference in these positions for different values for each simulation window. Thus, the z-position of the interface is equal to the value of the surface height of the interface at the point ($x=0, y=0, z_{interface}$), and the z-position of the ion center is identically the z-position of the ion. If the interface is between the ion and the protein, we will see a positive value; if the interface moves “through the ion”, we will get zero; if the ion resides between the interface and the protein, the value will be negative.

In our system, due to the strong restraint applied on the ion, the distribution of the corresponding ion’s z-position (z_{ion}) in each simulation window is narrow (0.1 Å). Consequently, for each window, by subtracting basically the same z_{ion} , the distribution of the instantaneous interface’s z-position ($z_{interface}$), which correlates with the interfacial fluctuation in our manuscript, essentially has the same width of the distribution for ($z_{ion}-z_{interface}$). The question arises whether the algorithm we use artificially includes all three scenarios ($z_{ion}-z_{interface} > 0, = 0, < 0$) in some simulation windows, and in this way suggesting larger fluctuations. We will show that even when all $z_{interface}$ values are distributed on one side of the ion (all positive/negative values for $z_{ion}-z_{interface}$), the distribution of $z_{interface}$ is not necessarily small, i.e. the induced fluctuations are non-artificial. Figure 4.13 shows that for just about all positions of I^- greater than 16.5Å, the interface resides between the protein and the ion. The interface does not pass through the ion center. There are some values less than zero when the ion z-position is 16.5Å, but at this point, we see suppression of interface fluctuations (Figure 4.12). Finally, we consider the same analysis taking the interface position to be the height of the surface at different x and y positions (in addition to a variety of z-positions). This is shown in Figure 4.14. This again shows the same behavior as Figure 4.13. Based on this analysis and to the best of our ability at this time, we believe that the induced fluctuations we report are reliable and robust.

To close this section, we attempt to evaluate hydrophobic interface fluctuations

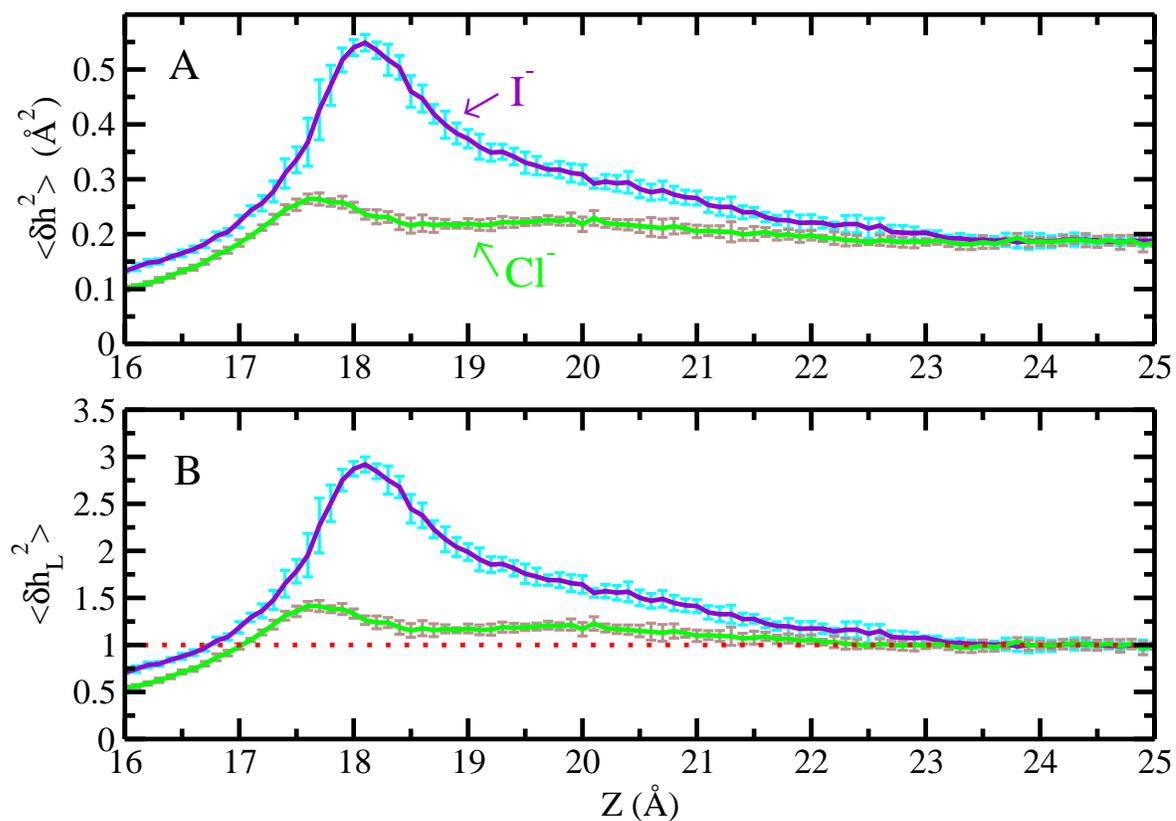


Figure 4.12: (A) Hydrophobic interface fluctuation at $(x = 0, y = 0)$ as a function of anion restrained position for Cl^- and I^- (B) Normalized interface fluctuation at $(x = 0, y = 0)$ as a function of anion restrained position for Cl^- and I^- .

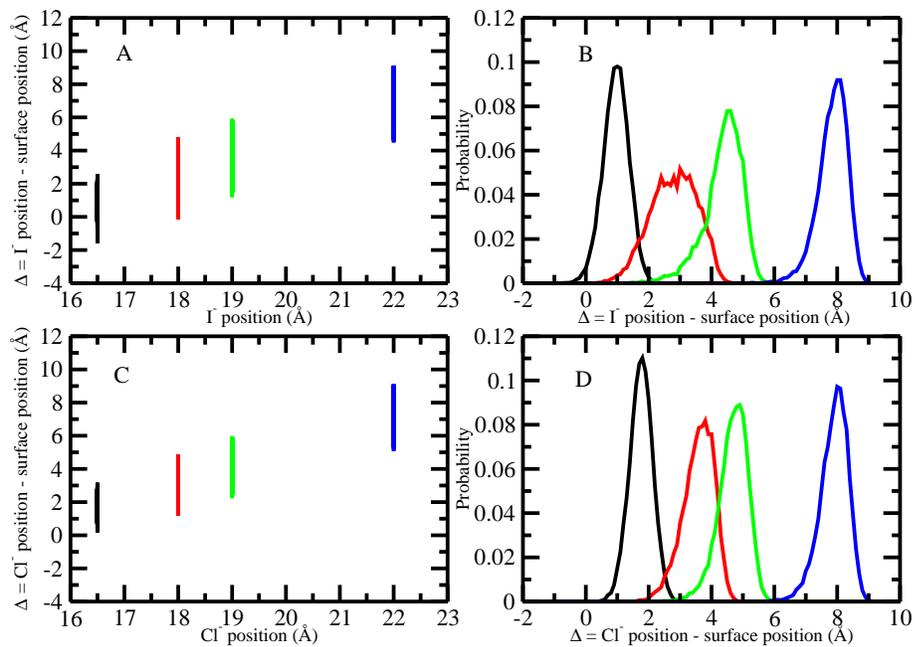


Figure 4.13: Differences between ion position and surface position (surface position is defined by the surface height at position $x = 0$, $y = 0$) for I^- and Cl^- at various positions from 16.5 (black), 18 (red), 19 (green) and 22 (blue).

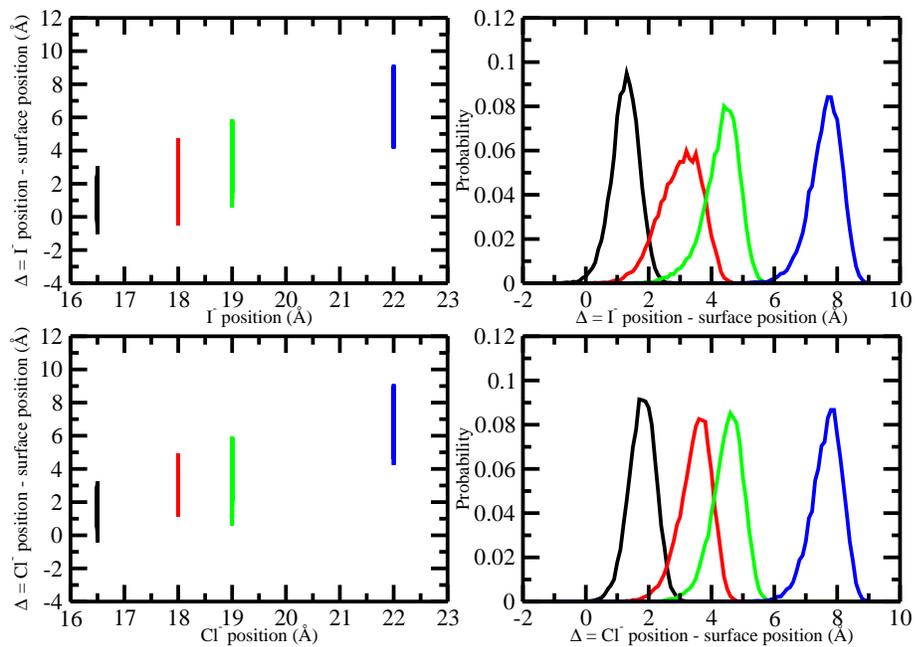


Figure 4.14: Differences between ion position and surface position (surface position is defined by the surface height around position $x = 3$, $y = 0.6$) for I^- and Cl^- at various positions from 16.5 (black), 18 (red), 19 (green) and 22 (blue).

allowing for protein flexibility. Instead of freezing all protein atoms, we allow modest vibrational degrees of freedom of the protein. Since the real proteins in biological system are not motionless, it is meaningful to address whether the different perturbations of interfacial fluctuations induced by Cl^- and I^- persist in the case of a flexible protein surface. For the convenience of evaluating the interface fluctuation around specific regions of the protein in the external coordinate system without worrying about translation and rotation of the protein in space, translational and rotational motions of the protein were first removed from the MD trajectory by using “MERGE ORIENT” module of CHARMM. RMSD based on the backbone protein atoms are shown in Figure 4.15. The RMSD values are less than 2.5 Å in all cases as Cl^-/I^- locate around protein surface and in the bulk. Aqueous protein interface was constructed using new trajectories based on the same protocol. Figure 4.16 shows the hydrophobic interface fluctuation profiles at $x = 0$ and $y = 0$ as a function of z -position of Cl^- and I^- approaching the flexible protein. When the single anion is in the bulk, fluctuation is about 0.3 \AA^2 for both anions, higher than the inherent fluctuation of the interface around the fixed protein, which is about 0.2 \AA^2 . This makes sense since inherent fluctuation of the protein interface is not only derived from thermal motion of water, but also from that of protein itself. Consistent with the fixed protein outcomes, I^- induces larger fluctuations than Cl^- nearing the patch, with the maximum value of 0.56 \AA^2 higher than that of Cl^- 0.43 \AA^2 at the location of $z = 20 \text{ \AA}$.

4.3.4 Less Hydrophobic and Hydrophilic Protein Interface

We now turn to the process single Cl^-/I^- approaches the aqueous protein interfaces with different hydrophobicity. We also start with PMF, representing the reversible work for Cl^-/I^- transferring from the bulk to the regions around the protein-water interfaces that we are interested in. Figure 4.17A presents the PMF for single Cl^-/I^- approaching the less hydrophobic protein-solvent interfaces. The PMF shows a minimum of $-0.06 \pm 0.04 \text{ kcal/mol}$ for the single Cl^- and $-0.16 \pm 0.04 \text{ kcal/mol}$ for the single I^- at position around 20 \AA for both, which is further emphasized in the small inset.

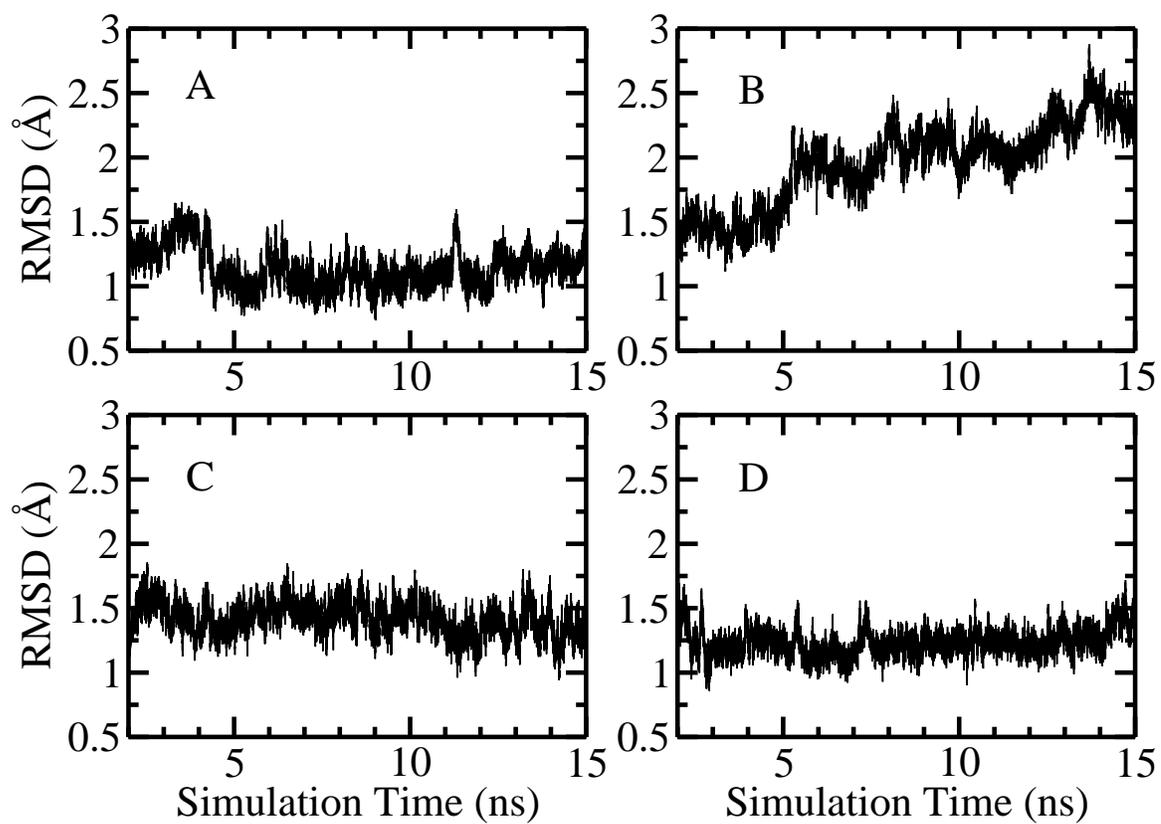


Figure 4.15: Evolution of protein backbone RMSD as (A) Single Cl⁻ locates at $z = 16 \text{ \AA}$ (B) Single I⁻ locates at $z = 16 \text{ \AA}$ (C) Single Cl⁻ locates at $z = 25 \text{ \AA}$ (D) Single I⁻ locates at $z = 25 \text{ \AA}$.

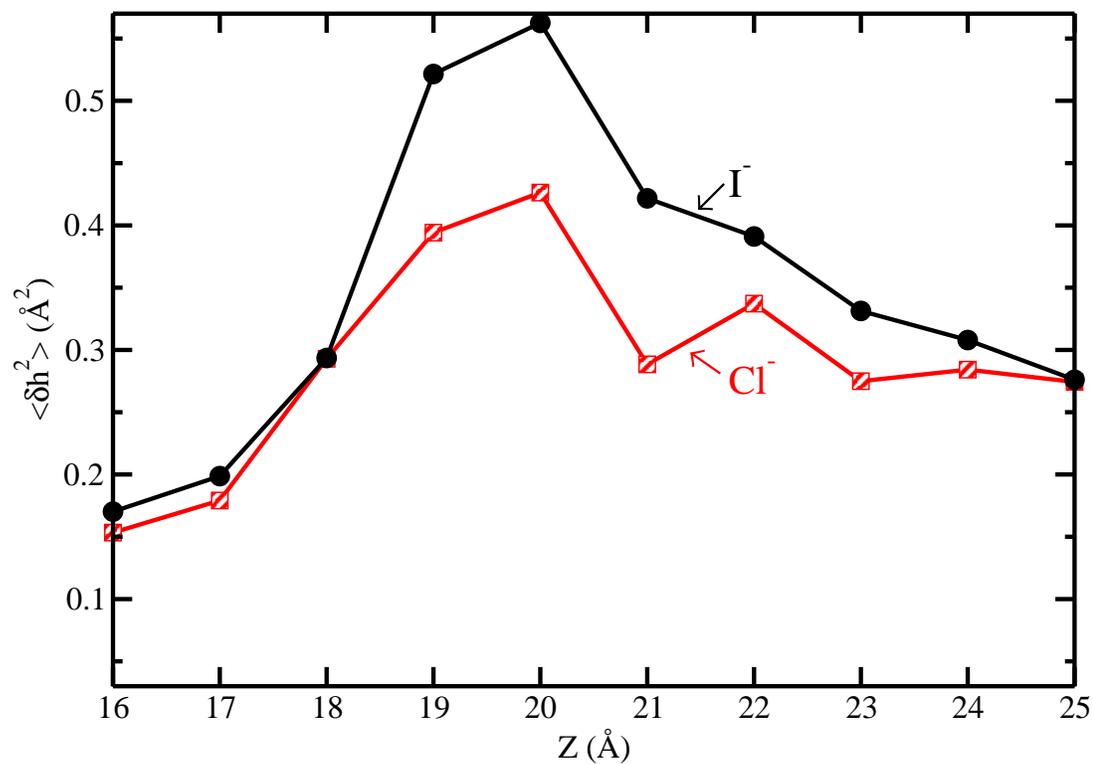


Figure 4.16: Hydrophobic interface fluctuation at $(x = 0, y = 0)$ as a function of anion Z -position for Cl^- and I^- in the case of flexible protein.

Relative to the state with ion in bulk, there is effectively no stabilization. The main differences in PMF between Cl^- and I^- appear in the range from $z = 15.5 \text{ \AA}$ to $z = 17.0 \text{ \AA}$. Unlike the Cl^- PMF in this range, which continues increasing, there is a minimum in the PMF profile for I^- . Consistently, the PMF for I^- shows slightly higher stability than that of Cl^- in this range. Figure 4.17B shows the PMFs for Cl^- and I^- approaching a hydrophilic region. The PMF shows a global minimum of -0.35 ± 0.06 for single Cl^- and -0.24 ± 0.05 for single I^- at position 14.7 \AA and 15.1 \AA respectively as they approach the hydrophilic protein-solvent interfaces. (shown more clearly in the inset). They suggest a modest stabilization effect from both Cl^- and I^- as they are in the vicinity of the hydrophilic region around protein interfaces. In summary of the PMF as single Cl^-/I^- approaches three different regions on the protein interfaces with different hydrophobicity, we find significant differences arising as single Cl^-/I^- is close to the interfaces from $z = 14 \text{ \AA}$ to $z = 17 \text{ \AA}$. For Cl^- , when it is close to the hydrophobic and less hydrophobic region, there are no free energy minima, and the free energy values are positive. For I^- , although the free energy values are still positive, they are lower (with the largest difference about 1 kcal/mol) than those of Cl^- . Minima are observed in this region for the I^- but not for Cl^- . However, around hydrophilic interfaces, both Cl^- and I^- have minima. This reflects the fact that for both Cl^- and I^- , there are more free energetic advantages as they are close to the hydrophilic regions, compared with the hydrophobic ones of HFBII protein, which may due to the favorable direct anion-charged residue interactions around the hydrophilic protein interfaces. Interestingly, our results of PMF for Cl^-/I^- when they are around hydrophobic and hydrophilic residues of HFBII protein follow the similar trend for the previous published work by Lund et al. [60]. They compared the free energetics of F^- and I^- around a spherical model macromolecule. Here, F^- is a small, highly charge-dense and fully hydrated anion similar to Cl^- . They suggest that when the macromolecule is uncharged and considered as a hydrophobic particle, I^- has more free energy advantage than F^- for being near the interface. When the macromolecule is positive charged and considered as a hydrophilic particle, the trend reverses, F^- is

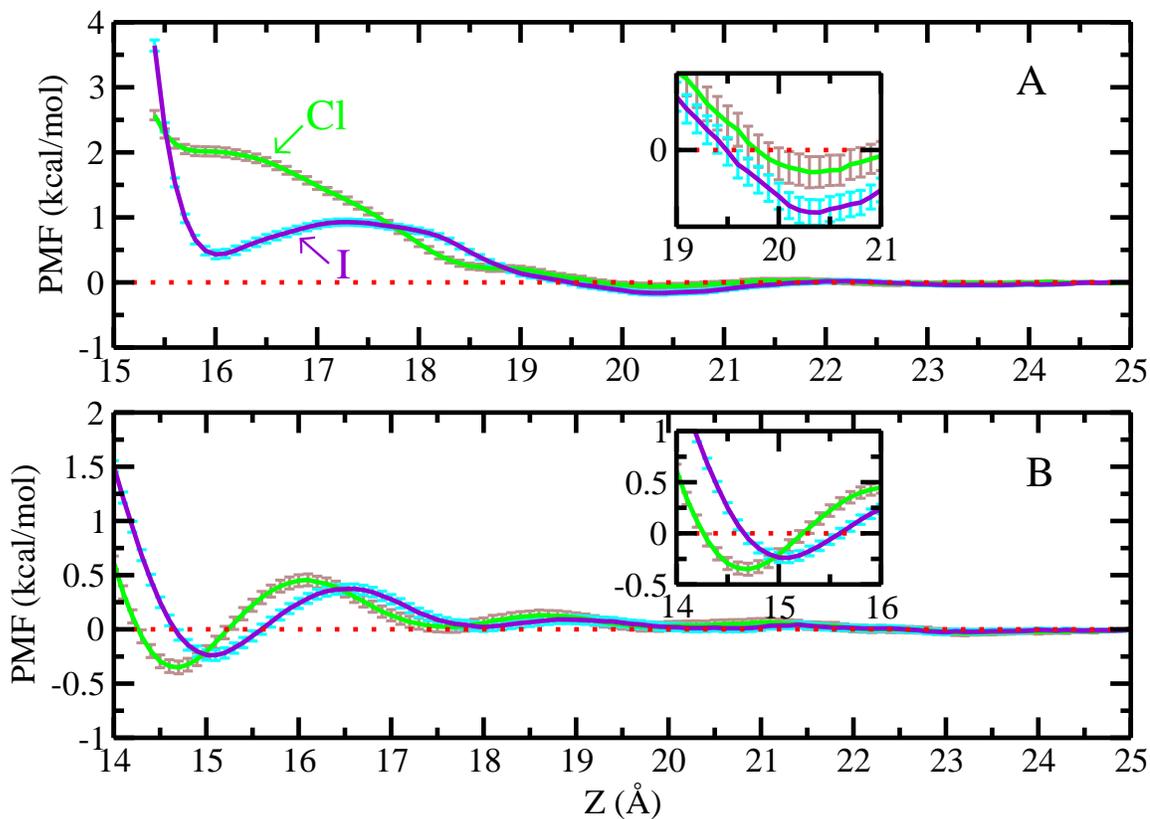


Figure 4.17: (A) PMF for single Cl^-/I^- approaching the less hydrophobic protein-solvent interfaces (B) PMF for single Cl^-/I^- approaching the hydrophilic protein-solvent interfaces.

more favorable around the macromolecule. Also, comparing the free energetics of the same anion around the hydrophobic and hydrophilic sphere, Lund et al find that both F^- and I^- are more stable around the hydrophilic particle.

Next, we consider interface fluctuations. First we evaluate the inherent fluctuations (absence of anions) of different interfacial regions of the protein as reference. Figure 4.18 shows a colored map of HFBII protein interface based on the magnitude of interface fluctuations in TIP3P water. The color scheme from red to blue represents fluctuation spectrum from higher to lower values. Since there are no other impurities in the system, the inherent interface fluctuations are derived from the thermal fluctuations of the water. As shown in Panel A, regions defined as hydrophobic interfaces (V18, L19,

L21, I22, V24, V54, A61, L62 and L63) possess the largest fluctuations while the selected hydrophilic interfaces in Panel D (D25, C26, K27, T28, A58, D59, Q60) manifest the lowest fluctuations. The less hydrophobic interface (Panel C) displays a moderate fluctuation. This suggests that the magnitude of interface fluctuation correlates with the surface hydrophobicity. This is consistent with Garde’s insights [28] that density fluctuations are enhanced near hydrophobic surfaces while reduced with increasing hydrophilicity. This enhanced density fluctuation is explained as a consequence of more facile cavity formation, increased compressibility of hydration water, and more favorable binding of hydrophobic solutes. Although in this work the fluctuation we address is based on the aqueous protein interface height, which is not exactly the same as water density fluctuation Garde et al[28]. apply, it reflects similar information about the malleable nature of the water around hydrophobic patch, considering that the aqueous protein interfaces we construct were based on the coarse-grained solvent densities at each space-time point.

We now address fluctuations induced by the anions. Figures 4.19A and B show fluctuation profiles as Cl^-/I^- approach the less hydrophobic and hydrophilic protein interfaces, respectively. Compared the fluctuations of distinct protein interfaces as anions in the bulk, in previous section we note this value for hydrophobic region is about 0.2 \AA^2 ; in the less hydrophobic interface, it is about 0.1 \AA^2 ; and in the hydrophilic interface, it is about 0.07 \AA^2 . These differences correlate with the inherent protein interface fluctuations of Figure 4.18. As single Cl^-/I^- moves closer to the less hydrophobic interface, I^- induces more interfacial fluctuation than Cl^- , especially in the range from $z = 18$ to $z = 19$. The magnitude of the difference is up to 0.2 \AA^2 , similar to the hydrophobic interface value of 0.3 \AA^2 . Comparing this profile with that of the hydrophobic interface in Figure 4.12A, the induced fluctuations are significant from I^- while marginal from Cl^- ; global maxima can be detected in the I^- fluctuation profiles at the location of $z = 18.0$ and 18.5 for hydrophobic interface and less hydrophobic interface respectively. In the case of the hydrophilic interface, both Cl^- and I^- have inappreciable effect on hydrophilic interfacial fluctuations. Although I^- may

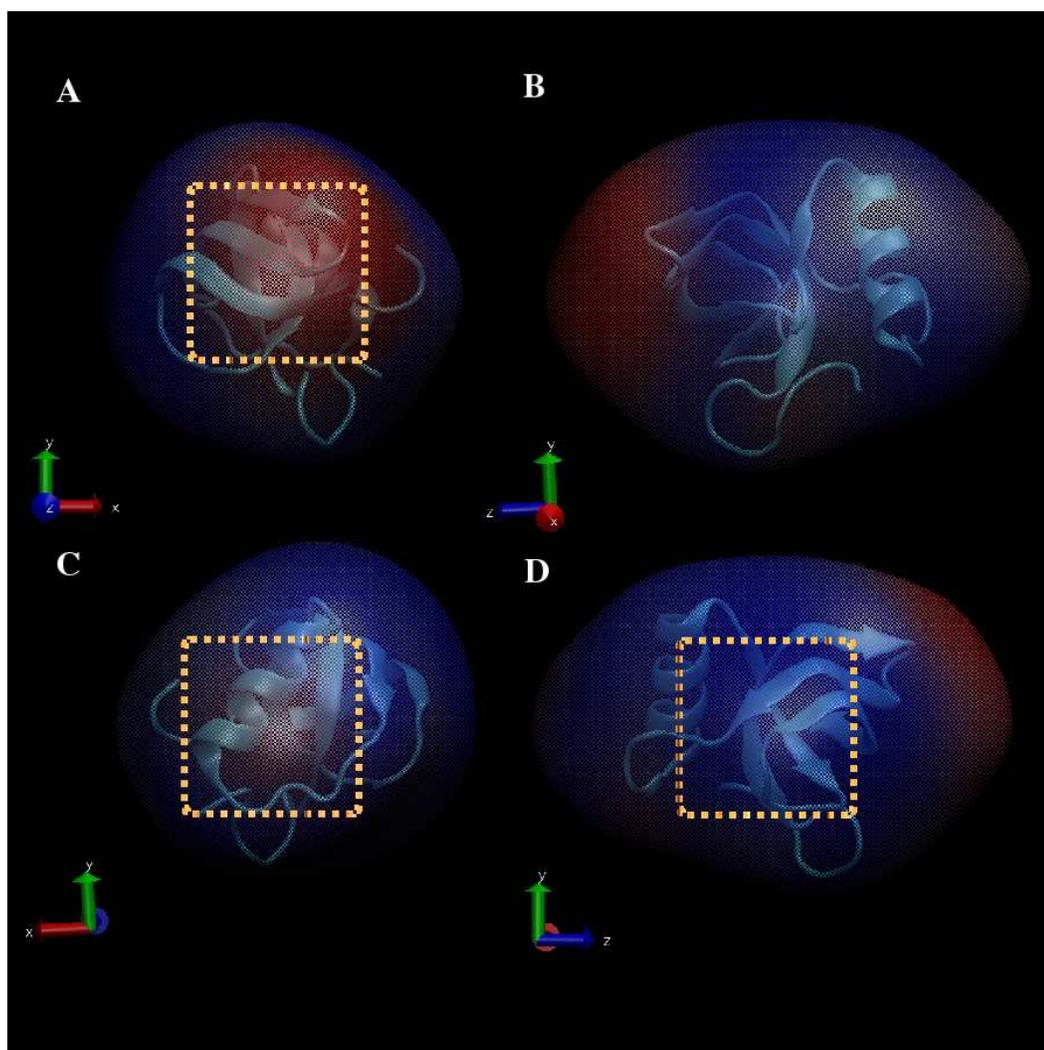


Figure 4.18: Inherent interface fluctuations of HFBII. For A, B, C and D, each one depicts one side of the protein interface with a rotation of 90° respectively. Red colour represents larger fluctuations, while blue colour represents smaller fluctuations. The highlight regions in A, C and D corresponds to the hydrophobic, less hydrophobic and hydrophilic regions that we define in this study.

induce a little larger fluctuation compared with Cl^- as it moves closer to the interface, the differences are quite small, with a value of 0.02 \AA^2 , only one-tenth of that from the less hydrophobic interface. In this picture, our suggestion is that the extent of the difference is highly related to the nature of the protein interface. The hydrophilic interface borders a rigid water environment that is difficult to couple with both the hydration shells of Cl^- and I^- . Consequently, Cl^-/I^- approaching the hydrophilic interface induce marginal interfacial fluctuations, and the difference between induced fluctuations of the two anions is less; however, for the hydrophobic interface and less hydrophobic interface we defined, the water shells around these regions are malleable, so they can exchange solvation water with that of I^- , which also possesses a less rigid solvation shell. However, due to the more severe ordering of water around Cl^- , it is not possible for water around hydrophobic interface to perturb the solvent around Cl^- . Therefore, as Cl^- and I^- approach this type of hydrophobic interface, significant differences appear in their ability to induce hydrophobic interfacial fluctuations.

4.4 Summary and Conclusions

Building upon the insights gained from the vast studies of specific ion behaviors at aqueous liquid-vapor interfaces, we have presented here a discussion regarding the unique fluctuation inducing properties of two anions for which the degree of induced interfacial fluctuations correlates with stability at the interface. Our major conclusions are for hydrophobic protein-water interfaces, and this particular nature of the interface is chosen as it is a logical extension of the ideally hydrophobic interface presented by the aqueous liquid-vapor context. Our control system, the aqueous liquid-vapor interface, recapitulates earlier specific ion behavior, namely that the less-charge dense, larger iodide anion demonstrates a slight surface propensity as embodied in a free energy stable state compared to chloride. Moreover, our results for the anions at the aqueous liquid-vapor interface recapitulate recent studies correlating surface propensity to ability to induce interface fluctuations[55, 57, 56]. At the interface between a hydrophobic region of a protein, in this case HFBII, and the aqueous solvent, we

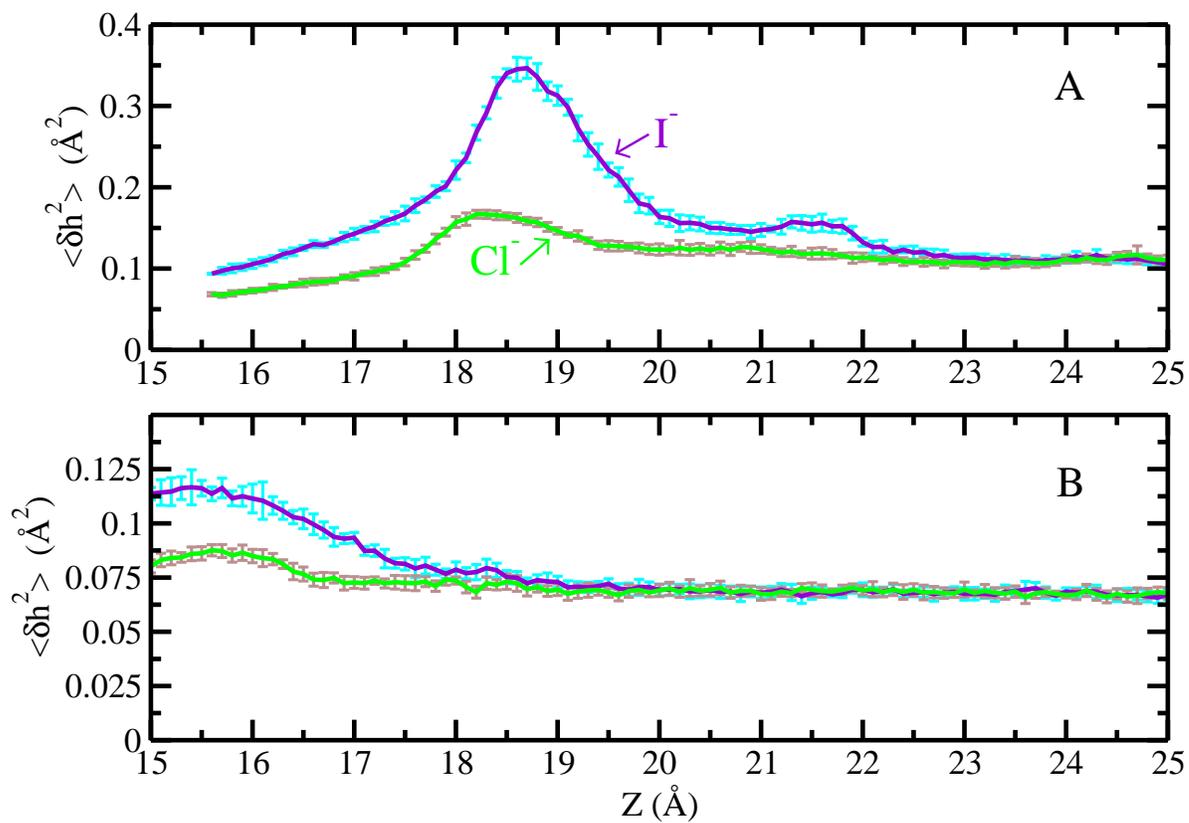


Figure 4.19: (A) Less hydrophobic interface fluctuation at $(x = 0, y = 0)$ as a function of anion restrained position for Cl^- and I^- (B) Hydrophilic interface fluctuation at $(x = 0, y = 0)$ as a function of anion restrained position for Cl^- and I^- .

find that potential of mean force calculations reveal a lower free energy state for iodide than chloride, the trends qualitatively consistent with those observed at the liquid vapor interface. Furthermore, we find that the more surface stable iodide also induces significantly larger interface fluctuations on approach to the interface compared to the smaller, more charge-dense chloride; this is again in keeping with observations at the aqueous liquid-vapor interface. These behaviors approaching the hydrophobic interface, are related to the coupling of local hydration water in the vicinity of the protein with the hydration water around the individual anions; specifically, the differential ability of the water environments to couple with one another in the case of chloride and iodide leads to the specific-ion behavior as it is related to induced interfacial fluctuations. Approaching interfaces at the other extreme, hydrophilic interfaces, we observe that both anions display similar behaviors in terms of surface stability and induced interface fluctuations. These differences offer a view of the anions as having different characters in different contexts. Where strong local interactions are not dominant, as in the case of hydrophobic surfaces that lead to higher fluctuations in general (i.e., higher solvent density fluctuations[36]), the anions tend to differentiate themselves based on their “hydrophobicity”; the large, less charge-dense iodide has a higher propensity to associate with hydrophobic regions due to its inherent higher “hydrophobicity”. The smaller, more charge-dense, less hydrophobic chloride is not a stable at a hydrophobic interface. The idea of specific-ion behaviors at interfaces being related to hydrophobic solvation has been put forth recently, and we suggest that the current results present another manifestation of the differential hydrophobic character of ions at specific interfaces[51]. In the case of hydrophilic interfaces presenting highly polar and charged species, the strong charge-dipole and charge-charge interactions dominate and equalize the stabilities and interface perturbing effects of both ions.

Chapter 5

ORIENTATIONAL PREFERENCE OF GUANIDINIUM CATION AND UREA DENATURANTS AROUND EFFECTIVE HYDROPHOBIC REGIONS OF PROTEIN SURFACE

Reproduced with permission from

Di Cui, Shuching Ou, Sandeep Patel. "Protein Denaturants at Aqueous-Hydrophobic Interfaces: Self-Consistent Correlation between Induced Interfacial Fluctuations and Denaturant Stability at the Interface." *Journal of Physical Chemistry B*. **2015**, 119 (1), 164-178. Copyright © 2014, American Chemical Society

5.1 Introduction

The pursuit for a global and self-consistent conceptual, mechanistic, and theoretical framework within which to discuss the denaturing properties and behaviors of cosolvents such as urea and guanidinium chloride (GdmCl) continues to garner a significant amount of scientific curiosity and effort [224, 225, 226, 227, 88, 228]. The quest for a fundamental understanding of protein denaturation has a long and rich history, to which the reader is referred [67, 68, 69, 70, 71, 72, 61, 73, 74, 75, 76, 77, 78, 79, 80, 43, 81]. Based on recent experimental and molecular simulation studies, the notion of direct interactions of denaturants with proteins in solution has come to be accepted in consensus. Since common denaturants used in practical situations are needed in significantly high concentrations, i.e., 5 M urea for instance, the notion that there are no direct interactions between denaturant and protein becomes less justifiable [88]. Within the context of direct interactions, one of two major mechanisms for denaturation involves the lessening of the hydrophobic effect as it relates to the formation of a compact "pre-folded" ensemble of states where protein hydrophobic surface exposure to solvent is

reduced in relation to the purely unfolded ensemble of states. The idea is that by associating with hydrophobic regions of the protein (specific residues, clusters of residues forming extended topographical “surfaces”, hydrophobic sidechains, etc), denaturant molecules can shield the hydrophobic surface area even in unfolded or extended configurations of the peptide/polymer. This chemical denaturation mechanism naturally involves direct interaction of the cosolvent molecule with regions of the protein surface. A particular aspect of this interaction deals with the precise nature of association geometries and the associated free energetics; specifically, molecules such as urea, and more so guanidinium cation (Gdm^+), can present several predominant relative orientations to the protein surface through which the interaction is mediated. In general, it is proposed that a dominant interaction of urea with surface groups in protein simulations involves hydrogen bonding with polar side-chain functions [83, 84], while the unique hydration properties of the Gdm^+ [85] support alternative interaction modes involving stacking with side-chain planar and hydrophobic groups. However, we should note that the nature of the relative orientations would be dictated in part by the nature of solvation and hydrophobic effects as they pertain uniquely to each denaturant molecule. Understanding of the precise geometrical and associated free energetic properties of denaturant-protein interactions is important as a piece in a more complete understanding of the denaturation process from a molecular perspective. Previous studies have shown that cosolvents such as Gdm^+ adopt orientations relative to “flat”, model hydrophobic surfaces that are planar. These hydrophobic surfaces include the aqueous liquid-vapor interface [86, 87], flat hydrophobic plate [75, 88] and hydrophobic polymer surface [77]. However, there is a lack of direct evidence for similar orientational behavior of Gdm^+ upon approaching more complex aqueous protein interfaces. The inherent chemical and topographical heterogeneity of protein surface makes it difficult to find a qualitatively rigorous approach to evaluate the relative orientation between the surface of Gdm^+ and the protein. To fill this gap, we apply molecular dynamics simulations investigating the association of Gdm^+ cation with a specific protein with a relatively flat surface region consisting of hydrophobic residues. In the context of

chemical denaturation via direct association, we ask here about the orientations that Gdm^+ and urea adopt when interacting with hydrophobic regions of proteins. The combination of this analysis addresses ideas of direct interaction as well as hydrophobic effects as they pertain to the denaturation process. Furthermore, there is sentiment in the literature demonstrating the importance of solvent fluctuations and their relation to what is called the hydrophobic nature of solutes. For example, using molecular dynamics simulations, Godawat et al [36] found that water density near the surfaces of self-assembled monolayers (SAMs) with hydrophobic head groups ($-\text{CF}_3$, $-\text{CH}_3$) shows a poor distinction from that of SAMs with hydrophilic head groups ($-\text{OH}$, $-\text{CONH}_2$). However, differences arise when considering the fluctuations of water density near the two regions. Enhanced fluctuations, reflected by the broad probability distributions of water number density are observed around hydrophobic surfaces compared with the bulk solution and hydrophilic surfaces [28, 34]. Moreover, the enhanced density fluctuations around hydrophobic surfaces are further characterized by more compressible hydration shells and increased cavity formation, [167, 168] indicating that the nature of hydration shells around hydrophobic surfaces are softer and more flickering than near hydrophilic ones. Since the long-ranged solute-induced perturbations of aqueous protein interfaces involve the coupling of local hydration shells of the solutes with distant hydration shells around protein surfaces, the natures of both would affect the extent of induced interfacial fluctuations. It would be interesting to compare the interface height fluctuations as Gdm^+ /urea approaches the hydrophobic/hydrophilic protein regions. We note that the interface height fluctuations we are pursuing here are conceptually different from the density fluctuations of References [28, 34, 27, 229], though both reflect the nature of hydration water around the protein surfaces. It is natural here to investigate the nature of induced fluctuations of the solvent at the protein-water interface via consideration of the fluctuations of the height of this interface (once defined in a well-controlled manner) upon approach of a denaturant molecule to a hydrophobic protein region as well as when the denaturant resides at very close separation to the protein-water interface.

The particular protein on which we are focusing in this study is hydrophobin-II (HFBII), a small protein expressed by filamentous fungi. The protein is known for its ability to form hydrophobic coatings on surfaces and self-assembles into monolayers on hydrophobic/hydrophilic interfaces such as the water/air interface [183, 230, 231, 232]. These behaviors are mainly determined by the protein’s amphiphilicity. Acharya et al. [27] mapped the effective hydrophobic regions and effective hydrophilic regions of HFBII by considering the density of small probe hydrophobic solutes around each region of the protein. They selected three regions with different hydrophobicity based on this and further monitored the density fluctuations in the vicinity of these regions. Their calculation shows that the largest density fluctuations occur around the most hydrophobic region whereas the least density fluctuations are detected around most hydrophilic region. This particular observation suggests hydrophobins as useful candidate proteins for comparing behaviors at hydrophobic and hydrophilic interfaces as denaturant molecules approach. We note that the purpose of this study is to demonstrate the specific denaturant’s stability and orientational preference around regions with different hydrophobicity of the protein with implication of the direct interaction as well as hydrophobic effects for the association between denaturant and the protein. The aim of this study does not focus on the denaturation process by these denaturants, so we use the totally fixed protein in the simulation along with quite low concentration of denaturants (1 M and an extreme case, single solute) compared with the significant high concentration (up to 5 M) in the actual denaturation experiments. We further emphasize that by using the single solute in this study, it is possible for us to systematically distinguish underlying characters of stability for different species (Gdm⁺ and urea) and orientational preference for different orientations (parallel and perpendicular relative to the regions of interest).

The paper is organized as follows. In Section 5.2 we discuss the simulation protocols and computational details of the liquid-vapor interface and aqueous protein interfaces. Our results are presented in Section 5.3 and are organized into four topics.

We begin with discussion of potentials of mean force (PMFs) and interfacial fluctuations as single Gdm^+ /urea cross the aqueous liquid-vapor interface. We consider Gdm^+ /urea density distributions around the aqueous HFBII hydrophobic interface in 1.0 molal solutions in the second part. We further investigate the PMFs and interfacial fluctuations as single Gdm^+ /urea approach this aqueous protein hydrophobic interface, demonstrating the resemblance between liquid-vapor interface and hydrophobic protein interface in terms of solute specific effect and orientational preferences. We finish this section by examining single Gdm^+ /urea approaching another region, which is considered a hydrophilic patch compared with the hydrophobic region we initially study. We address our conclusions and general discussion in Section 5.4.

5.2 Method

All the simulations in this study were performed with MD program NAMD 2.9b3 [184, 185], using CHARMM22 all-atom force fields with CMAP backbone torsion correction term [187]. Simulations of single Gdm^+ /urea approaching the liquid-vapor interfaces were performed in the *NVT* ensemble. The simulation cell was rectangular with dimensions $40 \text{ \AA} \times 40 \text{ \AA} \times 150 \text{ \AA}$, in which z is the direction normal to the liquid-vapor interface. The system contained one single Gdm^+ /urea and 1977 nonpolarizable TIP3P water model [190] water molecules. A rigid water geometry is enforced using SHAKE [192] constraints and an integration time step of 1.0 fs was used. The temperature was kept constant at 300 K by applying the Langevin friction force scheme with a damping coefficient of 5ps^{-1} . A switching distance of 10 \AA , non-bonded real-space cutoff of 12 \AA and pairlist generation distance of 14 \AA were used for the van der Waals interactions, and the particle mesh Ewald (PME) [191] method was employed for the calculation of conditionally-convergent electrostatic interactions. The grid size of PME in x -dimension is 40, in y -dimension is 40, and in z -dimension is 150 (as close to a 1 \AA grid point separation as possible). In order to obtain the PMF for transferring single Gdm^+ /urea from bulk aqueous environment to the liquid-vapor interface, we define a collective variable, which is based on the Cartesian z -component of the separation

between the water slab center of mass and single Gdm⁺/urea central carbon, describing this pseudo-chemical reaction path. To enhance sampling of the distribution of configurations where the collective variable holds a particular value, relevant restraint potentials were introduced on the collective variable in order to prevent it from moving outside of the desired range. In this case, we constructed 31 continuous "windows" with width 1.0 Å. In each window, central carbon of single Gdm⁺/urea was restrained to z -positions from 0 Å to 30 Å relative to the water slab center of mass using a harmonic potential $U_{\text{restraint}}(z; z_{\text{relative,ref}}) = \frac{1}{2}k_{\text{restraint}}(z - z_{\text{relative,ref}})^2$ with the force constant of 4 (kcal/mol)/Å². To consider the orientational dependence of Gdm⁺ around interface, we further desired to compare the free energetics of single Gdm⁺/urea transferring from the bulk with two distinctive orientations: the planar ring of Gdm⁺/urea parallel to the liquid-vapor interface and perpendicular to the liquid-vapor interface. Here the orientations were defined based on identical definitions from previous publications [86, 87] in which the angle θ between the vector normal to the molecular ring and the z -axis was computed. Gdm⁺/urea was considered as parallel (as shown in Figure 5.1A) and perpendicular (as shown in Figure 5.1B) to the liquid-vapor interface when $\theta = 0^\circ$ and $\theta = 90^\circ$ respectively. We note that for the parallel orientation, the normal vector of the molecular ring is along z -direction; for the perpendicular orientation, the normal vector of the molecular ring can either be along x or along y direction. Due to the homogeneous nature of liquid-vapor interface and the identical setup in x and y dimensions in the simulation, here we only need to consider one (when the normal vector is along y direction) of these two configurations in the perpendicular orientation case. In these two sets of simulations, initially, the parallel and perpendicular configurations of Gdm⁺/urea were selected as starting structures respectively and the orientations were maintained by restraining the directions of central carbon-nitrogen vectors. Based on the definition of orientations above, Gdm⁺ with parallel configuration has all three central carbon-nitrogen vectors in the plane of XY , with the magnitude along z -direction being zero. Therefore, harmonic potentials with force constant $k = 1000$ (kcal/mol)/Å² were applied to keep the magnitudes of z components of two

of the three central carbon-nitrogen vectors as zero. With this restraint protocol, we can ensure the parallel orientation of single Gdm⁺ with respect to the liquid-vapor interface. For Gdm⁺ with perpendicular orientation, all three central carbon-nitrogen vectors are in the plane of XZ . To maintain this orientation, harmonic potentials with force constant $k = 1000$ (kcal/mol)/Å² were applied to restrain the magnitudes of y components of the carbon-nitrogen vectors as zero. These restraint protocols were also applied to single urea molecule by considering only the two central carbon-nitrogen vectors. Apart from the orientational restraints, identical choice of collective variable and setup of simulation windows were applied. Total sampling time for each window was 15 ns for all the simulations and properties were calculated from all but the initial 1 ns, which was treated as equilibration.

Simulations of HFBII in 1.0 molal concentration of GdmCl/urea aqueous solutions were performed in the NPT ensemble using a cubic cell with a box size 60 Å × 60 Å × 60 Å. *NPT* ensemble was used to eliminate the liquid-vapor interfaces, so only the protein-water interfaces were considered in the system. The protein structure was based on the ultra-high resolution structure of HFBII, with PDB code 2B97 and it was constructed using CHARMM-GUI [189]. Monomer of this HFBII protein, which is composed of 70 residues, was placed in the center of the box and fully solvated with 6481 water molecules, along with 116 pairs of GdmCl or 116 urea molecules. The initial structure of the protein was arranged in a way that its largest hydrophobic patch, consisting of amino acid residues Val 18, Leu 19, Leu 21, Ile 22, Val 24, Val 54, Ala 61, Leu 62 and Leu 63 (the three letter representing the amino acid types and the number representing the position of the amino acid in the primary sequence), was nearly perpendicular to the z direction. The protein was rigidly fixed at the original configuration during the simulation while other system components were unrestrained. Temperature was maintained by Langevin bath at 300K and the pressure was kept constant by Langevin pressure control at 1 atm. A switching distance of 10 Å, non-bonded real-space cutoff of 12 Å and pairlist generation distance of 14 Å were used for the van der Waals interactions. For the grid size of PME setup, the values are changed

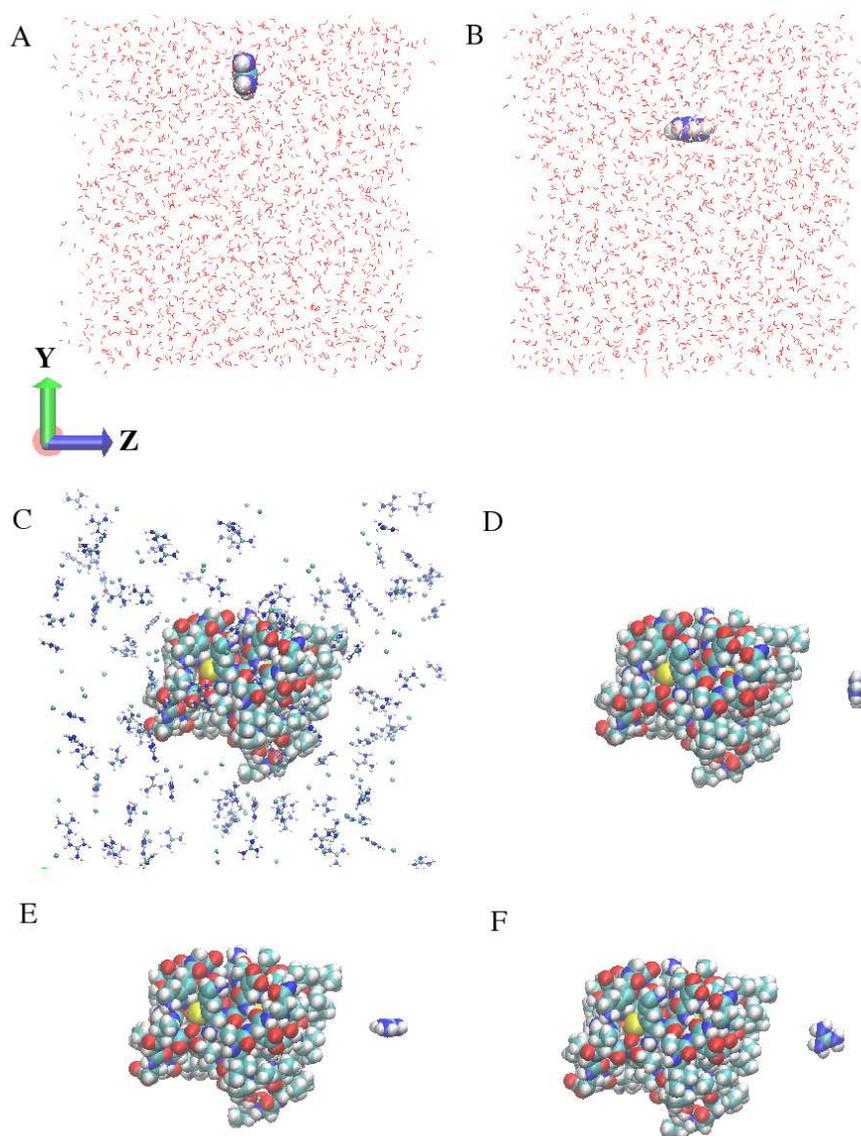


Figure 5.1: (A) Representative snapshot of single Gdm^+ with parallel orientation to the liquid-vapor interface (B) Single Gdm^+ with perpendicular orientation to the liquid-vapor interface (C) HFBII protein in 1.0 molal concentration of GdmCl aqueous solution (D) Single Gdm^+ with parallel orientation to the HFBII protein-solvent interface (E) Single Gdm^+ with perpendicular y orientation to the HFBII protein-solvent interface (F) Single Gdm^+ with perpendicular x orientation to the HFBII protein-solvent interface.

to 60 in all dimensions, corresponding to the cubic simulation cell in this case. Six different replicates were applied for each system and properties were computed based on at least 10 ns of production run for each replicate. A representative snapshot of the simulation system can be found in Figure 5.1C.

Furthermore, in order to illustrate the molecular details of orientation and free energetics of Gdm^+ /urea around protein interfaces, we simulated a system with single Gdm^+ /urea approaching the hydrophobic aqueous protein interface with different orientations. We use an identical protein starting structure as in the 1.0 molal solution case, with the hydrophobic interface of the protein nearly perpendicular to the z direction. In this way, similar to the liquid-vapor interface case, the relative orientations between single solute and protein interface can be defined in a straightforward way: when the normal vector of Gdm^+ /urea ring is along z direction, the solute is considered to be parallel to the hydrophobic protein interface as shown in Figure 5.1D; when the normal vector is along y direction (Figure 5.1E) or x direction (Figure 5.1F), the solute is considered to be perpendicular to the hydrophobic protein interface. Due to the asymmetry of hydrophobic protein interface, differences arise between these two perpendicular configurations. For the convenience of discussion, we denote the orientations in Figure 5.1E and 5.1F as perpendicular y orientation and perpendicular x orientation, indicating that the normal vector is along y direction and x direction, respectively. Here, we note that although the hydrophobic protein patch is commonly considered as flat, it still has some curvature. Hence strictly speaking, speaking of an actual parallel or perpendicular orientation of Gdm^+ plane relative to the protein patch is not rigorous. However, in this work, we aimed to study the contrasting hydration properties and surface fluctuations induced by different orientations of Gdm^+ (water-depleted flat faces versus the more strongly water-associated ring (edge-on) side of the cation) with respect to the hydrophobic patch of the protein. Therefore in this convention, parallel orientation simply indicates that Gdm^+ has more overlap with the protein patch in terms of their projections to XY plane, relative to the perpendicular orientation. The whole protein was fixed during the simulation with center of mass

located at ($x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$, $z = 0 \text{ \AA}$). A fixed Cl^- (at $x = 0 \text{ \AA}$, $y = 0 \text{ \AA}$, $z = -15 \text{ \AA}$) was added as the counter ion to neutralize the positive charge in the case of Gdm^+ . Similar to the liquid-vapor interface situation, for calculation of PMF, we use the Cartesian z component of the separation between the center of mass of protein and central carbon of single Gdm^+ /urea as the collective variable. Select configurations of single Gdm^+ /urea with parallel, perpendicular y and perpendicular x orientation were used as starting structures respectively with central carbon of the molecule located at $x = 0 \text{ \AA}$ and $y = 0 \text{ \AA}$. X component and y component of the solute's central carbon were restrained at this original position $x = 0 \text{ \AA}$ and $y = 0 \text{ \AA}$ during the simulation using NAMD's "selectConstraints" infrastructure, with sufficiently large force constant $k = 1000 \text{ (kcal/mol)/\AA}^2$. The orientations were maintained by restraining the directions of the central carbon-nitrogen vectors with the same protocol as liquid-vapor interface cases mentioned above. In this case, single Gdm^+ /urea will approach the specific spot on the patch ($x = 0 \text{ \AA}$ and $y = 0 \text{ \AA}$) with particular orientation while still keeping some rotational degree of freedom by using the normal vector to the molecular ring as rotation axis. We just centered on one specific region on the patch due to the fact that the interface is heterogeneous, resulting in the differences of the extent of inherent interface fluctuations at various locations (to be discussed further below). For a meaningful discussion of the molecule induced fluctuation (fluctuation in addition to the level inherent in pure water) as it approaches the hydrophobic interface, one representative spot with fixed position and unchanged inherent fluctuation had to be defined. In this case, along the positive z -direction, 49 continuous "windows" with width 0.2 \AA ranging from area around protein-solvent interfaces to bulk water region were constructed. The spans of the windows going from interfacial region to bulk region (in \AA) were: [15.4:15.6], [15.6:15.8], [15.8:16.0] [24.4:24.6], [24.6:24.8], [24.8:25.0]. In each window, a harmonic restraint potential with force constant of $10 \text{ (kcal/mol)/\AA}^2$ was applied. Other simulation conditions remain the same as that of the system of protein in 1 m concentration of GdmCl /urea aqueous solution. The first 2ns was allowed for equilibration before a total of 20ns production data was generated for each window.

For a complete understanding of the influence of the hydrophobicity of the protein patch on the orientational preference of Gdm⁺ solute, we considered another system in which single Gdm⁺ approaches a more hydrophilic protein region which consists of residues Asp 25, Cys 26, Lys 27, Thr 28, Ala 58, Asp 59, Gln 60. The simulation conditions remained identical except that the protein was posed in a different way in the simulation cell with the selected hydrophilic interface almost perpendicular to the z direction. The window setup ranged from [14.0:14.2], [14.2:14.4], [14.4:14.6] to [24.4:24.6], [24.6:24.8], [24.8:25.0] for a total of 56 windows.

5.3 Result and Discussion

5.3.1 Liquid-Vapor Interface

We first consider a single Gdm⁺/urea solute approaching the liquid-vapor interface. This analysis provides a reference context within which to discuss the results at a hydrophobic protein interface later. To first address solute orientational propensities as they vary along the order parameter, we compute orientationally resolved probability distribution profiles along the z-axis as a single Gdm⁺/urea approaches the liquid-vapor interface as shown in Figure 5.2A and C respectively. Here, in a statistical manner, we consider the probability of the single solute at position z with orientation θ , which is defined based on the following Equation 5.1[86]:

$$P(z, \theta) = \frac{\int_{z-\frac{\Delta z}{2}}^{z+\frac{\Delta z}{2}} dz \int_{\cos(\theta-\frac{\Delta\theta}{2})}^{\cos(\theta+\frac{\Delta\theta}{2})} d\cos\theta n(z, \theta)}{\int_{z-\frac{\Delta z}{2}}^{z+\frac{\Delta z}{2}} dz \int_0^1 d\cos\theta n(z, \theta)} \quad (5.1)$$

where $n(z, \theta)$ denotes the solute number count at position z with orientation θ . The numerator represents the number of solutes in a slab from $z - \frac{\Delta z}{2}$ to $z + \frac{\Delta z}{2}$, with select orientation within the range of $\cos(\theta - \frac{\Delta\theta}{2})$ and $\cos(\theta + \frac{\Delta\theta}{2})$. The denominator represents the total number of solutes in the slab region $z - \frac{\Delta z}{2}$ to $z + \frac{\Delta z}{2}$; this is used to normalize the probabilities in the relevant slab whose boundaries along the order parameter are $z - \frac{\Delta z}{2}$ to $z + \frac{\Delta z}{2}$. The limits $\cos(\theta) = 1$ and $\cos(\theta) = 0$ represent Gdm⁺ orientations that are parallel and perpendicular to the liquid-vapor interface, respectively. In bulk region

with $z < 13 \text{ \AA}$, the probabilities of Gdm^+ /urea with different orientations are identical, indicating no orientational preference; while in the interfacial region ($15 \text{ \AA} \leq z \leq 20 \text{ \AA}$), single Gdm^+ manifests a higher tendency to adopt the configuration that is parallel to the liquid-vapor interface. This observation is consistent with the result in our previous publication on an identical system using a polarizable force field (TIP4P-FQ), [87] and Wernersson et al.’s work using 1.1 m and 5.3 m GdmCl solutions. [86] Single urea also displays a marginal orientational preference for a parallel configuration as well; urea’s propensity for the parallel orientation is lower than that of Gdm^+ based on the lower intensity of the corresponding region in Figure 5.2C. We note that this higher probability of parallel orientation of single solute around interfacial region suggests a lower free energy of this configuration relative to the perpendicular. To further explore this difference, we consider potentials of mean force for single Gdm^+ /urea from bulk through liquid-vapor interface being restrained at particular orientations as shown in Figure 5.2B and D for Gdm^+ and urea respectively. In both panels, black lines represent conditional PMF profiles for single solute with parallel orientation; blue dashed lines represent conditional PMF profiles for single solute with perpendicular orientation. The orientation-averaged PMF profile (with no restraints on the orientation) is shown as a dotted green line. The weighted histogram analysis method (WHAM) was used for generating the final PMF in all cases. [145] The standard error was estimated by using the block averaging method obtained from each consecutive 0.5 ns time block in the production run of each umbrella sampling window. This selection will ensure the block size was significantly larger than the correlation time in each window. The PMF is defined to be zero when the solute is in the bulk, which is determined by window $z = 0 \text{ \AA}$. To better compare the interface stability among different orientations, in the large graph of Panel B and D, we emphasized the PMFs around the interfacial region while the entire PMF along the collective variable can be found in the inset. For single Gdm^+ , the parallel orientation shows a minimum of roughly -0.4 kcal/mol, with uncertainty about 0.1 kcal/mol, prior to the GDS at the separation of $z = 16.5 \text{ \AA}$, while the perpendicular orientation displays no surface stability at all. Overall, when

there is no orientational restraint applied on the Gdm^+ , as shown by the green line, no surface stability is found near the interface. The PMF is less repulsive in the case with no orientational restraint compared to the perpendicular orientation scenario; this is consistent considering that it is an average result from the contributions of all possible configurations. The PMF for no orientational restraints (green dotted line) shows a slight shoulder around $z = 16 \text{ \AA}$, indicating the effect of the parallel orientations. However, since the stability of the parallel orientation is rather small, and configurations differing from the parallel geometry are associated with significantly higher free energies at the interface, the overall effect leads to a PMF displaying no apparent interfacially-stable state. For single urea, the parallel orientation PMF shows a slight minimum of $0.04 \pm 0.07 \text{ kcal/mol}$. Considering the uncertainty here, whether parallel orientation of urea shows surface stability or not is debatable. However, we notice that in contrast to the perpendicular orientation, the parallel orientation is more free energetically favorable, although this trend is not as obvious as the case for Gdm^+ . All these PMF results are consistent with the probability distributions of orientations as discussed above. We note that this orientational preference of Gdm^+ around liquid-vapor interface may be related to the hydration structure of Gdm^+ as previously studied by Mason et al. [85, 73] and Cooper et al [233]. The hydration around Gdm^+ is anisotropic. In the molecular plane, the N-H group can serve as hydrogen bond donor, interacting with water molecules [233] as demonstrated in the gas-phase, while above or below the planar face, it is inadequate to serve either as hydrogen bond donor or acceptor. Therefore, when single Gdm^+ approaches the liquid-vapor interface with parallel orientation, it is easy for desolvation to occur, which is free energetically favorable. For the structurally analogous molecule urea, it still can serve as hydrogen bond donor above or below the planar face, so it is less facile for the parallel urea molecule to desolvate compared with that of Gdm^+ [234].

Recent studies have demonstrated an interesting connection between liquid-vapor interfacial stability of chemical species and the extent to which the presence of these molecular species in the vicinity of the interface induces collective fluctuations

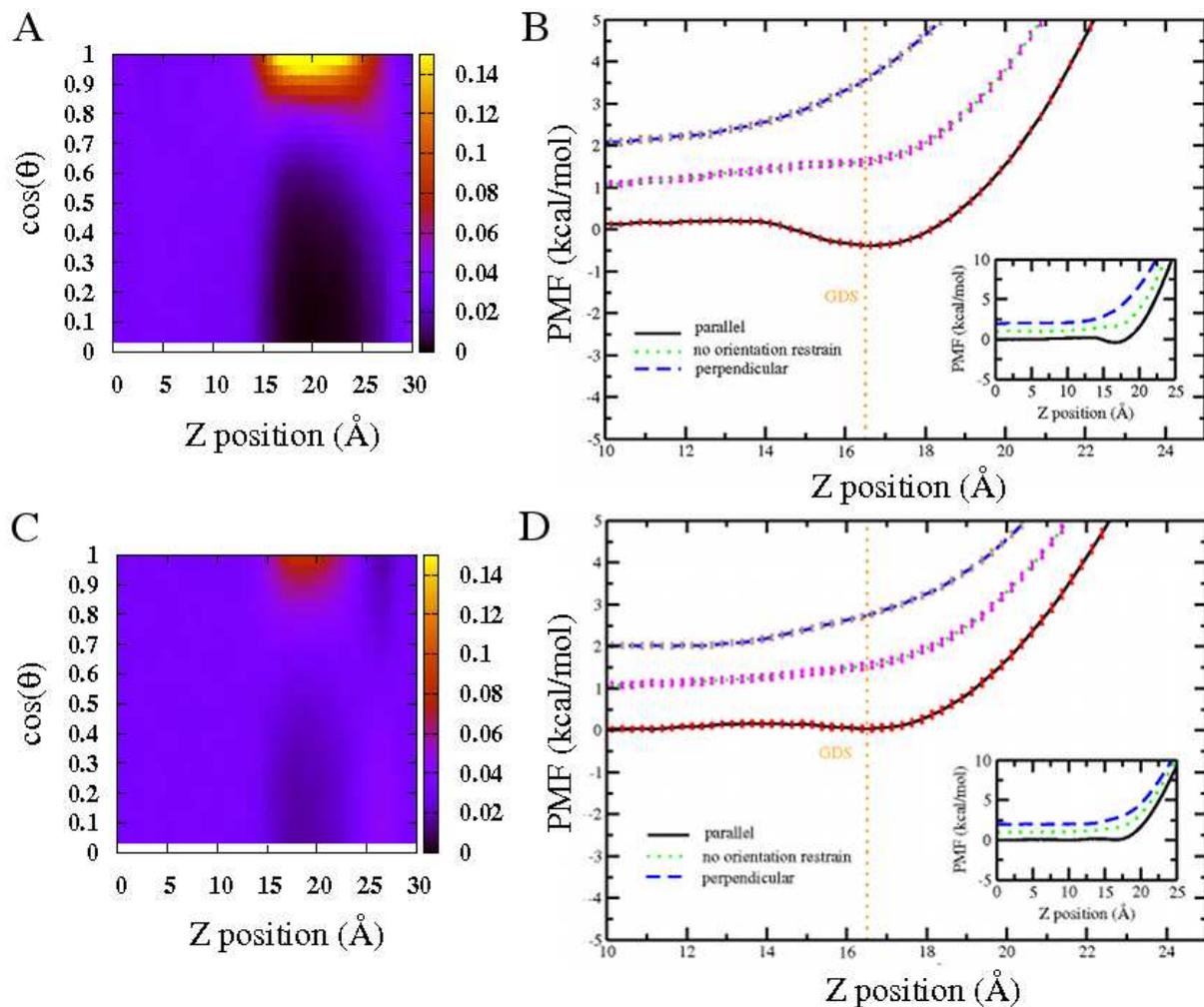


Figure 5.2: (A) Orientationally resolved probability map of single Gdm⁺ around liquid-vapor interface (B) PMF of single Gdm⁺ from bulk transporting through liquid-vapor interface with parallel orientation, perpendicular orientation and no orientational restraint (C) Orientationally resolved probability map of single urea around liquid-vapor interface (D) PMF of single urea from bulk transporting through liquid-vapor interface with parallel orientation, perpendicular orientation and no orientational restraint. For clarity, in (B) and (D), no orientational restraint profiles are shifted by 1 kcal/mol; perpendicular profiles are shifted by 2 kcal/mol. The GDS positions are denoted as orange dash lines in (B) and (D).

of the interface in addition to the level inherent in pure water (absence of the solute) due to thermal motion; these studies have focused on monovalent inorganic ions as initial test systems. [56, 57, 87] Initially, it is found that the species demonstrating an interfacial stability (eg, I^-) as demonstrated by free energy minima in the region of the GDS as evaluated via potentials of mean force, appear to enhance liquid-vapor interfacial fluctuations, while those that show no interfacial stability (eg, Cl^-) induce no or lesser extent of fluctuations. The differences in induced interfacial fluctuations by two representative ions, Cl^- and I^- , has been attributed to these two types of ions presenting distinct hydration shell environments. The first solvation shell of I^- is more malleable than that of Cl^- . The nature of the solvent structure around I^- determines that it is more amenable to inducing fluctuations of the interface as a consequence of a greater disruption of the solvent structure on approach to the interface. Inspired by this, we consider that differences in induced interfacial fluctuations should arise as the parallel and perpendicular orientations of Gdm^+ approach the liquid-vapor interface since these two configurations display distinct hydration shell environments with parallel orientation presenting a more malleable solvent environment and perpendicular orientation showing a more rigid hydration environment due to the more effective hydrogen bonding of water in the plane of the ring. The two orientations are associated with dramatically different free energetic profiles at the liquid-vapor interface. The surface height fluctuations were then computed. The mean surface height and surface height fluctuation when single Gdm^+ resides at the position of $z = 14 \text{ \AA}$ is shown in Figure 5.10. Both the mean surface profile and surface height fluctuation profile are radially symmetric, with the largest value at the position where the Gdm^+ is just approaching the point $(x = 0, y = 0)$. For convenience, we use this representative value $\langle \delta h^2(x = 0, y = 0) \rangle$ to compare the magnitude of interface fluctuations when the solute is restrained at different z -positions, with the result shown in Figure 5.3. Fluctuation profiles for Gdm^+ and urea with distinct orientations are presented in Panel A and C respectively. In the case of pure TIP3P water without the existence of solute, the inherent fluctuation for current system size is around 1.32 \AA^2 . Using this value as a

normalization factor, normalized fluctuation profiles were obtained, presented in Panel B and D for Gdm⁺ and urea respectively, which will display solute-induced contribution in a direct way. $\langle \delta h_L^2 \rangle > 1$ indicates the surface height fluctuation is enhanced relative to pure water; with $\langle \delta h_L^2 \rangle < 1$, it denoted that the surface height fluctuation is damped. In the Gdm⁺ case, the parallel orientation induces a large fluctuation, with the maximum normalized fluctuation value around 2.2 at the location of $z = 14$ Å, which is around 3 Å prior to the position of the free energy minimum. In stark contrast, no obvious enhancement of surface fluctuations is associated with perpendicular orientation. These trends are expected considering the distinct hydration structures for the parallel and perpendicular orientation of Gdm⁺. The nature of the malleable hydration shell around parallel orientation Gdm⁺ is similar to that of single, low charge density anions, like I⁻. The solvent structure is more amenable to inducing larger fluctuations due to the fact that it is more easily disrupted as the solute approaches the interface. On the other hand, the perpendicular orientation of Gdm⁺ showing no interface stability is more like Cl⁻, with more rigid solvation structure around the periphery of the ring due to the existence of hydrogen bonding. These results using a nonpolarizable force field are consistent with our previous work using the TIP4P-FQ polarizable force field, [87] indicating a force field independence of the fundamental, underlying physical origin of this correlative phenomenon. Previously, it has been pointed out the importance of considering polarizability in the ion-specific effect. [45, 47] By neglecting the polarizability in the force field, larger anion I⁻ may not show significant surface stability, giving a poor distinction with respect to Cl⁻. However, our results indicate that in case of Gdm⁺, the orientational preference is pronounced enough even in the nonpolarizable force field. In light of this, in the following section for the discussion of Gdm⁺ around the protein surface, we use this nonpolarizable force field. We also notice that differences in induced fluctuations also exist in the case of urea with dissimilar orientations. For the parallel orientation, the largest induced normalized fluctuation value is around 1.85, which is still larger than the fluctuation from perpendicular orientation 1.45. Interestingly, for the parallel orientation, the induced fluctuations from

Gdm⁺ is larger than that from urea, corresponding to Gdm⁺'s greater free energetic stabilization; the perpendicular orientations show a reverse trend as the induced fluctuation from Gdm⁺ is smaller than that from urea, which correlates with the PMF trend that Gdm⁺ is more repulsive in this case. Again, the smaller differences in interfacial stabilization and induced fluctuation between parallel and perpendicular orientation is related to the spatial location of hydrogen bonding network, either below or above the planar face, leading to the closer solvation structure of the parallel oriented and perpendicular oriented urea. Overall, the differences in orientational preference around liquid-vapor interface, interfacial stability and induced fluctuation between Gdm⁺ and urea may possibly be connected to the efficiency of these two solutes as denaturants via direct interactions with hydrophobic side chains and surface regions of proteins. For a further understanding of this, we attempt to extend this investigation from the ideally hydrophobic aqueous liquid-vapor interface to a somewhat more realistic and more complex aqueous protein hydrophobic interface.

5.3.2 Aqueous Protein Interface

Before we consider the free energetics of single Gdm⁺/urea approaching the hydrophobic protein interfacial region, we provide a general overview of distributions of solute orientation relative to the hydrophobic protein patch. To probe this, generally, we define a sampling volume in the Cartesian space corresponding to the hydrophobic patch around the protein. This sampling volume is shown in Figure 5.4A, within the range of $-8 \text{ \AA} \leq x \leq 8 \text{ \AA}$, $-8 \text{ \AA} \leq y \leq 8 \text{ \AA}$ and $12 \text{ \AA} \leq z \leq 25 \text{ \AA}$, roughly including residues Val 18, Leu 19, Leu 21, Ile 22, Val 24, Val 54, Ala 61, Leu 62 and Leu 63. Orientationally-resolved probability distribution of Gdm⁺ around this defined region is shown in Figure 5.4B. The probability at position z with orientation θ in this case is defined in the same way as that in liquid-vapor interface system in Equation 5.1. Around the selected hydrophobic protein surface, Gdm⁺ displays a higher propensity for the parallel configuration. We note that this marked tendency has previously been noticed in proximity to hydrophobic surface. England et al. [75, 88] found that Gdm⁺

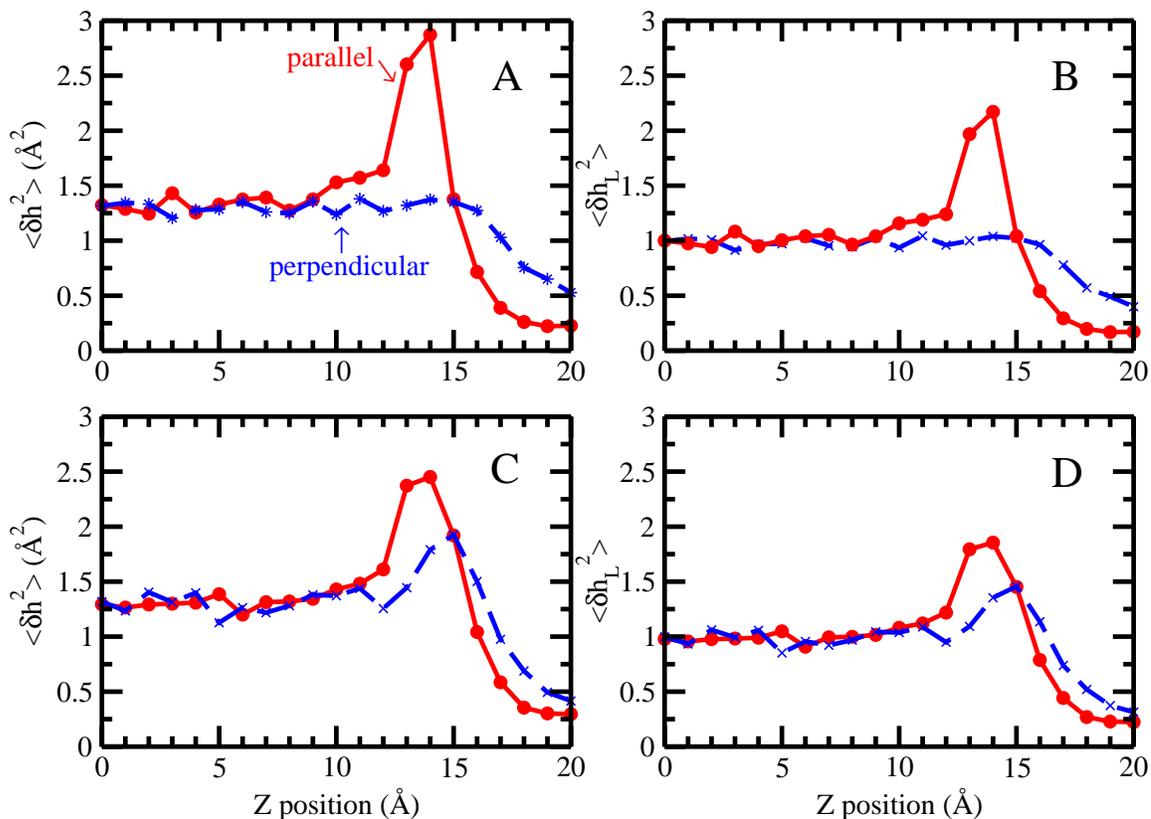


Figure 5.3: (A) Surface height fluctuation for liquid-vapor interface at ($x = 0, y = 0$) as a function of position of single Gdm^+ (B) Normalized surface height fluctuation for liquid-vapor interface at ($x = 0, y = 0$) as a function of position of single Gdm^+ (C) Surface height fluctuation for liquid-vapor interface at ($x = 0, y = 0$) as a function of position of single urea (D) Normalized surface height fluctuation for liquid-vapor interface at ($x = 0, y = 0$) as a function of position of single urea.

accumulates in the vicinity of flat hydrophobic plate in a roughly parallel way. Godawat et al. [77] also mentioned the Gdm^+ has a preference for parallel stacking with the hydrophobic polymer surface from the snapshots of their simulations. Here, using a simple approach we defined the relative orientations between Gdm^+ solute and hydrophobic protein surface and showed the orientational preference of Gdm^+ around the hydrophobic patch of HFBII. Furthermore, we consider orientationally-resolved probability distributions of urea in an identical probe volume in Figure 5.4C. Parallel oriented urea is also preferred compared with the perpendicular configuration around the hydrophobic protein surface. However, this trend is less intense in the case of urea compared with Gdm^+ , which is similar to the situation at the liquid-vapor interface. This is consistent with the previous report that urea, compared with Gdm^+ , displays more orientational diversity around hydrophobic plate-like surfaces[75]. So far, we only concentrated on solute distributions around the hydrophobic region of the protein and attempted to connect this with the similar observation around the liquid-vapor interface, which is one model of an ideal hydrophobic interface. A complementary study would be focusing on another distinct region around the protein surface with different hydrophobicity. Therefore, we define another sampling volume corresponding to the hydrophilic patch of the protein (including residues Asp 25, Cys 26, Lys 27, Thr 28, Ala 58, Asp 59, Gln 60) in Cartesian space within the range of $-8 \text{ \AA} \leq y \leq 8 \text{ \AA}$, $-8 \text{ \AA} \leq z \leq 8 \text{ \AA}$ and $12 \text{ \AA} \leq x \leq 25 \text{ \AA}$. The volume of this sampling region remains the same as that defined for the hydrophobic region, but the position of the probe region in Cartesian space is different. In Figure 5.4D, is shown the orientationally-resolved probability distribution for Gdm^+ around this hydrophilic region. Overall, no preference for parallel oriented Gdm^+ is observed in this case, although this preference can be detected in a small portion of the map with the separation of $z = 14 \text{ \AA}$. Compared to the orientationally-resolved probability maps for Gdm^+ around hydrophobic (Panel B) and hydrophilic area (Panel D), it is safe to claim that around more hydrophobic regions, there is a stronger tendency for the parallel stacking of Gdm^+ with protein surface, which is also been noticed in flat plate systems previously. [74] We note that

these general trends are robust as an identical patch with different sampling volume, as defined in the Figure 5.5 and Figure 5.6, further verifies this orientational preference around hydrophobic/hydrophilic regions.

The PMF profiles for single Gdm^+ /urea approaching the hydrophobic protein surface region are shown in Figure 5.7. Panel A shows the PMF of single Gdm^+ with parallel orientation (solid red line), perpendicular y orientation (dotted green line), and perpendicular x orientation (dashed blue line) moving towards the hydrophobic patch region from the bulk, which is located at $z = 25 \text{ \AA}$ in this case. The PMF profiles were generated by post-processing umbrella sampling MD trajectories with WHAM and the standard errors were estimated by using the block averaging method obtained from each consecutive 0.5 ns time block in the production run of each umbrella sampling window. The parallel configuration gives rise to a PMF minimum of $-2.85 \pm 0.04 \text{ kcal/mol}$ as it nears the hydrophobic patch at a separation of $z = 15.7 \text{ \AA}$; a shallow second minimum can be observed at a separation of $z = 19 \text{ \AA}$. A free energetic barrier can be observed between these two minima, which may be related to the dramatic change of the number of water molecules within the first hydration shell of the solute as shown in Figure 5.8. There is a shallow minimum with free energy $-0.48 \pm 0.04 \text{ kcal/mol}$ around $z = 17 \text{ \AA}$ for the perpendicular y orientation Gdm^+ ; while for perpendicular x Gdm^+ , a monotonically repulsive trend was observed. This difference may be determined by the exact composition and local spatial arrangement of the residues on and near the hydrophobic patch as shown in Figure 5.13A and 5.13B. When Gdm^+ with perpendicular x orientation approaches the patch, there is a repulsive interaction between NH groups of Gdm^+ and side chains of residues Ile 22 and Leu 63 on the patch. More importantly, we notice that compared with PMFs for perpendicular oriented Gdm^+ showing marginal or no stability, PMF's for parallel orientation Gdm^+ are much more free energetically favorable. This further echoes the result shown in Figure 5.4B, indicating that Gdm^+ prefers to associate with the hydrophobic protein patch with its more hydrophobic, easily desolvated, parallel orientation. This preference is explained by England et al. [75, 88] as hydrophobicity-driven stacking interaction in their study

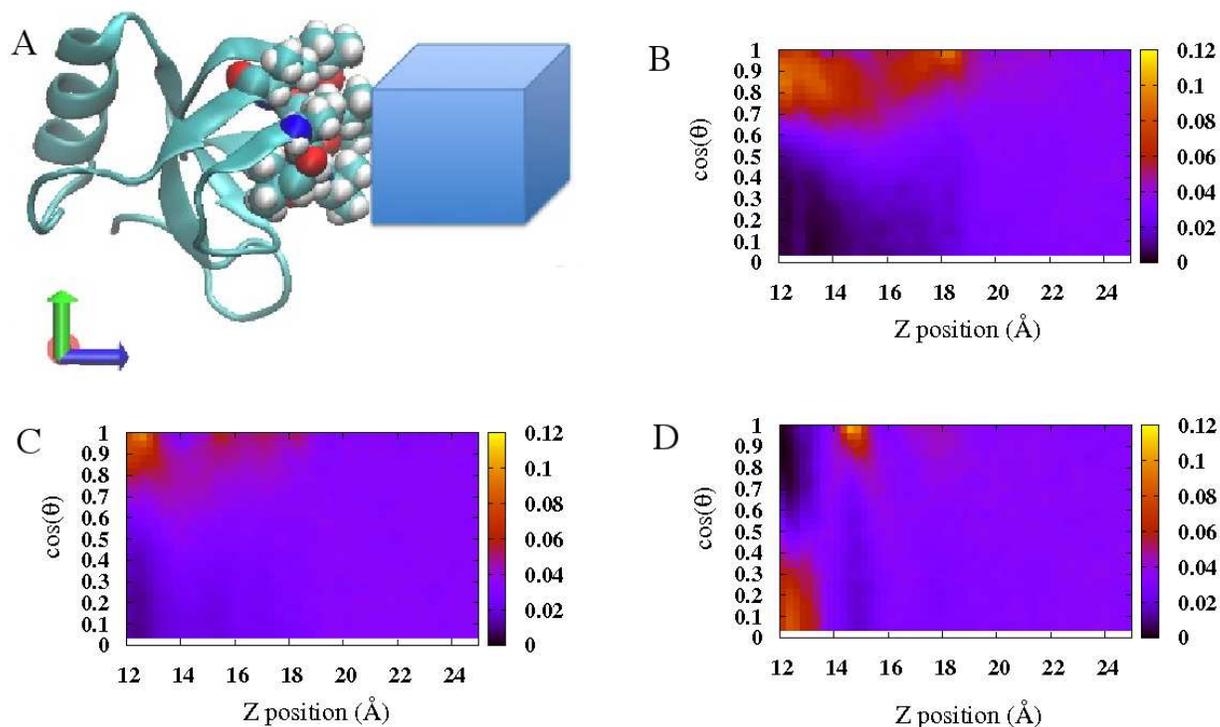


Figure 5.4: (A) Representative of sampling volume for probing orientational resolved probability of solute around certain region of protein interface (B) Orientational resolved probability distribution of Gdm⁺ around hydrophobic protein interface in 1.0 molal GdmCl solution (C) Orientational resolved probability distribution of urea around hydrophobic protein interface in 1.0 molal urea solution (D) Orientational resolved probability distribution of Gdm⁺ around hydrophilic protein interface in 1.0 molal GdmCl solution.

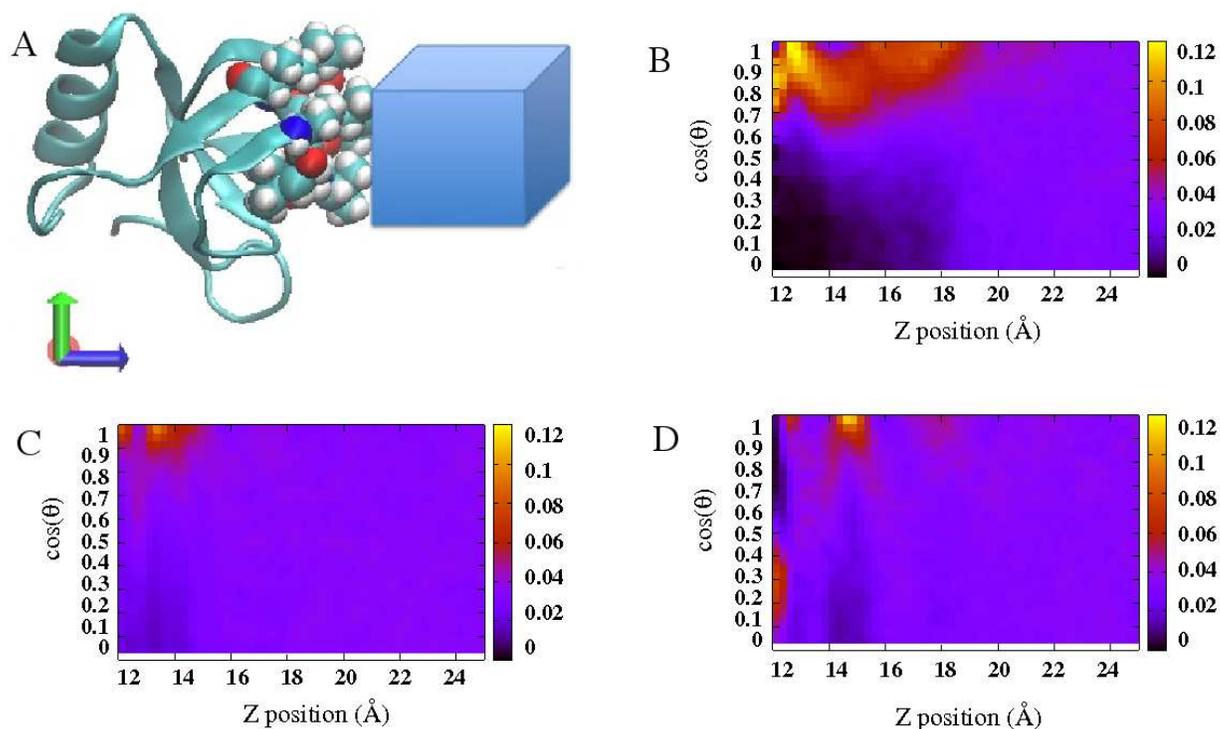


Figure 5.5: (A) Representative of sampling volume for probing orientational resolved probability of solute around certain region of protein interface with the range of $-6 \text{ \AA} \leq x \leq 6 \text{ \AA}$, $-6 \text{ \AA} \leq y \leq 6 \text{ \AA}$ and $12 \text{ \AA} \leq z \leq 25 \text{ \AA}$ (B) Orientational resolved probability distribution of Gdm⁺ around hydrophobic protein interface in 1.0 molal GdmCl solution (C) Orientational resolved probability distribution of urea around hydrophobic protein interface in 1.0 molal urea solution (D) Orientational resolved probability distribution of Gdm⁺ around hydrophilic protein interface in 1.0 molal GdmCl solution.

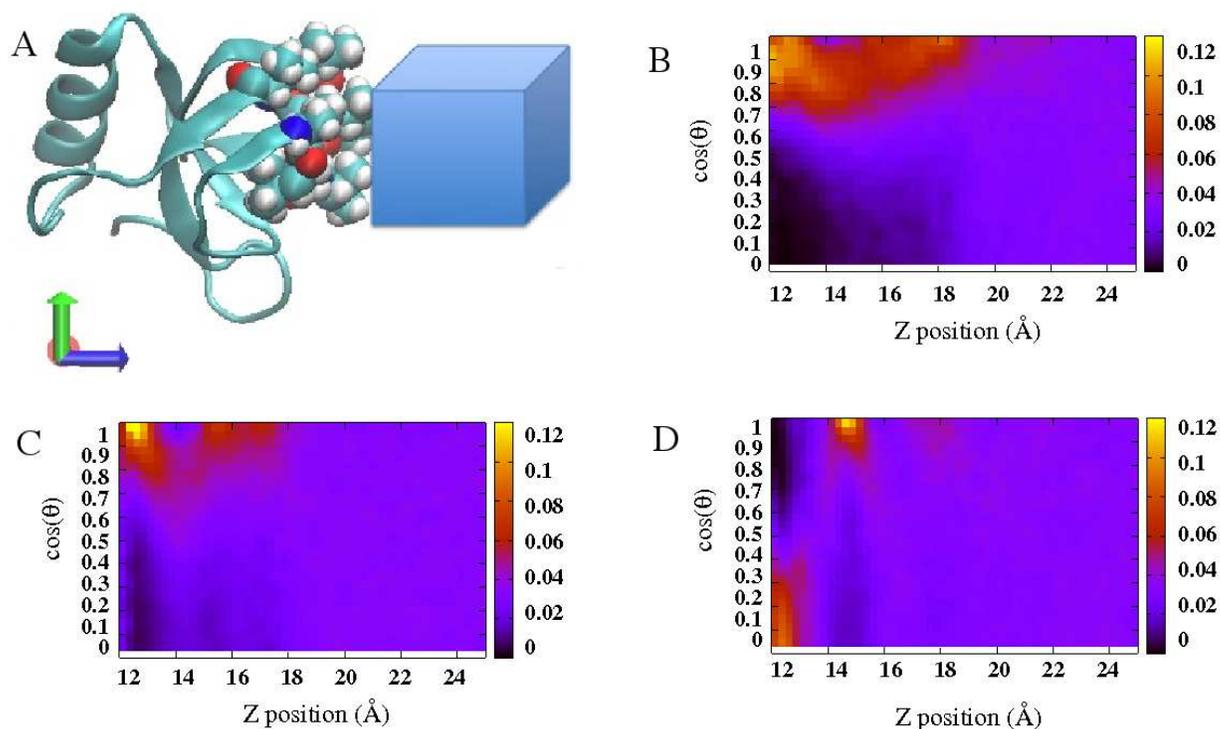


Figure 5.6: (A) Representative of sampling volume for probing orientational resolved probability of solute around certain region of protein interface with the range of $-7 \text{ \AA} \leq x \leq 7 \text{ \AA}$, $-7 \text{ \AA} \leq y \leq 7 \text{ \AA}$ and $12 \text{ \AA} \leq z \leq 25 \text{ \AA}$ (B) Orientational resolved probability distribution of Gdm⁺ around hydrophobic protein interface in 1.0 molal GdmCl solution (C) Orientational resolved probability distribution of urea around hydrophobic protein interface in 1.0 molal urea solution (D) Orientational resolved probability distribution of Gdm⁺ around hydrophilic protein interface in 1.0 molal GdmCl solution.

using hydrophobic plate. Due to the inability of hydrogen bond formation between hydrophobic surface and water molecules, the hydrophobic surface has a stronger tendency to minimize the exposed area in the aqueous environment. This can be achieved by the face-on coating by Gdm^+ of the surface, which is free energetically favorable. The observed stacking mode of self-association among Gdm^+ [73, 78, 235] can also be considered as hydrophobically driven interaction. Instead of association with large hydrophobic plate or protein surface, in this case, Gdm^+ pairs with another Gdm^+ by maximize the overlapping of their hydrophobic planar rings. We further verify this by considering the PMFs of single Gdm^+ approaching another Gdm^+ with different relative orientations as shown in Figure 5.9. Besides, In Figure 5.7B, PMF profiles for urea with different orientations moving towards the identical hydrophobic patch are presented. Again, parallel oriented urea molecule shows the most free energy stability with a value around -2 kcal/mol compared with the two perpendicular orientations. However, comparing parallel urea association free energy around hydrophobic protein patch with that of parallel Gdm^+ , we find that it is less favorable, which is due to the lower hydrophobicity of urea's planar surface as discussed earlier. These behaviors are consistent with the results from orientation distribution maps in Figure 5.4.

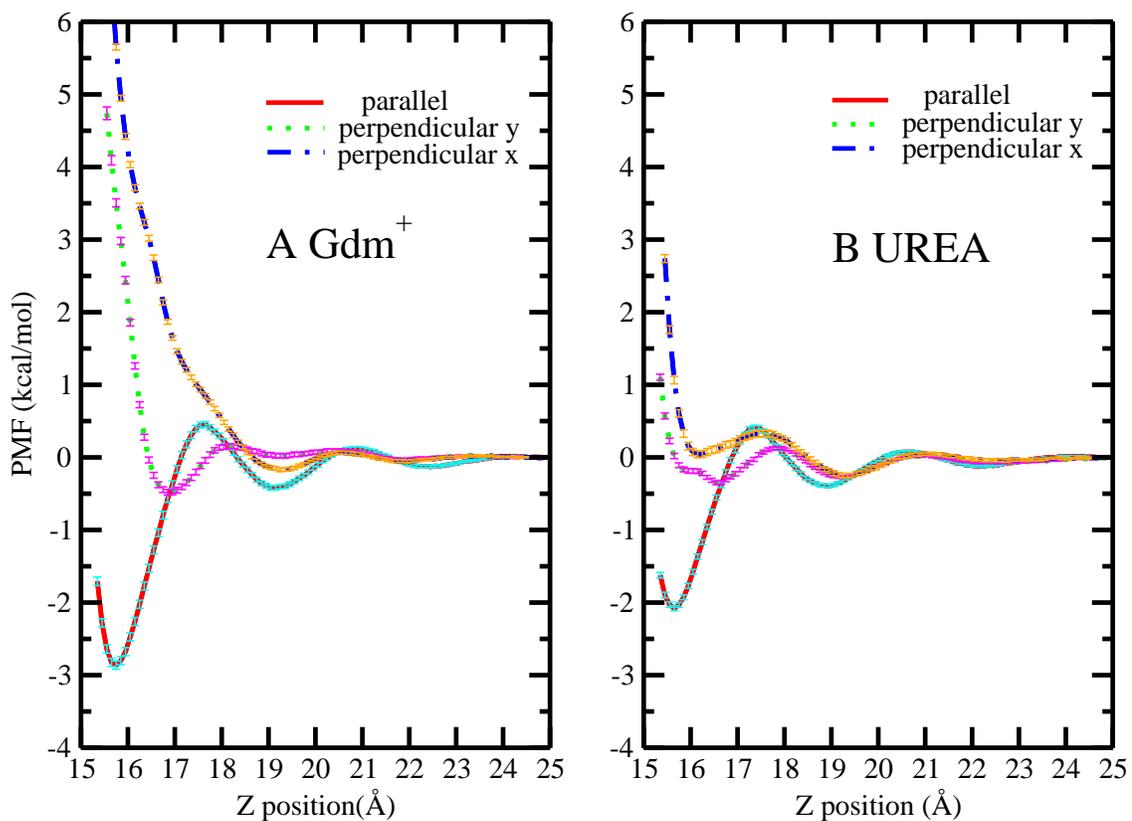


Figure 5.7: (A) PMF for single Gdm^+ with parallel orientation, perpendicular y and perpendicular x orientation from bulk approaching the hydrophobic protein-solvent interface (B) PMF for single urea with parallel orientation, perpendicular y and perpendicular x orientation from bulk approaching the hydrophobic protein-solvent interface.

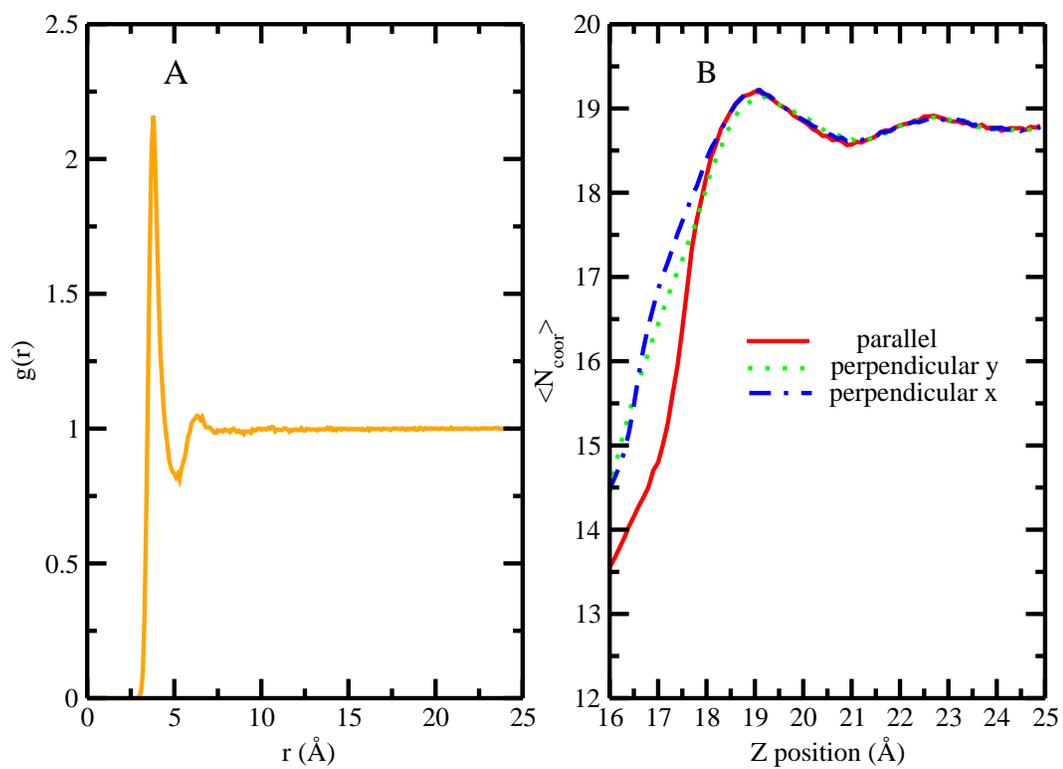


Figure 5.8: (A) RDFs between central carbon of Gdm^+ and oxygen atoms of water molecules (B) Coordinate water numbers within the first hydration shell of Gdm^+ restraining at different orientations as a function of Z position of central carbon of Gdm^+ .

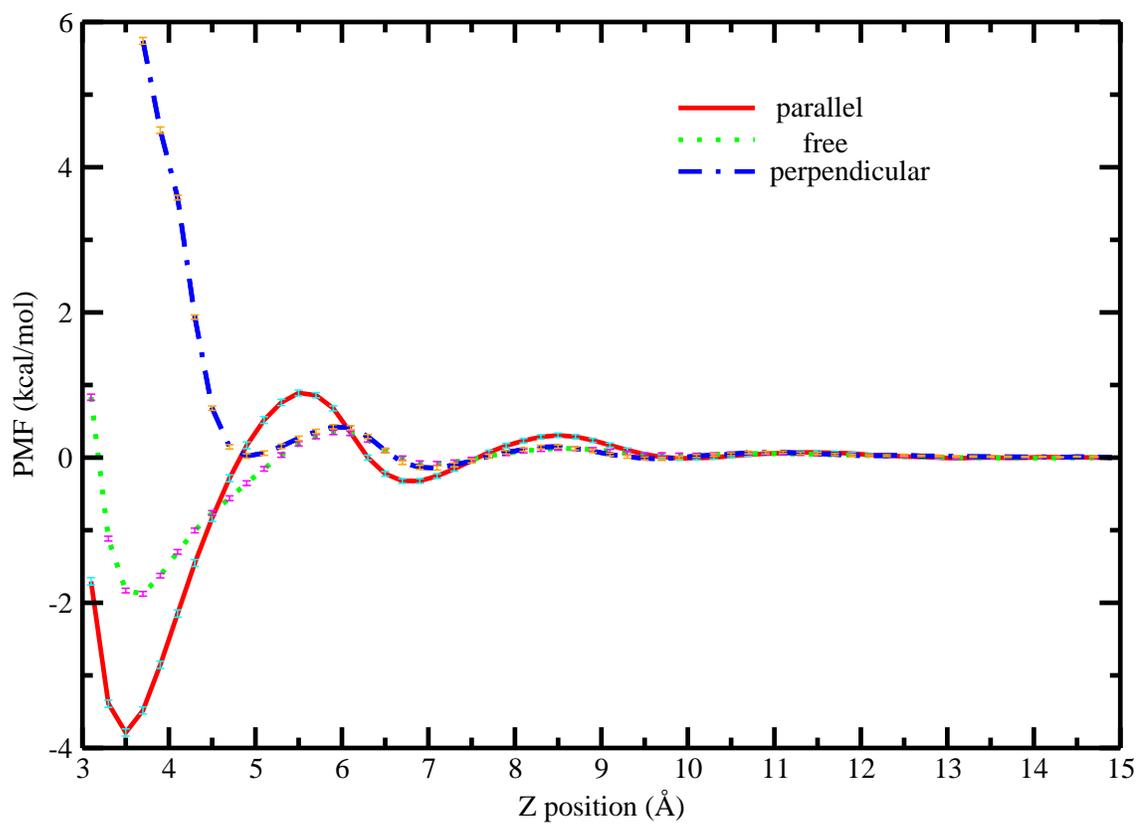


Figure 5.9: PMF profiles for Gdm^+ approaching another Gdm^+ with parallel orientation (red), perpendicular orientation (blue) and no orientational restraint (green).

The stability of a single solute at the protein interface correlates with the induced interfacial fluctuations as the solute approaches. The protein-solvent interface was constructed based on our defined protocol. A representative average protein-solvent interface is shown in Figure 5.10C and the corresponding height fluctuation is shown in Figure 5.10D in the Supporting Information. Overall, the contour of the mean interface is reasonable considering that the shape of the protein is globular. The magnitude of interface fluctuations can be judged by the color scale in Panel D, in which case a single Gdm⁺ is located at $z = 18 \text{ \AA}$ right above the position $x = 0, y = 0$. We consider this point as a reference point since it displays the largest induced fluctuations compared with other regions on the protein surface as indicated by the bright ring in Figure 5.10D. To better illustrate the change in interface fluctuation magnitude as single Gdm⁺/urea approaches the patch, the induced fluctuation at this reference point, $\langle \delta h^2(x = 0, y = 0) \rangle$, as a function of z position of the central carbon of the solute is plotted in Figure 5.11. In Figure 5.11A, at large separations of the single Gdm⁺ from the hydrophobic patch region, none of the configurations of Gdm⁺ shows increased perturbation of the aqueous protein interface. The magnitude of undulations of the protein-solvent interface solely comes from the inherent, thermal fluctuation. As the restrained Gdm⁺ with distinct orientations approaches the hydrophobic patch, the induced fluctuation profiles exhibit striking differences. Parallel-oriented Gdm⁺ induces large fluctuations of the interface (0.95 \AA^2) at the separation of $z = 18 \text{ \AA}$, which is around 5 times that of the inherent interfacial fluctuation (0.19 \AA^2). In the case of perpendicular-oriented Gdm⁺, maxima in the fluctuation profiles can also be found at the same separation of $z = 18 \text{ \AA}$. However, the extent of the induced interfacial fluctuation is smaller compared with that of the parallel orientation, with perpendicular y giving a value of 0.4 \AA^2 (2 times of inherent fluctuation) and perpendicular x giving a value of 0.35 \AA^2 (1.8 times of inherent fluctuation). This is consistent with the trends at the liquid-vapor interface showing that interfacially stable parallel configurations of Gdm⁺ induce larger interfacial fluctuations than the perpendicular, less interface-stable orientations of Gdm⁺. As expected, parallel orientations urea induces a larger

extent of fluctuation (around 0.75 \AA^2) than the perpendicular ones (around 0.45 \AA^2), corresponding to the greater free energy stability of the parallel configurations around the interface shown in Figure 5.7. Comparing the induced fluctuation values between parallel configurations of the two solutes, the more hydrophobic and more surface stable Gdm^+ gives a higher level of enhanced fluctuation. These results support the argument that the more hydrophobic nature of the parallel-oriented Gdm^+ makes the hydration shell weakly-bound and less-ordered, so that it has more tendency to break and couple with the hydration water in the vicinity of hydrophobic protein patch region, which will cause a large perturbation of the protein-solvent interface in addition to the level present in pure water. According to the previous studies [56, 55, 57, 169], this enhanced fluctuation represents an increase of interface entropy, which may contribute to differentially stabilizing configurations where the parallel orientation Gdm^+ is closer to the interface compared to other configurations of the solute.

To close this discussion about induced interfacial fluctuations, we address potential artifacts in our algorithm for computing interfacial fluctuations. One may ask whether the instantaneous coarse-grained interface we construct can artificially pass "through" the solute, thus giving rise to artificially large fluctuations. To explore this, we plot the difference Δ in the z -position of the central carbon of Gdm^+ and the z -position of the interface ($z_{\text{interface}}$) as Gdm^+ moves toward the protein patch along the z -axis. Here, the z -position of the interface is equal to the value of the surface height of the interface at the point ($x = 0, y = 0, z_{\text{interface}}$). $\Delta > 0$ means the interface is between single Gdm^+ and the protein patch; while $\Delta = 0$ indicates that interface just passes through the Gdm^+ ; $\Delta < 0$ implies that the Gdm^+ resides between the interface and the protein. We will show that even when all the $z_{\text{interface}}$ values are distributed on one side of the solute with Δ value constantly being larger or smaller than zero, the distribution of $z_{\text{interface}}$ is not necessarily small. This would suggest that the induced fluctuations are non-artifactual and the higher fluctuation values are not due to the combination of three different scenarios ($\Delta > 0, = 0, < 0$). Figure 5.14 displays the

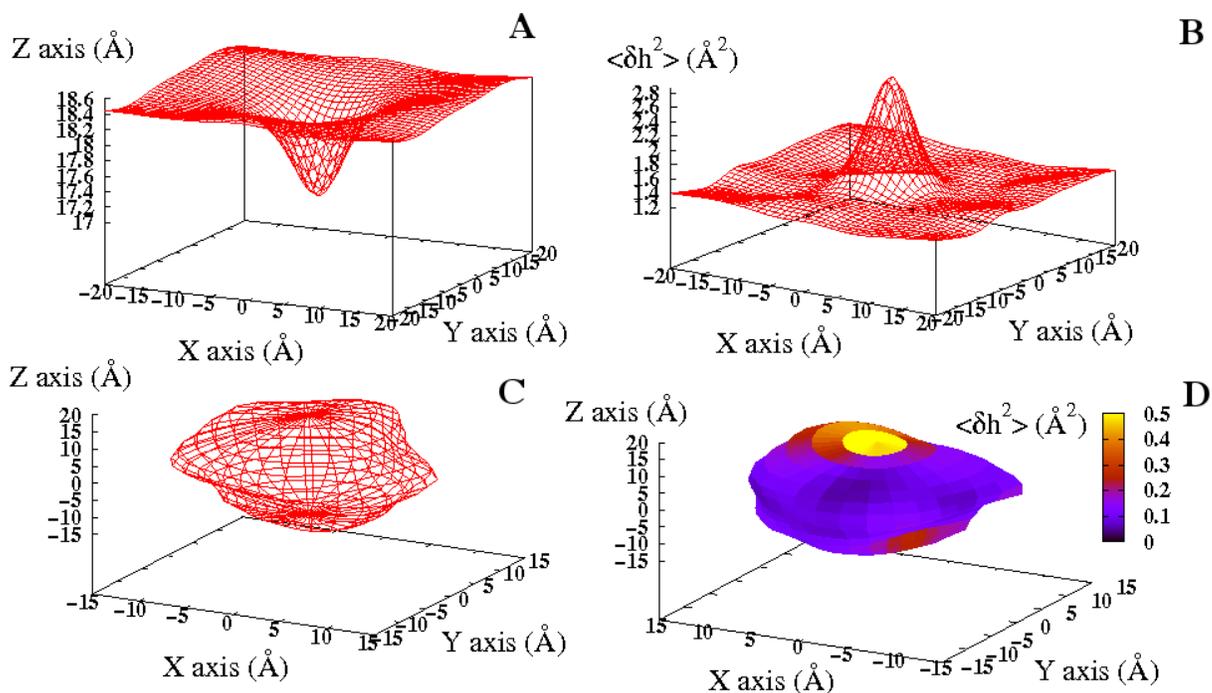


Figure 5.10: (A) Mean height of liquid-vapor interface as single Gdm⁺ locates at $z = 14$ Å (B) Surface height fluctuation of liquid-vapor system as single Gdm⁺ locates at $z = 14$ Å (C) Mean protein-solvent interface height as single Gdm⁺ locates at $z = 18$ Å (D) Height fluctuation of protein-solvent interface as single Gdm⁺ locates at $z = 18$ Å, the color scales representing the magnitude of fluctuation.

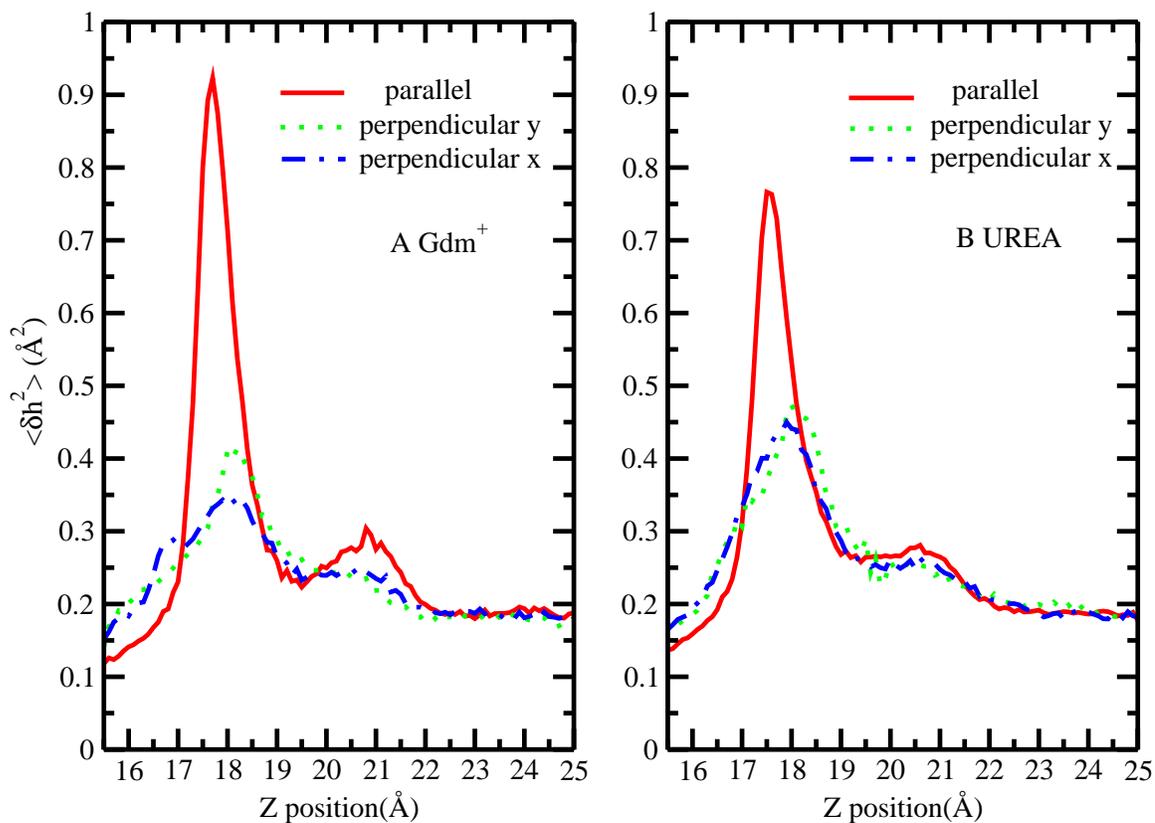


Figure 5.11: (A) Surface height fluctuation for hydrophobic protein interface at ($x = 0, y = 0$) as a function of restrain z position of single Gdm⁺ with parallel orientation, perpendicular y and perpendicular x orientation (B) Surface height fluctuation for hydrophobic protein interface at ($x = 0, y = 0$) as a function of restrain z position of single urea with parallel orientation, perpendicular y and perpendicular x orientation.

values of Δ and the distributions of Δ as single Gdm^+ is located at three representative positions, $z = 16 \text{ \AA}$, 18 \AA and 19.5 \AA . Panel A and D correspond to the parallel configuration; Panel B and E correspond to the perpendicular y configuration; Panel C and F correspond to the perpendicular x configuration. At a separation of $z = 18 \text{ \AA}$, where the Gdm^+ induces a large interfacial fluctuation, we observe that Δ is always larger than zero for all the three cases, indicating that the interfaces will always reside between the protein and the solute, so there is no artifact where the surface passes through the solute. The same applies to the situation that Gdm^+ is located at $z = 19.5 \text{ \AA}$ as indicated by the blue line. Although at the separation of $z = 16 \text{ \AA}$ there are some Δ values less than zero, at this point the interfacial fluctuation is suppressed by the presence of the solute. This suggests further that enhanced fluctuations are not influenced by the interface fluctuating on both sides of the solute. Furthermore, at closer separations of Gdm^+ and the protein-water interface, the meaning of the local interface becomes ambiguous perhaps, but this is not a serious issue as the major differences in interface effects occur well-before the solute arrives at the interface. An additional point worth addressing is that at a separation of $z = 18 \text{ \AA}$, we observe that the parallel configuration of Gdm^+ exhibits a wider distribution of Δ values as shown in Panel D green line. This is consistent with the earlier result of Figure 5.11A that parallel orientations of Gdm^+ induce larger fluctuations of hydrophobic protein-solvent interface compared to the perpendicular ones. In a recent study, Patel et al discussed that water near hydrophobic surfaces can be described as being near a phase transition characterized by enhanced fluctuations in relevant order parameters. [236, 35] The relevant order parameter is solvent density in their work, the distribution of which varies from unimodal (when two hydrophobic interfaces are far apart) to bimodal at separations where the volume between surfaces fluctuations between wet and dry states to unimodal once the inter-solute space is completely dry (post dewetting transition). In this work, as solutes approach the hydrophobic surface as a perturbation to the interfacial water, a different order parameter based on the interfacial height is considered.

Ideally, this interfacial height should display the same signatures of bimodal distribution as the solute reside at the position that induces the largest surface fluctuation. From previous discussion, Δ is the difference between z_{Gdm} and $z_{interface}$. Since z_{Gdm} is almost constant (this is the fixed position of the solute), the distribution profiles of Δ and $z_{interface}$ should be identical except the shift along X axis. We will just use the distribution of Δ in the following discussion. A fat tail in the distribution profile in Figure 5.11D is observed as the solute is located at a separation of $z = 18 \text{ \AA}$, which is near the position of largest fluctuation. Furthermore, the distribution profiles of Δ for Gdm^+ with parallel orientation at the separation of $z = 17.5 \text{ \AA}$ and $z = 17.7 \text{ \AA}$ are shown in Figure 5.12A. Interestingly, at the exact separation where solute induces the largest fluctuation, $z = 17.7 \text{ \AA}$, a prominent bimodal distribution is observed, which is consistent with the view that at this position, there would be transitions between a "wet" and "dry" region between the solute and the protein-water interface. Figure 5.12B shows the log probability of Δ versus Δ , analogous to Figure 3C in Patel et al [236]. The present probability distributions for the interface position recapitulate the results of Patel et al in a rather dramatic fashion. This further speaks to the notion that water near hydrophobic interfaces, even on the smaller scales of specific regions of biomolecules, is poised close to phase transitions, which upon perturbation by external potentials (in this case, a solute approaching the interface and perturbing the solvent density near the protein surface as a consequence of the nature of the solute's hydration shell) undergoes a transition. This transition is now considered as an alternative signature of the hydrophobic effect.

Finally, we consider PMF's of single Gdm^+ approaching a hydrophilic region on the protein surface. Figure 5.15A shows the PMF profiles of a single Gdm^+ approaching the hydrophilic protein patch with parallel orientation (red), perpendicular y orientation (green) and perpendicular x orientation (blue). A slight free energy stabilization is observed in all the cases, which may due to the electrostatic interaction between -NH group of Gdm^+ and side chain of hydrophilic residues (like D25) on the patch as shown in Figure 5.13C. However, compared with the free energy of Gdm^+ with

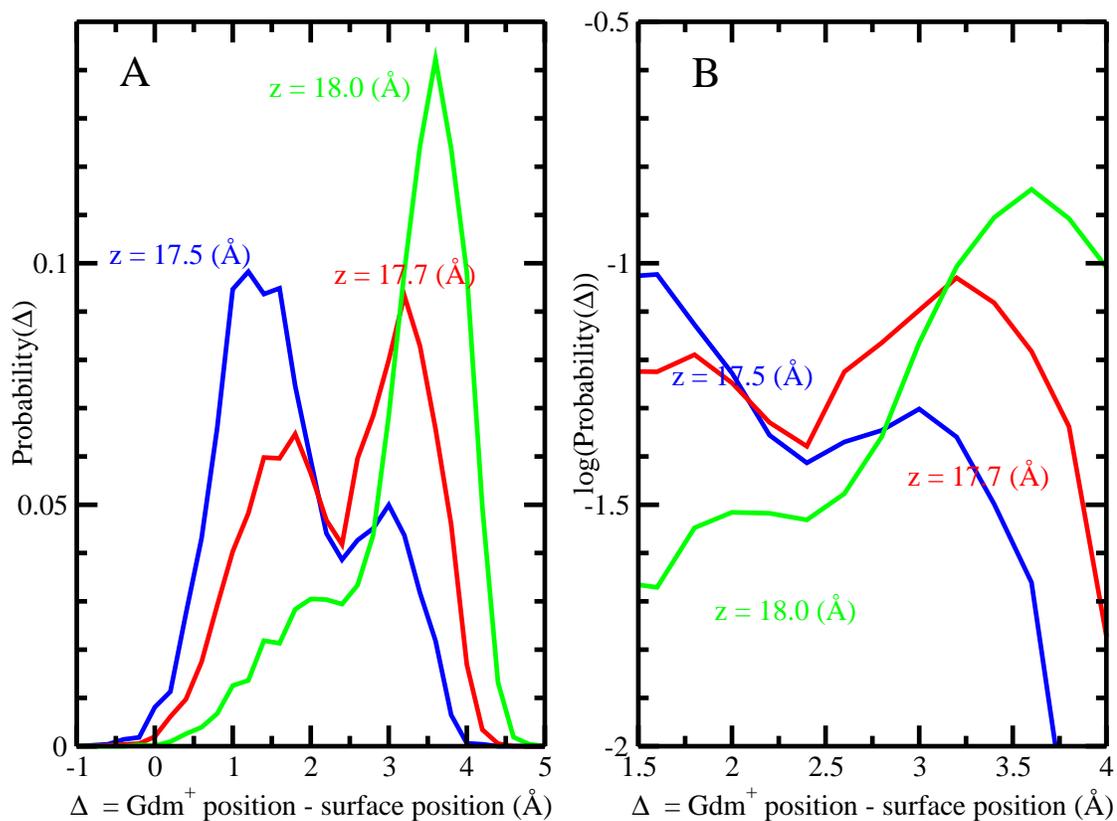


Figure 5.12: (A) Probability distributions of Δ for Gdm^+ with parallel restrained orientation locating at various positions close to the peak of largest fluctuation. (B) Probability distributions (log scale) of Δ for Gdm^+ with parallel restrained orientation locating at various positions close to the peak of largest fluctuation.

most favorable parallel configuration approaching the hydrophobic protein region (-3 ± 0.15 kcal/mol), the free energetic advantages from Gdm^+ with all three configurations approaching the hydrophilic patch are quite small (around -0.5 ± 0.15 kcal/mol). Furthermore, all three configurations show little to no difference in free energy suggesting that the orientational preference of Gdm^+ around certain types of surfaces is highly dependent on the effective hydrophobicity of the region, with significant orientational preference of Gdm^+ occurring around the more hydrophobic surface regions. Furthermore, the induced interfacial fluctuation profiles of single Gdm^+ with these distinct orientations approaching this hydrophilic region are shown in Figure 5.15B. Previously we have reported that for the protein in pure water regions with different hydrophobicity will display dissimilar inherent interface height fluctuations. [169] The larger magnitude of fluctuations are related to the malleable nature of the water and facile cavity formation around hydrophobic patches. [28] When Gdm^+ is located far from the patch, in all three cases, an inherent fluctuation value of 0.07 \AA^2 is detected, which is lower compared with the inherent fluctuation value around hydrophobic protein region 0.18 \AA^2 . As Gdm^+ moves closer to the hydrophilic interface, both parallel and perpendicular orientation have inappreciable effect on hydrophilic interfacial fluctuations. Although parallel configuration may induce a little larger fluctuation compared with the perpendicular one, the difference is quite small, around 0.02 \AA^2 . Such negligible differences in inducing fluctuations among these configurations corresponds to the marginal differences of free energies around hydrophilic interface.

5.4 Summary and Conclusions

In this article, we continue to explore and demonstrate a connection between interfacial stability and induced interfacial fluctuations as interfacially-stable solutes approach ostensibly hydrophobic aqueous-hydrophobe interfaces. The context in which we consider the present work is relevant for discussion of the nature of direct chemical interactions between typical chemical denaturants of proteins, Gdm^+ and urea,

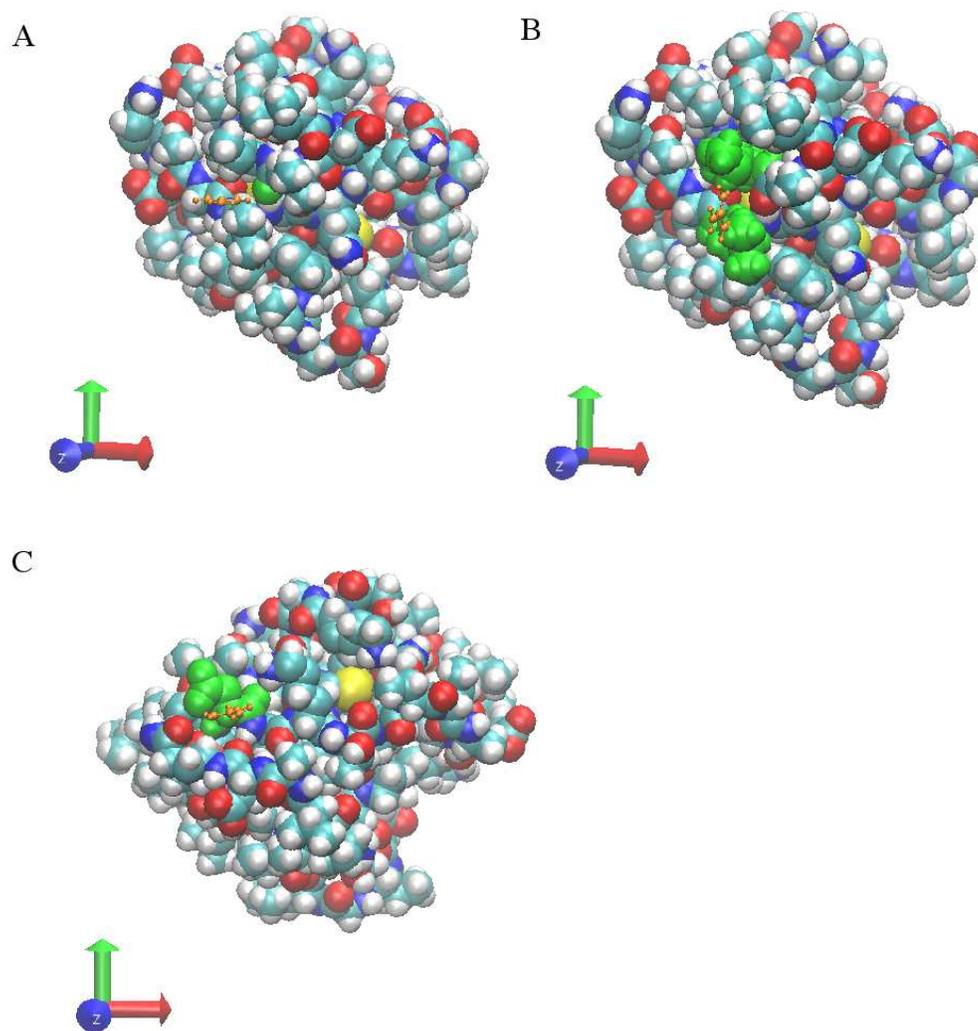


Figure 5.13: (A) Gdm⁺ with perpendicular y orientation (orange) around the hydrophobic protein patch (B) Gdm⁺ with perpendicular x orientation (orange) around the hydrophobic protein patch, residue I22 and L63 are shown in green (C) Gdm⁺ with perpendicular y orientation (orange) around the hydrophilic protein patch, residues D25 is shown in green.

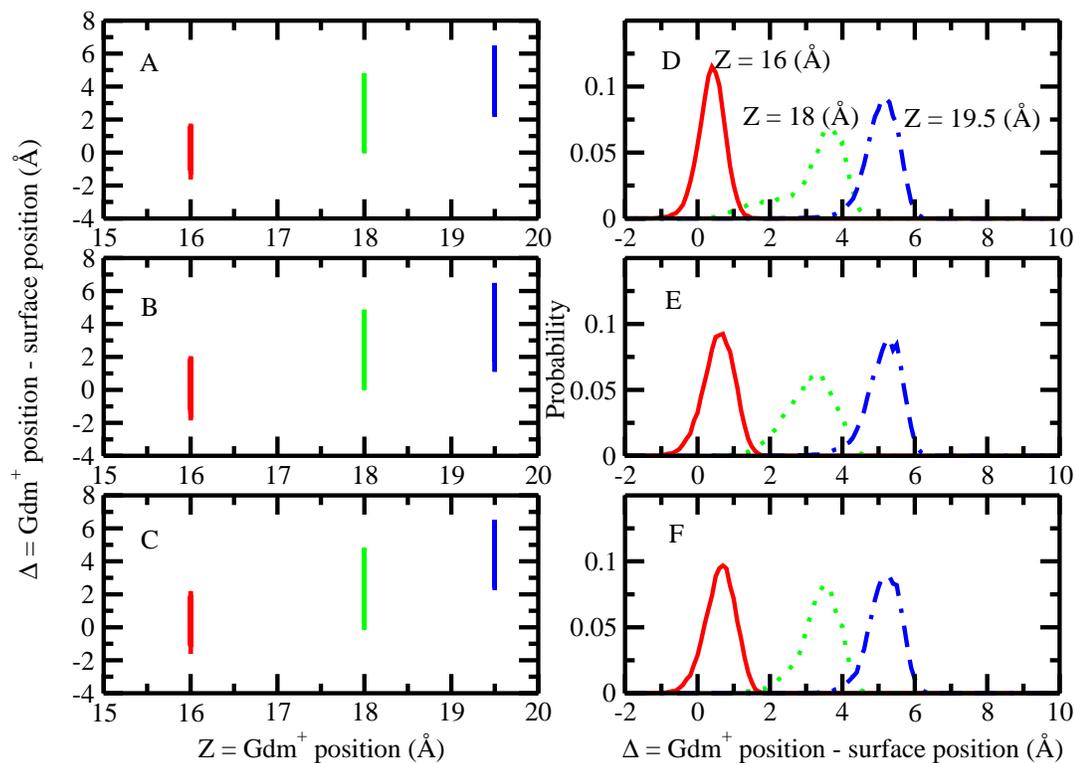


Figure 5.14: (A)-(C) Differences between single Gdm^+ position and surface position (surface position is defined by the surface height at position $x = 0$, $y = 0$) Δ for Gdm^+ with different restrained orientations locating at various positions, (A) parallel orientation (B) perpendicular y orientation (C) perpendicular x orientation (D)-(F) Probability distributions of Δ for Gdm^+ with different restrained orientations locating at various positions, (D) parallel orientation (E) perpendicular y orientation (F) perpendicular x orientation.

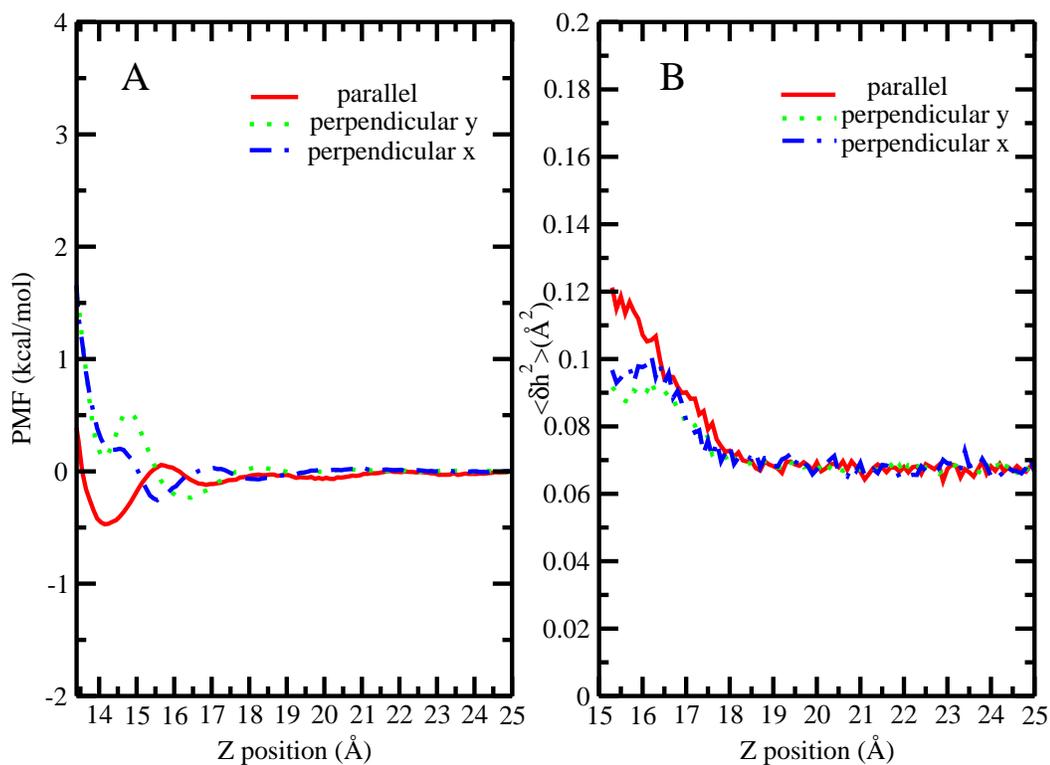


Figure 5.15: (A) PMF for single Gdm^+ with parallel orientation, perpendicular y and perpendicular x orientation from bulk approaching the hydrophilic protein-solvent interface (B) Surface height fluctuation for hydrophilic protein interface at $(x = 0, y = 0)$ as a function of restrain z position of single Gdm^+ with parallel orientation, perpendicular y and perpendicular x orientation.

specifically at hydrophobic regions of a model protein, HFBII. Our calculations of potentials of mean force indicate that Gdm^+ and urea exhibit non-trivial stability at the aqueous-hydrophobe interface as indicated by Figure 5.7 of this paper. Furthermore, we observe a richer subdivision of the contributions to the total free energy arising from two relative orientations of the solute that we have chosen to study, the parallel and perpendicular orientations as defined relative to the surface of the protein. Though the protein surface is not quite parallel to the axis chosen as our order parameter for calculations of potentials of mean force, the selected definitions, we feel, suffice for the current purposes. With respect to the orientation-free energy correlation, our calculations indicate that the orientations of both solutes in which the solute is parallel to the interface are associated with stronger free energy minima compared to configurations where the solutes approach in a perpendicular orientation. These two orientations appear to envelope the total free energy profiles (though we cannot say with certainty what contributions intermediate orientations would offer; however we stress that in this study, our aim is to demonstrate the self-consistency of the free energy profiles computed via the potentials of mean force with the orientation probability densities determined from free, solute-unrestrained MD simulations of the solutes in solution with the protein). Furthermore, we find that the correlative behavior between solute orientation and free energy stability (using the current force field combinations for water, solute, protein, and ions) mimics that observed at the aqueous liquid-vapor interface (Figure 5.2 and 5.4, probability distribution maps). Our results for both the protein-water interface and the pure liquid-vapor interface are in agreement with previous studies. [86, 87]

Recent simulations have highlighted the unique nature of hydrophobic interfaces as it relates to the fluctuations induced in solvent density vicinal to the interface (refer to Garde et al.'s work [36, 237]). Complementary studies have illuminated the fluctuations of aqueous-hydrophobe interfaces as simpler atomic species (monovalent ions) and slightly more complicated molecular species approach such interfaces. Both these approaches ostensibly define a further characteristic property associated with hydrophobic

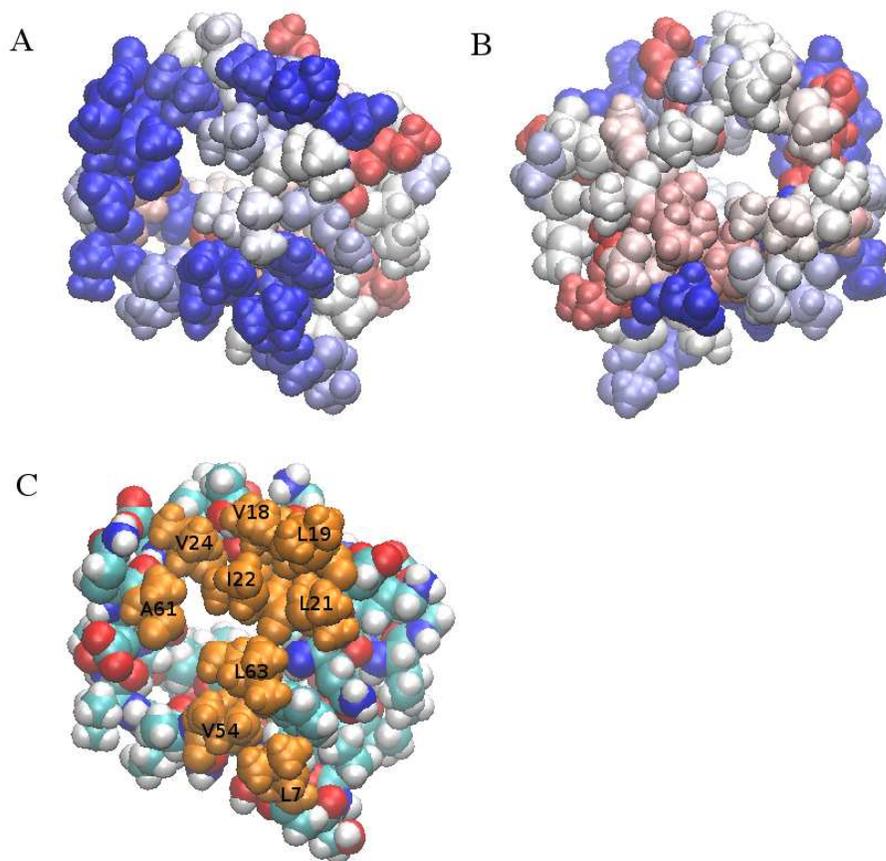


Figure 5.16: (A) Gdm⁺ number density map around HFBII protein (hydrophobic side). Blue represents higher number density, while red represents lower number density (B) Gdm⁺ number density map around HFBII protein (opposite side) (C) Representation of hydrophobic protein patch of HFBII with orange highlighting each hydrophobic residues on the patch.

solutes (and perhaps the hydrophobic effect in general). The present calculations indicate that associated with interfacial stability of the chosen chemical denaturants is an induced fluctuation of the interface upon approach of the solute to the interface. We stress that the induced fluctuations of the interface formed between the hydrophobic region of the protein and solvent occur before the solute resides directly at the interface. This is an important detail, as it speaks to the somewhat long-ranged nature of the effects generated by certain solutes prior to any direct interaction being realized. That solutes can affect an interface from a distance is a subtle though non-negligible effect we suggest. Moreover, the present results suggest that denaturant orientations that are parallel to the interface (vis-a-vis, display interfacial stability) are the orientations that induce the largest fluctuations of the interface (and hence the solvent density). The relation between solute orientation and induced fluctuations is related to the nature of the solvation shells of the solute presented towards the interface upon approach of the solute. In the case of Gdm^+ approaching the interface in a parallel orientation, the solvation shell presented is a more "malleable" one, where the solvent is more labile and free to rearrange. This leads to greater solvent density fluctuations and hence, higher interfacial induced fluctuations. In the case of the perpendicular orientations of Gdm^+ and urea, the tighter hydrogen bonding patterns of water (as demonstrated in previous studies[233] create a more rigid, well-defined solvation shell that is not easily disrupted. This translates to lower solvent density fluctuations, and hence lower induced fluctuations (or even suppression of interfacial fluctuations). The present results are thus consistent with recent work and provides yet another example of the relation of hydrophobic effect, solvent fluctuations, and interfacial stability. This relationship appears to be common across a series of atomic and molecular species, as well as encompassing charged, polar, and non-polar characteristics of the solutes considered. These observations suggest that molecular ions, such as Gdm^+ , as well as polar molecules with heterogeneous charge distributions (at least in the context of empirical molecular mechanics force fields) inherently have built into them regions of high

and low charge density. The dependence of local solvation structure on this heterogeneous (or asymmetric) charge density is to a large extent involved in determining the propensities of the modalities involved in specific association of molecules with specific types of interfaces. Observations based on classical simulations as well as recent DFT-based calculations [238] thus suggest an intriguing fundamental underlying theme. These ideas call for further study regarding specific details about the nature of the relationship between fluctuations, degree of solute hydrophobicity, solute solvation/hydration shell properties, and interfacial stability. Finally, Figure 5.16 shows the number density of Gdm^+ molecules in the vicinity of the canonical hydrophobic region of HFBII as well as on the side opposite to this hydrophobic patch (the opposite side not being hydrophobic to any significant extent). Our analysis of simulation data from 1M Gdm^+ solutions with no restraints demonstrates a propensity for the Gdm^+ to the hydrophobic region. This is consistent with the analyses presented in this work.

Chapter 6

A BIOCHEMICAL MODEL FOR BINDING MEDIATED BY HYDROPHOBIC INTERACTION - ASSOCIATION BETWEEN UBIQUITIN AND UBIQUITIN BINDING DOMAINS

Reproduced with permission from Di Cui, Shuching Ou, Sandeep Patel. "Free Energetics of Rigid Body Association of Ubiquitin Binding Domains: A Biochemical Model for Binding Mediated by Hydrophobic Interaction." *Proteins: Structure, Function, and Bioinformatics*. 2014,82 (7), 1453-1468. Copyright © 2014, Wiley Periodicals Inc.

6.1 Introduction

Weak intermolecular interactions underlie numerous molecular recognition processes;[\[239, 240\]](#) these can involve noncovalent association of two species (i.e., protein and ligand, protein and protein) with a wide range of specificity. Recent studies have explored the effects of weak intermolecular association such as hydrogen bonding and hydrophobic interactions and their role in stabilizing ligands at protein surfaces.[\[162\]](#) The role of hydrogen bonding in association processes has been extensively studied. [\[241\]](#) However, the elucidation of a general description of hydrophobic interactions between protein and ligand has proven elusive; it is still poorly understood how association is mediated by hydrophobic interactions in terms of specificity and affinity. There is general agreement that the contribution of hydrophobicity to recognition of protein to small (molecular) and large (protein) ligands is an important and relevant factor.[\[242\]](#) Although the importance of hydrophobic interactions in the binding process is well-appreciated, it is often assumed that this contribution is nonspecific compared to hydrogen bonding which confers specificity to protein-ligand association as a consequence of the inherent directionality of hydrogen bonds. However, recent investigation of structures of

protein-ligand complexes has enabled the reevaluation of the role of hydrophobic interactions in helping to define specificity of recognition. [243] A major issue concerning hydrophobic mediated binding processes is the driving force for protein-ligand interaction. A classical view of the characteristic thermodynamic signature of hydrophobic association is one of an entropy-driven process.[15, 91, 89] That is, the overall stabilizing contribution to the complexed state is an increase in total entropy (the origins of which generally are associated with the release of water degrees of freedom upon association, thus allowing more configurational states). In contrast, association between hydrophilic binding pockets and hydrophilic binding ligands is generally regarded as enthalpy-driven. A good model serving for interactions between hydrophilic ligands and hydrophilic binding pockets is galactose binding to Arabinose Binding Protein (ABP). Isothermal titration calorimetric (ITC) experiments demonstrate that the interactions are enthalpy-driven with a change of enthalpy about -95 kJ/mol. [90] However, when it comes to the case of association process between hydrophobic binding patch of protein and hydrophobic ligand, the observed underlying thermodynamic signatures are more complicated and surprisingly broad. Investigation of nanoscopic hydrophobic association [92, 94, 93] has shown that the process is highly entropically favorable; association of model hydrophobic surfaces, solutes, and (bio)macromolecules near hydrophobic surfaces have been extensively explored using molecular simulations providing a general picture of the entropy-driven association processes in these model systems. The explanation for this is that hydrophobic solute will disrupt the structure of bulk water and since it is incapable of hydrogen bonding with water, water-water hydrogen bonds are reoriented along such a surface in order to minimize disruption of the three-dimensional hydrogen bonded network of water molecules. This leads to a structured water cage around the hydrophobe surface, the classical Frank and Evans model.[8] The propensity of water molecules to predominantly adopt a subset of configurations to maximize interaction leads to significant loss of configurational entropy of water molecules. Such unfavorable effects can be minimized if hydrophobe molecules aggregate. Upon aggregation, water molecules form one larger cage surrounding the

hydrophobic aggregate and the surface area of such aggregate is smaller than the sum of surface areas of individual solutes. This makes the entropic contribution less unfavorable and, hence, makes the free energy more favorable. If this mechanism was the sole driving force for protein associations, all binding events involving hydrophobic binding partners would be entropy-driven. This is not the case. Recent experiment has investigated the binding process of nonpolar ligand to the poorly solvated pocket of the mouse major urinary protein-1(MUP-1). [96, 97, 98] Despite the apparent hydrophobic character of the binding partners, ITC data indicated that the binding was enthalpy-driven and accompanied by an unfavorable entropy change. The authors suggest that by virtue of poor solvation of the binding pocket, protein-ligand interactions (dispersion) arising from complexation compensate favorably any interactions lost between binding-site residues and solvent prior to the association. In stark contrast, the binding of a series of hydrophobically modified benzamidine chloride inhibitors to trypsin, which has a well-solvated binding site, is strongly entropy driven at a number of temperatures. [99] It suggests that the characteristic thermodynamic signature of hydrophobic association in solution will depend on the degree of solvation of the binding pocket. Molecular dynamics simulation study has been applied to investigate the enthalpy driven hydrophobic association by Setny. [100] A model for nonpolar cavity-ligand association is used in their molecular dynamics simulation. Thermodynamic contributions, including free energy, entropy and enthalpy along the binding coordinate have been investigated. The results show that the favorable driving force for this process is from enthalpy change among the release of water molecules from the hydrophobic environment to the bulk water. Although there are some controversial issues like the origin of the unfavorable entropic component for the hydrophobic association process, it is generally accepted that the enthalpy driven hydrophobic association usually involves the receptor protein that has poorly solvated binding sites. These poorly solvated binding sites harbor water that is termed “disorganized”; this reflects the notion that the water is not able to recoup energetically favorable water-water interactions via sufficient orientational restriction. More importantly, the above experimental and

simulation results suggest that the underlying signatures of hydrophobic interactions (mediated through the association via regions of macromolecules canonically labelled as “hydrophobic”) are by no means absolute and generally observed.

In part, a goal of this study is to extend the analysis and discussion of underlying signatures of the hydrophobic association in a system widely considered to be dominated by hydrophobic interactions, the ubiquitin interaction with a particular binding partner, UIM. MD simulations were performed to study the binding process between ubiquitin and its binding partner - ubiquitin interacting motif (UIM). It has been extensively characterized so far that ubiquitin binds to different kinds of UIM domains through one particular hydrophobic patch/region that includes residues L8-I44-H68-V70. [244, 245, 246] The binding domain we investigated is UIM of Vps27, which adopts a helical conformation. The helix is markedly amphiphilic with a hydrophobic stripe along one side which interacts with the complementary hydrophobic Leu8-Ile44-Val70 region of ubiquitin. [247] Mutation of the selected residues on both ubiquitin [170] and UIM [248] have suggested that these regions are involved intimately in the binding interface. Based on these observations, we consider that hydrophobic interactions may play a role in mediating the association process between the ubiquitin and UIM. Our aim presently is to investigate the thermodynamic signature of the association of these two proteins via their hydrophobic regions. We aim to use molecular dynamics simulations in conjunction with free energy sampling methods to calculate the potential of mean force (PMF) for reversible association of the two proteins taken to be semi-rigid bodies. Arguments for considering a rigid-body process are discussed in the Methods section. We also propose to evaluate the influence of hydration level around the binding patch on the driving force for association. We proceed to decompose the computed reversible work (PMF), which reflects the free energy difference between the associated and dissociated states of the protein-protein complex, into enthalpic and entropic components. Such decomposition enjoys a long history in its application to the study of the thermodynamics of association processes in solution. We further explore the dependence of relative orientations of ubiquitin

and the UIM domain on the potential of mean force. These latter calculations also provide indication of the correlative capability of current force field methods to predict the experimental structure of the bound complex as a free energy minimum. This is in the spirit of recent studies using coarse-grained Martini protein-lipid-solvent models that have shown that the binding interfaces of G-protein coupled receptors in model bilayers are associated with global free energy minima with respect to orientations of the two proteins taken as rigid bodies[249].

6.2 Materials and Methods

6.2.1 Simulation Details

Molecular dynamics simulations were performed with NAMD, version 2.9b3,[184, 185] using the CHARMM 22 all-atom force field (Chemistry at Harvard molecular mechanics) [186] with CMAP backbone torsion correction term.[187] Isothermal - isobaric ensemble (NPT) simulations were performed using a rectangular cell with a box size $60 \text{ \AA} \times 60 \text{ \AA} \times 100 \text{ \AA}$ as shown in Figure 6.1. A rectangular system was selected in order to probe the association along a distance between the centers of mass of the protein and binding domain, while minimizing the computational overhead involved in computing interaction forces between more waters included in a larger cubic box. [250] The initial structures of the complexes were constructed using CHARMM-GUI. [189] The ubiquitin-UIM complex with PDB code 1Q0W was placed in the center of the box, surrounded by 10738 TIP3P model[190] water molecule and 3 K^+ , which are used to neutralize the -3e charge of the UIM. Potentials of mean force (to be discussed further below) were computed along a reaction coordinate defined as the distance between the centers of mass of the two associating proteins. The larger ubiquitin was biased to remain in a single orientation and its center of mass at a specific position, chosen as $(x=0 \text{ \AA}, y=0 \text{ \AA}, z=0 \text{ \AA})$ via the use of strong restraining potentials. Using NAMD's collective variable infrastructure, ubiquitin's center of mass was restrained at $(x=0 \text{ \AA}, y=0 \text{ \AA}, z=0 \text{ \AA})$ using a force constant of $500 \text{ (kcal/mol)/\AA}^2$, and its orientation was

restrained about the NMR-based orientation using a harmonic restraint potential with force constant of 5000 (kcal/mol)/Å².

The helical structure of the UIM was maintained by harmonically restraining backbone atoms with a force constant of 100 (kcal/mol)/Å². Since we are only interested in the potential of mean force along one particular path, we chose to maintain the UIM helix as more or less rigid as this freezes some of the orthogonal degrees of freedom (orientational degrees of freedom for instance) and facilitates convergence along the chosen reaction coordinate. Furthermore, experimentally, Nuclear Overhauser Effect (NOE) measurements indicate that the polypeptide backbone within the helical region is relatively constrained compared with other regions of the UIM. [170] Also, from molecular dynamics simulations of the free protein in solution, we observe that the helix backbone RMSD is low as shown in Figure 6.2, reflecting this rigidity of the helix relative to other flexible regions of the short peptide. We do acknowledge that the simulation reflects the bias for helical propensity introduced by the CMAP correction; the force field, over the time scales we are concerned with, nevertheless suggests a low drift from the experimental structure, and thus we claim this as sufficient to warrant treating the UIM helix as a rigid body for the purposes of this work.

Temperature was maintained by Langevin bath at 300K, and the pressure was kept constant by Langevin pressure control at 1 atm. A switching distance of 10 Å, non-bonded real-space cutoff of 12 Å and pairlist generation distance of 14 Å were used for the van der Waals interactions, and the particle mesh Ewald (PME) method was employed for the calculation of conditionally-convergent electrostatic interactions.[191] The grid size of PME in x dimension is 60, in y dimension is 60, and in z dimension is 100 (as close to a 1Å grid point separation as possible). The SHAKE algorithm [192] was used to constrain bond lengths involving hydrogen atoms and an integration time step of 0.8 fs was used; this slightly lower time step is needed due to the PME update frequency of every 4 steps we use; we have ascertained that the use of this protocol does not affect the dynamics or energetics as the density of water in the bulk regions of the system are equivalent when using a more frequent PME update as shown in Figure

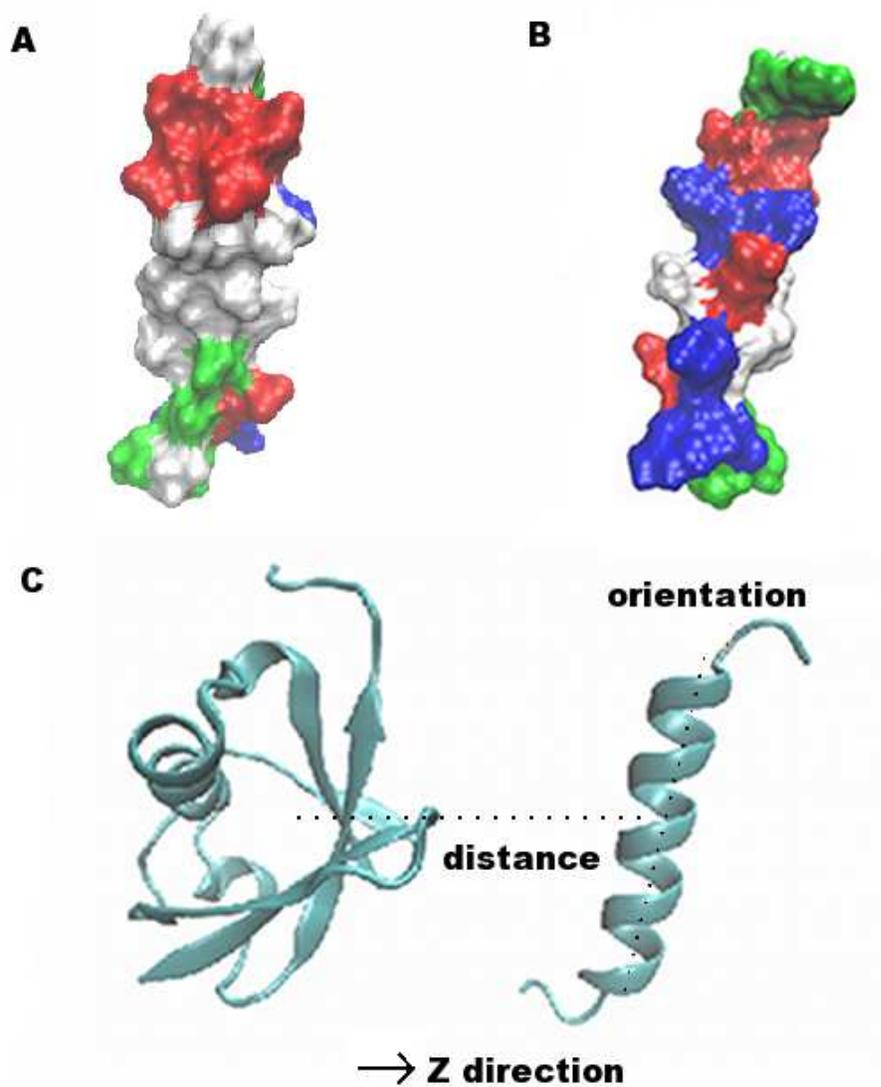


Figure 6.1: Representative snapshots of the system used in the study (A) representation of the hydrophobic side of the UIM helix. Non-polar residues, white; basic residues, blue; acidic residues, red; uncharged hydrophilic residues, green (B) representation of the hydrophilic side of the UIM helix (c) representation of the ubiquitin and UIM binding. The distances change from 15 Å to 34 Å. The orientations change from -150° to 180° .

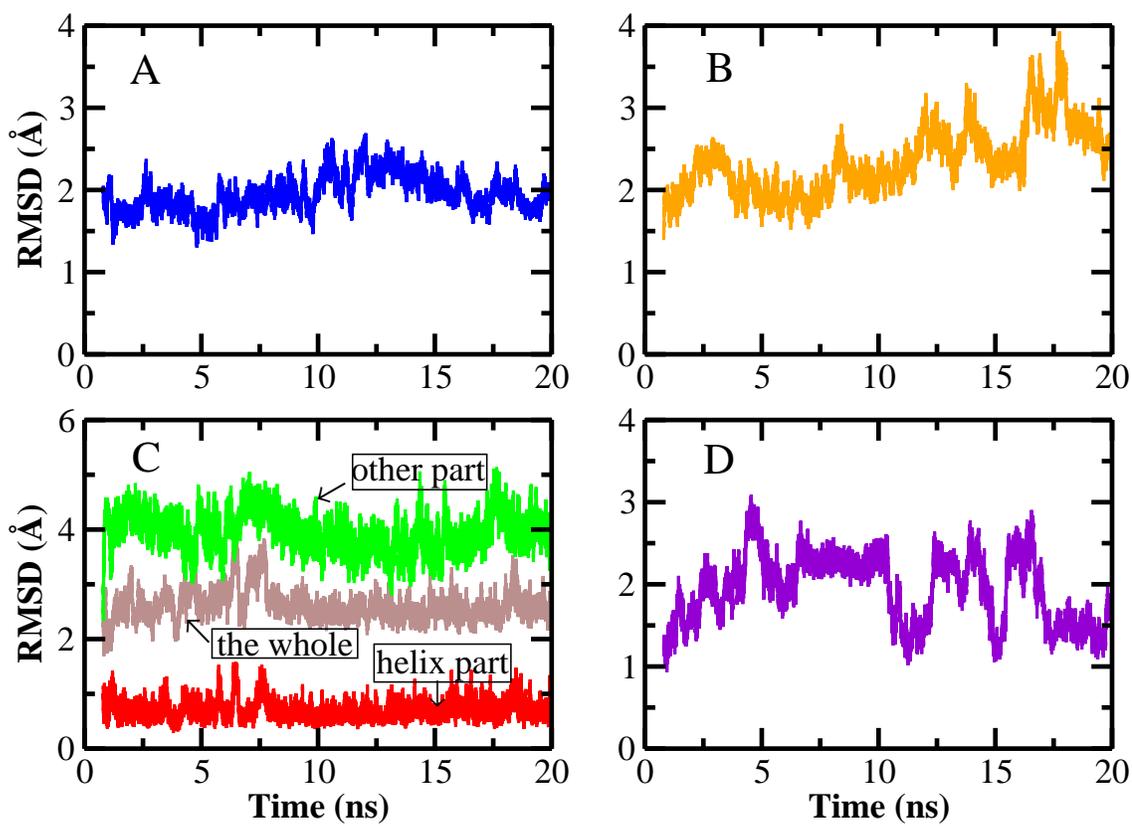


Figure 6.2: Evolution of protein backbone RMSD (A) Ubiquitin-UIM complex with 0° orientation (B) Ubiquitin-UIM complex with -30° orientation (C) UIM (D) Ubiquitin.

exclude	scaled1-4	langevinPistonPeriod	50
1-4scaling	1	langevinPistonDecay	25
COMmotion	no	langevinPistonTemp	300
zeroMomentum	yes	useFlexibleCell	no
dielectric	1.0	useGroupPressure	yes
switching	on	cellBasisVector1	60.00 0.00 0.00
switchdist	10	cellBasisVector2	0.00 60.00 0.00
cutoff	12	cellBasisVector3	0.00 0.00 100.00
pairlistdist	14	cellOrigin	0.00 0.00 0.00
timestep	0.8	wrapAll	on
stepspercycle	20	PME	yes
nonbondedFreq	2	PMEGridSizeX	60
fullElectFrequency	4	PMEGridSizeY	60
rigidBonds	all	PMEGridSizeZ	100
langevin	on	constraints	on
langevinDamping	5	selectConstraints	on
langevinTemp	300	selectconstrX	on
langevinHydrogen	off	selectconstrY	on
langevinPiston	on	colvars	on
langevinPistonTarget	1.01325	extraBonds	on

Table 6.1: NAMD input parameters for the simulations.

6.3. We provide the NAMD input file for our simulations in Table 6.1.

6.2.2 Potential of Mean Force

In order to determine the potential of mean force (PMF) describing the free energy of association of ubiquitin and UIM, a reaction coordinate defining this pseudo-chemical reaction must be defined. Presently, we consider a reaction coordinate, ξ , defined as the Cartesian z component of the separation between the center of mass of the UIM and a dummy atom located at the position ($x=0$, $y=0$, $z=0$); this point coincides with the restrained center of mass of the ubiquitin. In Figure 6.4, we show the deviation of ubiquitin center of mass from this point is sufficiently small so as not to incur any systematic error in the potential of mean force we compute. We use a dummy atom instead of the actual center of mass of ubiquitin as a collective variable

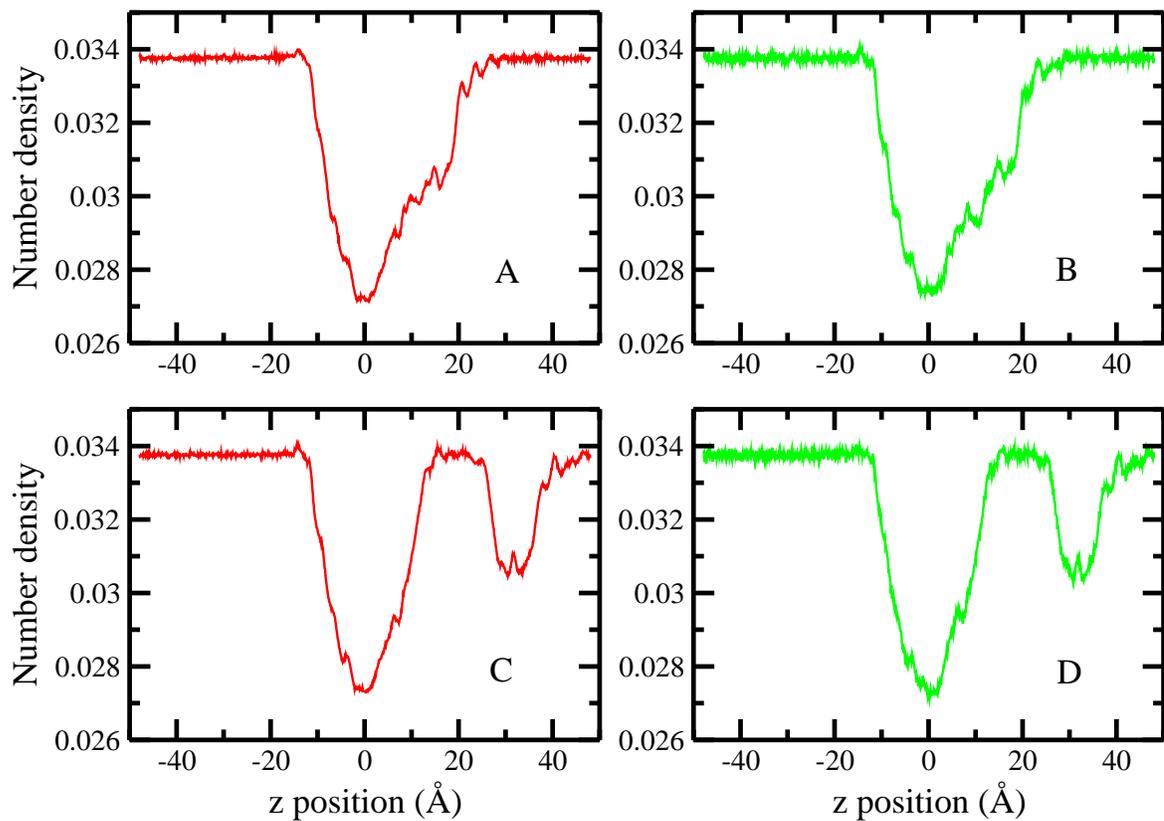


Figure 6.3: Number density profiles of water (A) associated state window [15:16] using timestep 0.8 fs (B) associated state window [15:16] using timestep 1.0 fs (C) separated state window [32:33] using timestep 0.8 fs (D) separated state window [32:33] using timestep 1.0 fs.

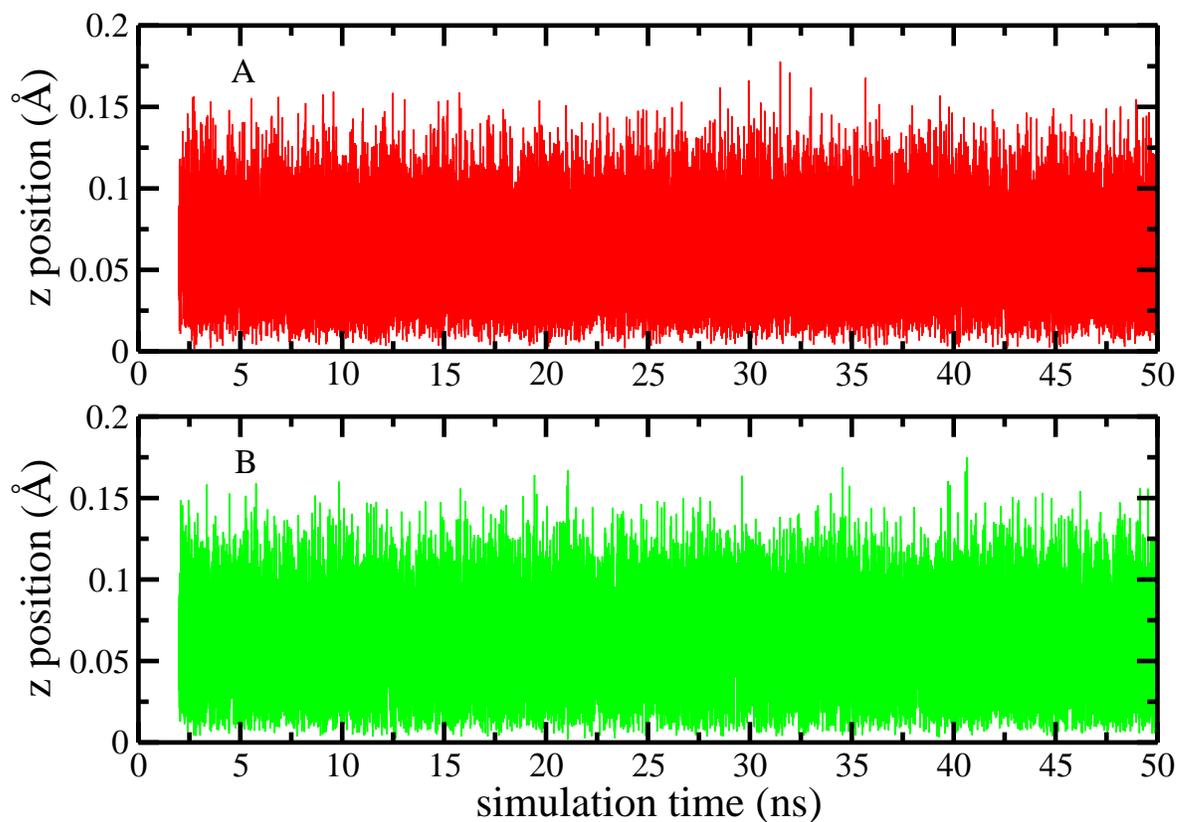


Figure 6.4: Evolution of z component of the center of mass of ubiquitin (A) associated state window [15:16] (B) separated state window [32:33].

group in NAMM because using a smaller group (smaller number of atoms) alleviate the loss of parallel performance in NAMM. [251]

NAMM provides several sampling methods[252] for calculation of PMF's. Here we use the Adaptive Biasing Force (ABF) method[139, 140, 141, 142]. For more detailed discussion of ABF, it can be found in Chapter Two. To enhance sampling of the distribution of configurations where the reaction coordinate holds a particular value, the reaction coordinate is restrained within a certain narrow range (instead of its entire span). At the boundaries of the narrow range of interest, relevant restraint potentials are introduced on the reaction coordinate in order to prevent it from moving outside of the desired range. In this work, we construct twenty continuous "windows" with

width 1.0 Å along the positive z-direction ranging from a separated to contact state. The spans of the windows going from separated to contact state (in Å) are: [34:35], [33:34], [32:33], [31:32], [30:31], [29:30], [28:29], [27:28], [26:27], [25:26], [24:25], [23:24], [22:23], [21:22], [20:21], [19:20], [18:19], [17:18], [16:17], [15:16]. Forces are accumulated in smaller bins of width 0.02 Å within each window as per the ABF protocol. During the production free energy calculations, the ABF method introduces a biasing force acting on the UIM; the bias force is equal in magnitude and opposite in sign to the actual mean force on the UIM. The mean force is an average over the preceding 500 molecular dynamics steps.

For the association process between ubiquitin and UIM, initially, we consider the situation that UIM approaches the ubiquitin with the orientation from NMR determined structure. The starting structures for each ABF window are constructed as follows. Initial solution structure of UIM-ubiquitin complex was arranged in a way that the center of mass of ubiquitin and that of UIM are both located on the z axis ($x=0$, $y=0$) and the relative position between the two proteins remains the same. Specifically, the center of mass of the ubiquitin was located at (0, 0, 0) and center of mass of the UIM was located at (0, 0, 15.25). Therefore, this initial structure was considered as the associated state window [15:16] of the ABF sampling coordinate. The starting coordinates of other windows were obtained by translating the UIM along the positive z axis while maintaining the position of ubiquitin the same (i.e., for window [16:17], coordinates of all the atoms of UIM translocating along z axis for 1 Å). We also consider how the binding PMF varies with the orientation of the UIM relative to the NMR-based structure orientation. We define the NMR orientation of UIM in the solution structure of the UIM-ubiquitin complex as 0° orientation. Other orientations were generated as UIM was rotated along an axis taken as the line connecting the center of mass of the first half of UIM helical region (including residues 4-12 in UIM) and the center of mass of the second half of UIM helical region (including residues 13-20 in UIM). Coordinates of all the atoms of UIM are rotated along this axis with the corresponding degree of 30°, 60°, 90°, 120°, 150°, 180°, -30°, -60°, -90°, -120° and -150° respectively. Positive

<i>Orientation</i> (°)	<i>Simulation Time</i> (ns)
-150	30
-120	30
-90	42
-60	50
-30	42
0	50
30	46
60	42
90	42
120	42
150	42
180	38

Table 6.2: Duration of the simulated trajectory of various orientations.

angles correspond to UIM rotated clockwise along the axis; negative angles correspond to anticlockwise rotation. Since the backbone atoms of UIM were restrained with a large force constant as previously discussed, the desired orientations were maintained during the simulations. For each window, we allowed at least 2ns of equilibration before considering the rest of simulation data as production data. Durations of the production trajectory for different orientations are reported in Table 6.2.

6.3 Results and Discussion

6.3.1 Free Energy, Enthalpy and Entropy Changes

We first consider potentials of mean force, and specifically that of semi-rigid UIM, with 0° orientation, associating with ubiquitin as shown in Figure 6.6 (blue solid line). The furthest right region represents the dissociated state and the left side represents the region of the associated state. Uncertainties in the free energy profile are determined as: [219]

$$\text{var}[G(\xi_N)] \approx \sum_{i=1}^N \text{var}[K \Delta \xi \bar{z}_i] \quad (6.1)$$

where $\text{var}[G(\xi_N)]$ is the variance, \bar{z}_i is the mean position of z in the i_{th} window, which can be obtained from block averages.[220] The standard deviation $\sigma[G(\xi_N)]$ is then the square root of $\text{var}[G(\xi_N)]$. The obtained free energy profile is flat from $d = 35 \text{ \AA}$ to $d = 21 \text{ \AA}$. At small separations, the PMF decreases monotonically, reaching a minimum at $d = 15.75 \text{ \AA}$; this position is the simulation-based contact state. We note that from the initial solution structure of the complex, the separation between the center of mass of ubiquitin and that of UIM is around 15.25 \AA ; the distance predicted by the combination of force field and simulation methodology agrees rather well to the NMR structure. After 50ns production of simulation, the free energy value at contact state is $-16.20 \pm 0.51 \text{ kcal/mol}$. To assess the convergence of the free energy profile, we show the time evolution of the free energy difference between contact and separated states in Figure 6.5. We further note that in this relative orientation of ubiquitin and UIM, there appears to be no free energy barrier to association; this is unlike common potentials of mean force of model hydrophobic entities associating in pure water which exhibit oscillatory barriers in the free energy profile itself as water layers gradually evacuate the region between hydrophobes.[253, 95] In prototypical models of hydrophobic association, the barriers are related to enthalpic contributions to the potential of mean force, analysis of which we now turn to.

We consider decomposition of the total free energy into enthalpic and entropic components to assess the relative contributions of these thermodynamic quantities to the overall association process. For a particular ubiquitin-UIM separation d , total system enthalpy relative to the separated state is computed as the difference of the average potential energy of the dissociated state (furthest separation studied) and the state at the separation of interest. As the difference in the pressure-volume term is close to zero for our system, we approximate the enthalpy at a distance, d , relative to the separated state as in Equation 6.2.

$$\Delta H(d) = H(d) - H(d_{\text{dissociated}}) = \Delta U(d) + \Delta(PV) \approx \Delta U(d) = U(d) - U(d_{\text{dissociated}}) \quad (6.2)$$

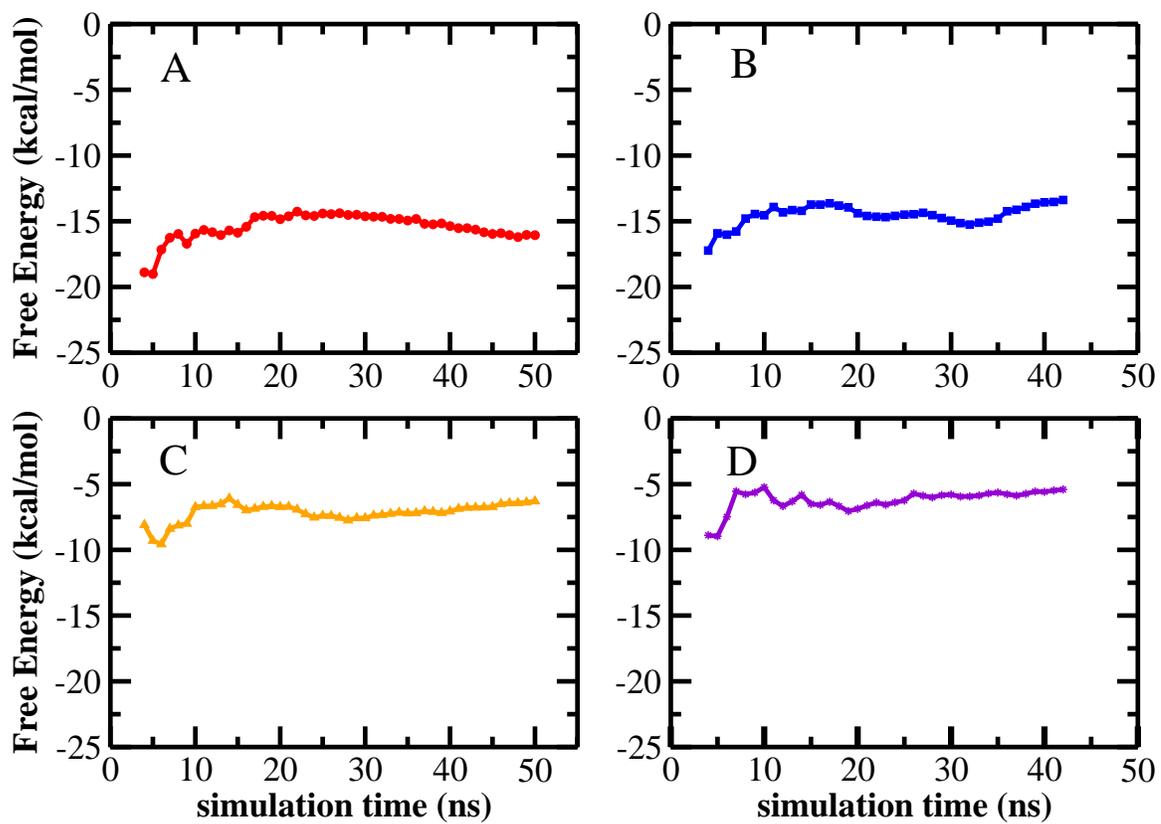


Figure 6.5: Evolution of the free energy for different orientations (A) 0° orientation (B) -30° orientation (C) -60° orientation (D) -90° orientation.

where $H(d)$ and $U(d)$ denote the enthalpy and potential energy of the system at separation d respectively.

System entropy is extracted via $-T\Delta S = \Delta G - \Delta H$. In Figure 6.6, the red dash line shows enthalpic component (ΔH) as a function of separation, and the green dash line shows the entropic component ($-T\Delta S$). The uncertainties of enthalpic component were again obtained based on blocking method [220] and the uncertainties of entropic component could be derived from the summation of free energy and enthalpy variances. Since entropic component is derived from free energy and enthalpic component, it is nature to expect that in the range of $d = 20 \text{ \AA}$ to $d = 35 \text{ \AA}$, where the free energy profile is flat, the enthalpic component and the entropic component are complementary. Within uncertainty, both the enthalpic component and entropic component are zero for the separated state from $d = 28 \text{ \AA}$ to $d = 35 \text{ \AA}$. Since the potential energy is a large, fluctuating value, the variation of ΔH for $d > 22$ is difficult to interpret. However, we note an observed substantial increment of the enthalpic component which is as high as $+20 \text{ kcal/mol}$ at a separation between $d = 20 \text{ \AA}$ and $d = 22 \text{ \AA}$. When the separation is closer, enthalpic component remains positive in sign and fluctuates about a statistically different mean value compared to states with larger separation. The enthalpy profile, taken at face value, appears qualitatively similar to profiles observed in association of hydrophobic surfaces and solutes. In such cases, as mentioned above, the emptying of the region between hydrophobes where water-water interactions are exaggerated (more favorable than in bulk) in order to compensate entropy loss, results in a net enthalpic destabilization upon hydrophobe association. We will further explore the behavior of water and details of the contributions to the enthalpy change further below. As expected, system entropy increases in the range of $d = 20 \text{ \AA}$ to $d = 22 \text{ \AA}$ and the entropic component remains negative (favorable) at smaller separations. At the contact state, $-T\Delta S = -26.70 \pm 1.62 \text{ kcal/mol}$, which is favorable for the association. Ostensibly, this entropic stabilization intimates that release of water degrees of freedom as water evacuates the inter-protein region sufficiently compensates any entropic losses from protein modes that are lost upon binding. We will comment on these issues

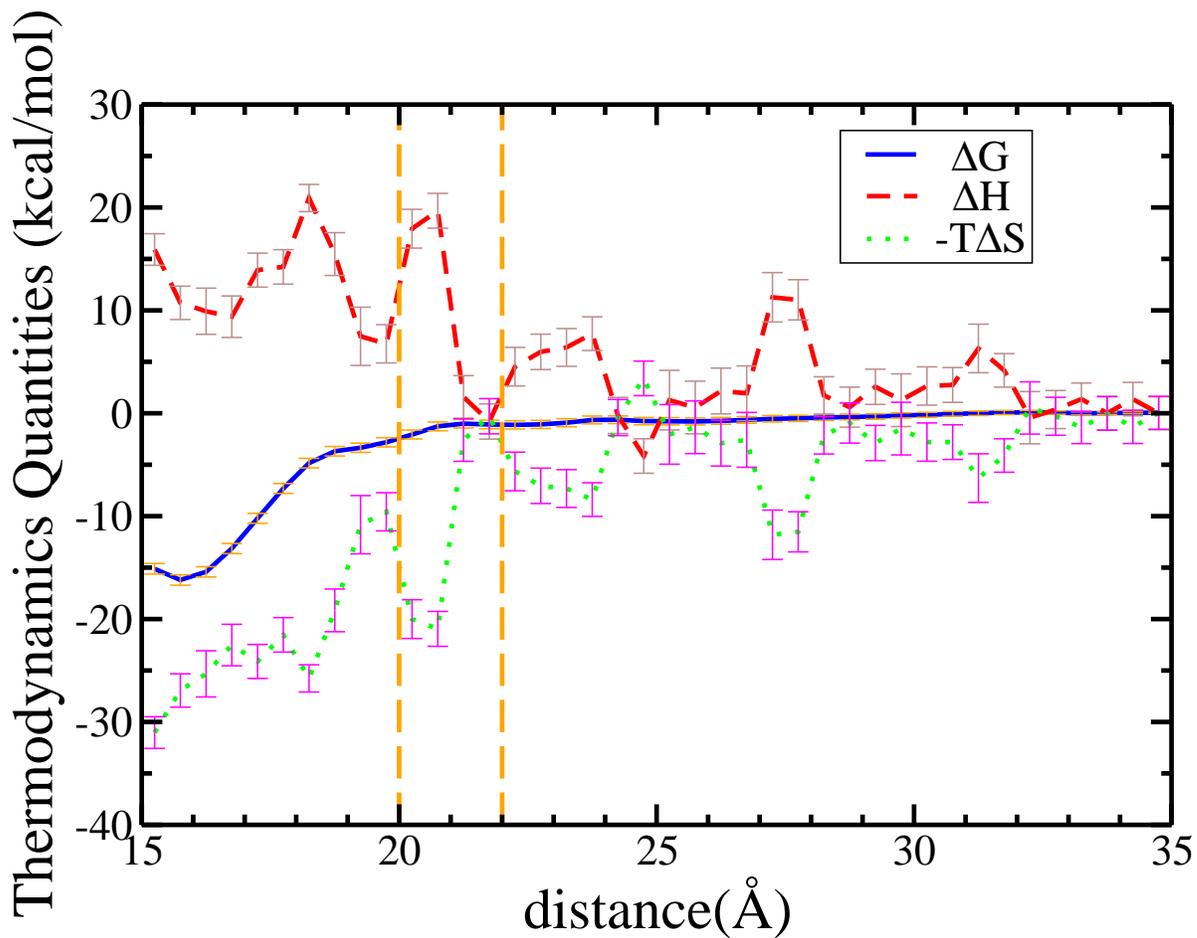


Figure 6.6: Thermodynamic contributions along the reaction coordinate. Relative Gibbs free energy, G (red), enthalpy, H (blue) and entropic term, $-TS$ (green), and their uncertainties (vertical bars).

further below.

We note that different contributions of enthalpy and entropy to overall thermodynamic stabilization of the complex represents entropy-enthalpy compensation, which is a general character of weak intermolecular interactions in biological systems. [254] Presently, free energy stabilization of the contact state in our model system arises from the entropic component. The favorable entropic contribution is large enough to stabilize the association process even after compensating the unfavorable enthalpic effect. The result here apparently recapitulates the thermodynamic signature for the classical hydrophobic effect mediated association process between two hydrophobic binding partners.[255, 99, 92, 93] Since it is argued that the characteristic thermodynamic signature of hydrophobic association in solution will depend on the degree of solvation of the binding region, the entropy-dominated association process between ubiquitin and UIM implies that the binding region of ubiquitin should be exposed on the protein surface and can be fully solvated in the absence of binding partner. Relating directly to previous studies on the association of nanoscopic hydrophobic graphene plates in water, Choudhury et al[94] obtained the same thermodynamic signature including highly favorable entropic contribution at the contact state (-460 kJ/mol) and the opposite unfavorable enthalpic contribution (250 kJ/mol). Interestingly, the entropic component contribution profiles in both the case of graphene plates association and that of ubiquitin-UIM association appear quite similar, exhibiting a virtually flat region at large separations and a sharp decrease at smaller separations. The observed changes of thermodynamic quantities are connected with the hydration level around the solutes. When the two hydrophobes contact one another, expulsion of water molecules from the intersolute region occurs. The release of the structured water into the bulk results in the increment of entropy (decrease in entropic component $-T\Delta S$). Simultaneously, the favorable interaction energy arising from attractive interactions between the solute and water in the confined region is lost, which accounts for the highly unfavorable enthalpic component.

6.3.2 Interaction Energies from Different Components

We address in further detail total system enthalpy along the separation distance by decomposing it into contributions from interactions between specific system components, namely, protein (ubiquitin), domain (UIM), water and counter ions (K^+). We consider four elements, protein-domain interactions, domain-water interactions, protein-water interactions, and water-water interactions as shown in Figure 6.7. Panel A and Panel B show the protein-water interaction energy and domain-water interaction energy, respectively. Approaching the contact state, both protein-water and UIM-water interactions are lost, relative to the separated state, resulting in each of these interaction components contributing large, unfavorable enthalpic contributions to the total free energy. The destabilizing contributions result from desolvation of the protein and domain in water as the UIM approaches ubiquitin, which has also been observed previously in the context of protein-ligand association. [255, 100] Panel C shows water-water interaction energy, and Panel D depicts the interaction energy of protein with domain. These two interaction energy profiles are quite similar, providing a stabilizing negative enthalpy of association in opposition to the protein-water and domain-water cases. The protein-domain interactions are attractive, globally arising from a combination of polar, charged, and hydrophobic residues interacting with one another; again, ostensibly, this appears much like the hydrophobe-hydrophobe interaction energy observed in numerous past simulation studies. When the separation is small enough ($d = 15 \text{ \AA}$), repulsive interactions dominate, indicated by the slight increase in protein-domain interaction energy at this region. Before the appearance of a global minimum at this region, there is another minimum between $d = 22 \text{ \AA}$ and $d = 23 \text{ \AA}$. This may relate to the favorable intermolecular electrostatic interaction between side chains of negatively charged glutamic acid residues at the N-terminus of the UIM and positively charged arginine residues on the surface of ubiquitin as mentioned in the literature. [170] Further evidence about this very specific interaction in this localized region of the reaction coordinate is shown in the Figure 6.8. At window [22:23], the relative orientation of ubiquitin and UIM allows the closest distance and most favorable

interaction between GLU 5 on the UIM and ARG 74 on the ubiquitin. In addition, interactions involving GLU 3, GLU 5, GLU 6, GLU 7 and ARG 42, ARG 72, ARG 74 also come to bear as shown in Figure 6.8C. The importance of neighboring hydrophilic residues cannot be trivialized and we acknowledge that there is a contribution from this type of residue, even in the present system. The importance of hydrophilic interactions in protein-protein association and folding processes has been discussed deeply in the recent literature by Ben-Naim[256, 257]. Based on our computational results, we can only go so far as to suggest the importance of hydrophilic interactions/residues in the ubiquitin-UIM association process; it is clear that the stabilizing driving force from these interactions is offset by the other system components. This may or may not be a consequence of improper balance of interaction energy scales in the force fields used, and this certainly warrants broader studies. The water-water interaction component also contributes favorably at the contact state, which arises from association of the protein and domain, squeezing out the water and enhancing the water-water interaction. Overall, the loss of water-protein and water-UIM interactions dominate the destabilizing contributions to the enthalpy of association; stabilizing, favorable water-water and protein-domain interactions are insufficient to compensate the loss of former interactions, at least in this system. Finally, we note that the interaction energies are computed using full PME energies; in Figure 6.9, we show that these energies are qualitatively (and in most cases quantitatively) similar to energy values computed using large cutoffs without PME.

6.3.3 System Component Contributions to Potential of Mean Force

We can now connect enthalpy decomposition results from the previous subsection to how water-protein and protein-protein interactions contribute to the free energy vis-a-vis the potential of mean force. By averaging the force on the UIM from individual system components, such as water, one can extract the potential of mean force contribution from that system component by integration of this average force. This is a well-defined protocol.[95, 258]. As shown in Reference[258], the average force along

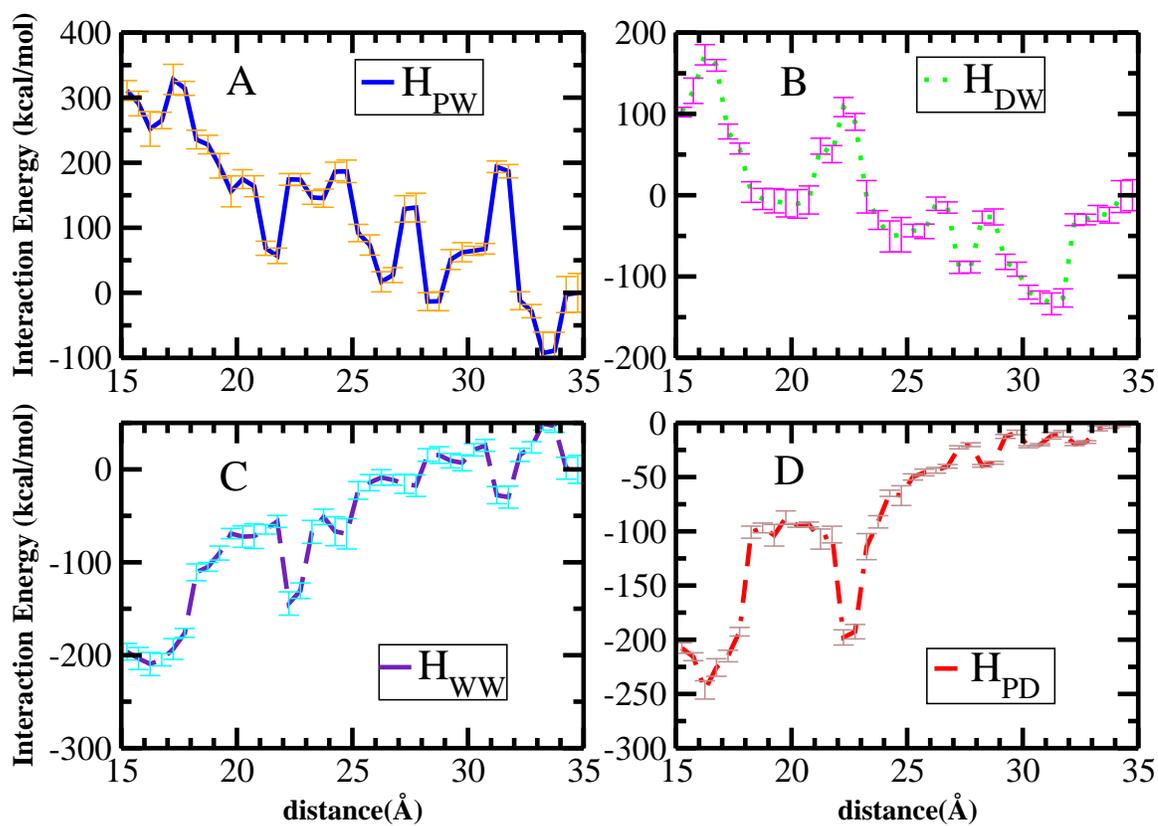


Figure 6.7: Decomposed interaction energies: (A) protein-water H_{PW} (B) domain-water H_{DW} (C) water-water H_{WW} (D) protein-domain H_{PD} .

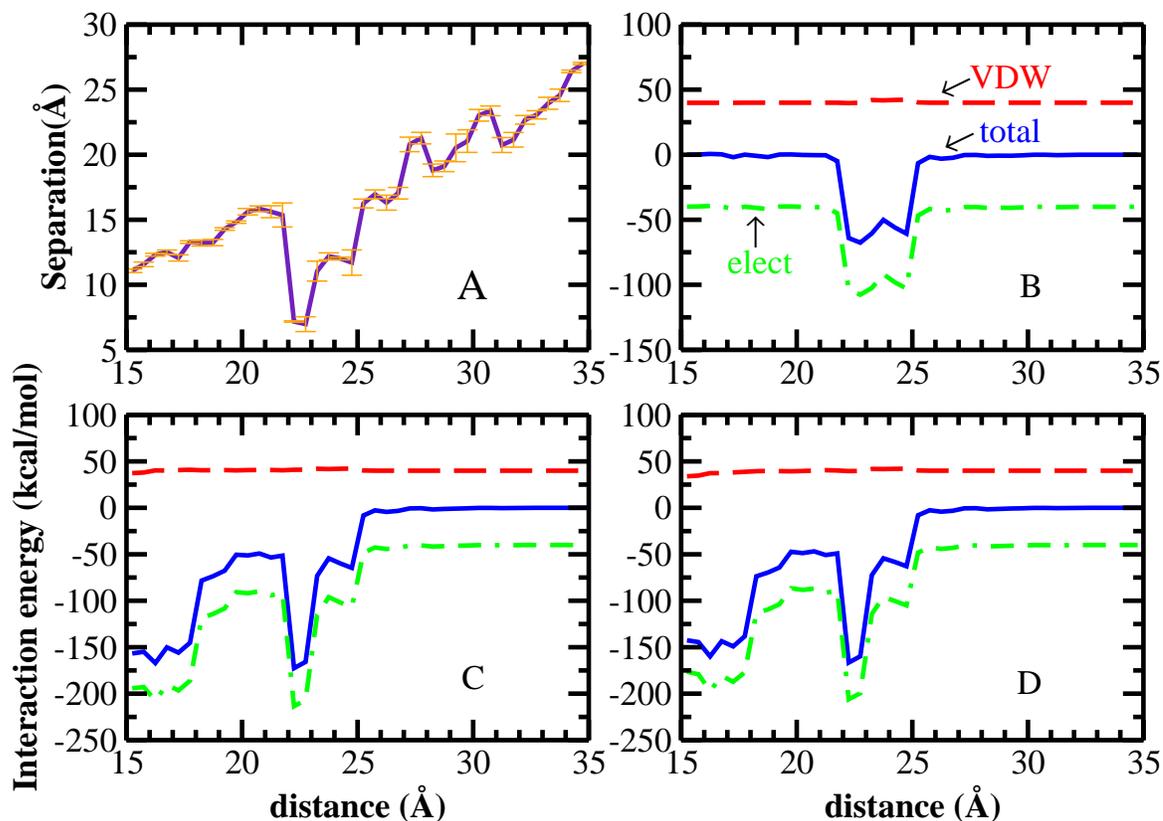


Figure 6.8: (A) Average separation between center of mass of R74 on ubiquitin and center of mass of E5 on the UIM as a function of reaction coordinate (B) Interaction energy between R74 and E5 as a function of reaction coordinate. (C) Interaction energy between R42, R72, R74 on the ubiquitin and E3, E5, E6, E7 on the UIM as a function of reaction coordinate (D) Interaction energy between R42, R72, R74 on the ubiquitin and all the residues on the UIM. In all cases, the VDW component is shifted by 40 kcal/mol and the electrostatic component is shifted by -40 kcal/mol for clarity.

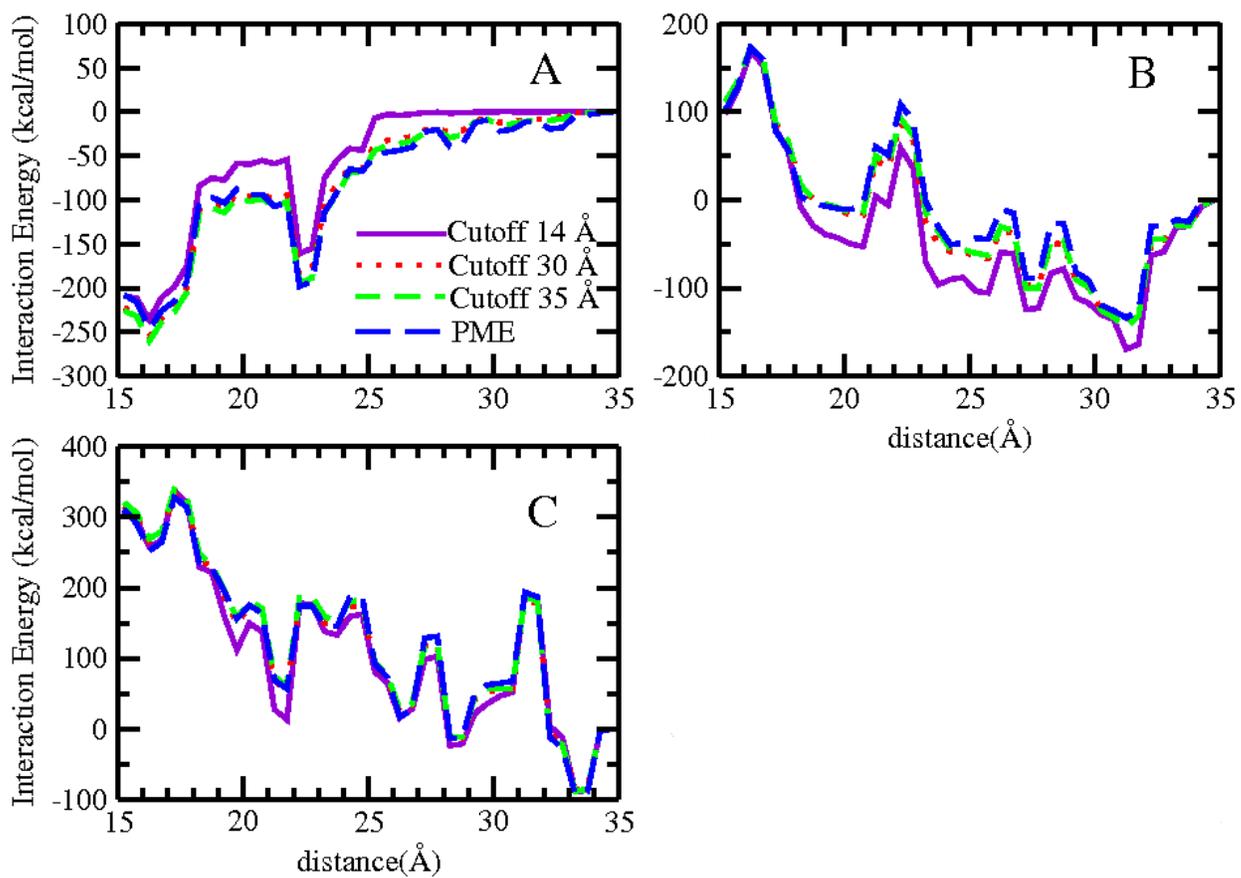


Figure 6.9: Comparison of interaction energies between different components using PME and different non-bonded cutoff values (A) Protein-domain interaction energy (B) Domain-water interaction energy (C) Protein-water interaction energy.

the reaction coordinate is the difference between the forces on the two protein centers of mass projected onto the reaction coordinate of interest. Converting from the collective variable reaction coordinate (relative distance along z-axis between the two centers of mass) to the absolute Cartesian z-coordinate, one obtains the relation that the component contribution along the reaction coordinate is simply the average force on one member of the two proteins with the other considered being fixed at the origin of the coordinate system; here we focus on the UIM domain. The potential of mean force contributions from the water, ubiquitin, counter ions, and the UIM complex itself is expressed as:

$$\Delta W = - \int \langle F_{ion}(\xi_0) \rangle d\xi_0 - \int \langle F_{wat}(\xi_0) \rangle d\xi_0 - \int \langle F_{pro}(\xi_0) \rangle d\xi_0 - \int \langle F_{UIM}(\xi_0) \rangle d\xi_0 \quad (6.3)$$

where F_{ion} , F_{wat} , F_{pro} and F_{UIM} denote as forces on the UIM domain arising from counter ions, water, protein and domain itself. Among them, the principal contributions come from protein and water, which are shown in Figure 6.10. The error bars here are computed via the same method as for the total PMF. The contribution from ubiquitin is largely negative and favorable for association. The favorable protein-protein interactions between ubiquitin and the UIM indicated by the increasingly negative relative enthalpy behavior in Figure 6.7 are the overriding element since association entails loss of protein conformational, translational, and vibrational entropy, which will be addressed later in the Entropy Analysis of Protein subsection. In contrast, the water contribution is repulsive and destabilizing. Relative to the separated state, the associated state lacks inter-region water molecules, as these are expelled from this region. Consequently, the predominantly electrostatic attractive interactions between polar protein and water draw each protein of the complex away from one another, leading to the destabilizing contribution of water to the PMF. Equivalently, water-protein interactions are lost upon association, giving rise to a destabilizing contribution from water. Previously reported studies of the association processes between hydrophobic

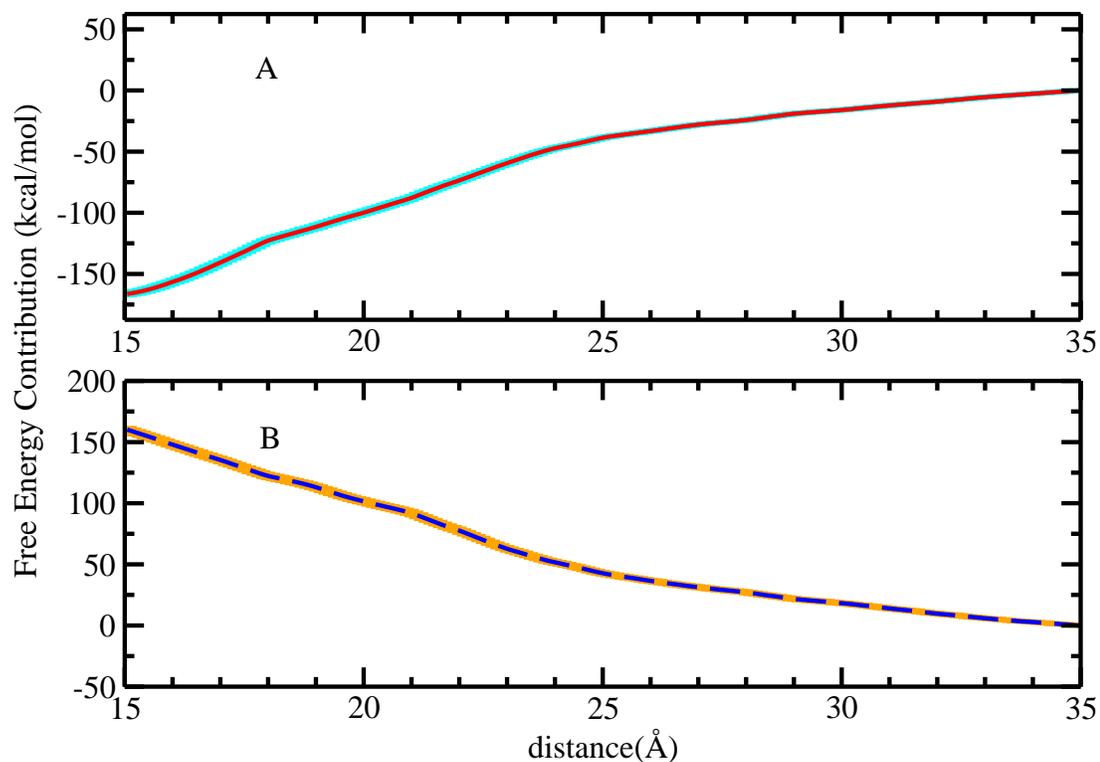


Figure 6.10: PMF contribution from (A) protein (B) water.

species, including carbon nanoparticles in water, [253, 259] graphene in water [94] and carbon nanotubes in water/aqueous ionic solutions, [95] have also demonstrated this feature, though in this case the similarity originates from different underlying physical interactions.

6.3.4 Water Density

Since a discussion of hydrophobic association implicates water, an analysis of spatial distributions of water around solutes is warranted and relevant. Because of the complexity of the chemical and topographical context on protein surfaces, water distribution around them will be complicated. A simplified water number density profile as a function of protein-protein separation only outlines the hydration along one dimension, which is insufficient to depict the full picture. So instead of showing

this in a general way, we decompose the water density distribution around the entire protein, considering the water density "slice by slice" as we move from one protein's center of mass to that of the binding partner. We define a series of thin volume slices perpendicular to the z axis along the positive z direction as shown in Figure 6.11A. Each slice S is defined by $S = [z_0, z_0 + dz]$; thus the width is dz . A water molecule is counted in the slice if the z -component of the water oxygen atom is within the range. We set $dz = 0.5 \text{ \AA}$ and z_0 as $(0, 0.5, 1.0, 1.5, \dots, 33.5, 34.0, 34.5) \text{ \AA}$, so that 70 consecutive slices were constructed. The slice at $z=0$, for example, represents a slice through ubiquitin, and the density profile for this slice would show the cross-sectional profile of water density. For each slice, a water oxygen number density map along x and y is constructed as shown in Figure 6.11. Panel B depicts the water density distribution map for window [32:33] with $z_0 = 0 \text{ \AA}$; this is essentially a cross-section around ubiquitin. For comparison, Panel C shows the water density map around the UIM domain at $z_0 = 32 \text{ \AA}$. The density plotted is the local water density normalized by the bulk value of 0.0334 \AA^{-3} . In regions far from the proteins, the Figures show density of 1, corresponding to the bulk value for TIP3P water. The existence of proteins (and solutes in general) affects water density around the solute, resulting in either lower or higher water density compared with that in bulk regions. Water molecules around these areas were regarded as hydration water. From a traditional point of view, these hydration water molecules are quite rigid, associating specifically with polar and charged residues on the protein surface. [260] Recently, it has been argued that these water molecules may be more flexible than originally believed [261] and the structure and dynamics of water in the vicinity of biomolecules has become an open question. [262] Nevertheless, it is generally accepted that there is a difference in the properties of hydration and bulk water, as has been indicated by both simulation [148] and experiment. [263] Therefore, the release of these hydration waters into the bulk region during the association may be relevant to the thermodynamic signature of the binding event. In view of this, we combine data about the position-dependent water density in the inter-protein region next, obtaining the number of hydration water molecules in this region as a function of protein-protein

separation. From the water density map of Figure 6.11B and C, water density at bulk region corresponds to a value ranging from 0.8 to 1.2. Therefore, the total number of inter-protein hydration water could be obtained by integration of water density that is lower than 0.8 or higher than 1.2 over the desired space as domain approaching the ubiquitin, which is expressed in Equation 6.4 and Equation 6.5.

$$N_{water} = \int_0^d \int_{-30}^{30} \int_{-30}^{30} \rho_c(x, y, z) \quad (6.4)$$

$$\rho_c(x, y, z) = \begin{cases} 0 & 0.8 \leq \rho(x, y, z) \leq 1.2 \\ \rho(x, y, z) & \rho(x, y, z) > 0.8 \text{ or } \rho(x, y, z) < 1.2 \end{cases} \quad (6.5)$$

where $\rho(x,y,z)$ is the water density at position (x,y,z) ; $\rho_c(x, y, z)$ represents the corresponding hydration water density at this position; and N_{water} is the inter-protein hydration water number. Alternatively, to consider only hydration waters, one could combine data from analysis of radial probabilities of water around the protein surface and consider only water molecules within a certain distance from, say, protein heavy atoms. For the qualitative arguments we pursue here, our approach appears sufficient. N_{water} was monitored as a function of distance in Figure 6.11D. As the proteins move toward each other from separated state, initially there is little relative variation in the number of hydration water, which is indicative of a release of bulk-like water in the middle of inter-protein region that is far away from each protein at large separations. As the UIM domain comes near the protein, N_{water} gradually decreases starting from a separation less than 22 Å, which is an indication of the beginning of hydration water release. This separation distance coincides with the increase of enthalpic component and decrease of the entropic components as $d < 22$ Å in Figure 6.6. These trends are in agreement with the traditional explanations of hydrophobic interactions which is based on the argument that there will be an entropic gain by expelling the hydration water molecules from a more rigid environment which is close to the hydrophobic region to the less-ordered bulk. However, in this case, we observe that the release of hydrating

waters, which may be associating with the protein, also give rise to large stabilizing entropic factors, resulting, in a fashion similar to hydrophobic association, to a free energy minimum. We acknowledge that this behavior may be dependent on specific proteins and the nature of their binding interfaces; however, in this case of a system widely held to associate via hydrophobic interactions, the thermodynamic signature, at a coarse resolution, appears very much like that for true hydrophobic interaction.

6.3.5 Free Energetics and UIM Helix Orientation Relative to Ubiquitin

Further exploration of free energetics of model, semi-rigid body association between ubiquitin and restrained UIM with different orientations will be discussed in this section. Free energy profiles are shown in Figure 6.12B, C, D and E. For each orientation, the profiles are calculated by using the ABF approach. Several orientations of the UIM helix (relative to a single orientation of ubiquitin), including 0° , -30° , 30° , -60° , 60° and -90° , exhibit a PMF minimum. The experimental NMR orientation is associated with the largest PMF minimum; this is an interesting, though admittedly anecdotal demonstration of the surprising accuracy of modern force fields in recapitulating detailed molecular interactions that define protein-protein interactions. Further similar studies are warranted and ongoing. Furthermore, we observe no barriers to association for the NMR orientation, as discussed previously. Several non-native orientations are associated with slight barriers. Figure 6.12A shows a 3-dimensional representation of the individual PMF's. Interestingly, the present results show that the NMR structure indeed is the most stable of the orientations probed. We summarized the outcomes of contact state distances $\xi_{contact}$, association free energies ΔG and differences of association free energies between each orientation and 0° orientation $\Delta\Delta G$ in Table 6.3. For 0° , -30° , 30° and 60° , ΔG is lower than -10 kcal/mol and favorable. For 120° , -120° , 150° , -150° and 180° , ΔG is only around -1 kcal/mol; with consideration of the uncertainty (around 0.5 kcal/mol), we suggest that in actuality, there is no free energy benefit for the UIM association with ubiquitin under these orientations. The dramatic difference in association free energy is consistent with the fact that the helix domain is

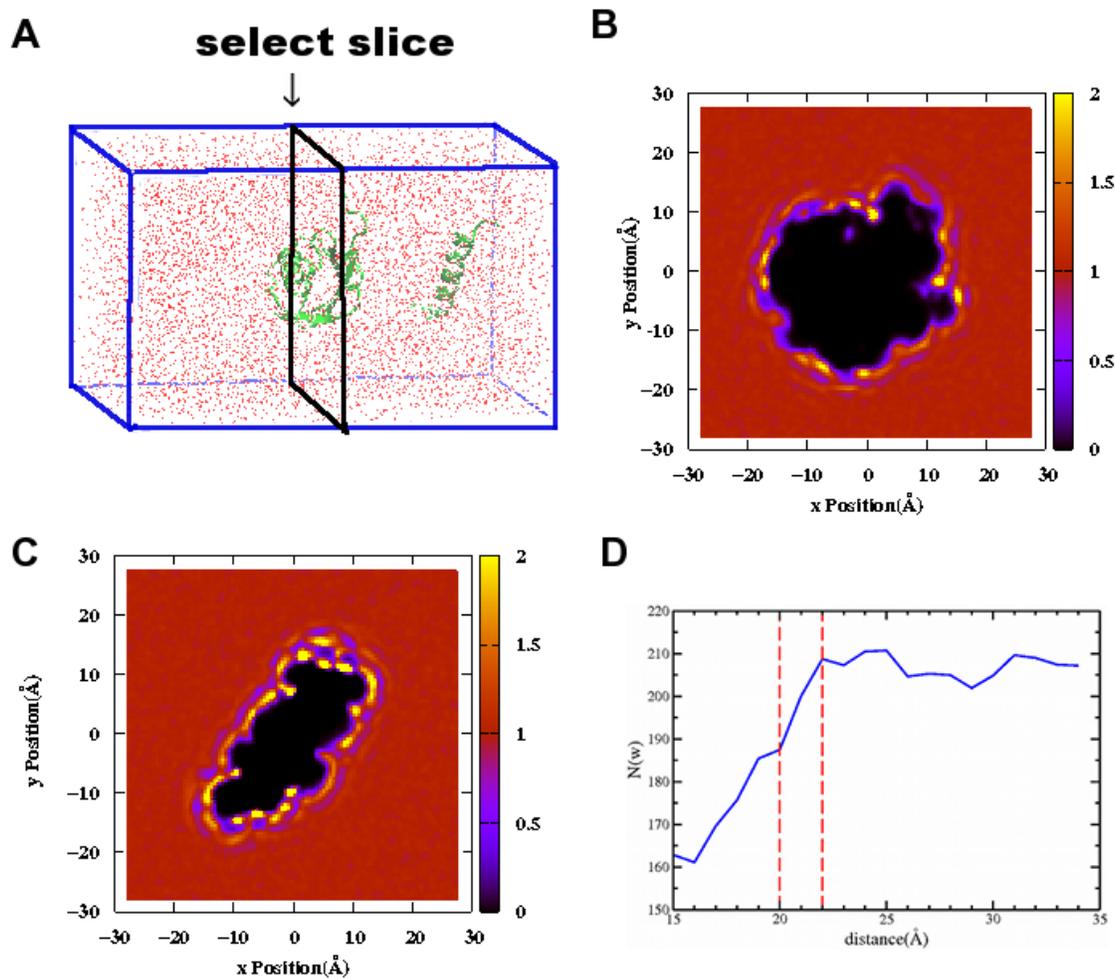


Figure 6.11: Water density distribution maps for a select slice along the z direction (A) representation of the select slice to consider water density (B) water density distribution map for $z_0 = 0 \text{ \AA}$ at window [32:33] (C) water density distribution map for $z_0 = 32 \text{ \AA}$ at window [32:33]. The density is normalized such that $\rho = 1$ corresponds to the bulk water density of 0.0334 \AA^{-3} .

<i>Orientation</i> (°)	$\xi_{contact}$ (Å)	ΔG (kcal/mol)	$\Delta\Delta G$ (kcal/mol)
0	15.75	-15.97(0.58)	0
30	16.99	-10.13	5.84
-30	16.01	-13.03(0.62)	2.94
60	16.75	-11.73	4.24
-60	18.59	-6.13(0.55)	9.84
90	15.49	-4.54	11.43
-90	17.19	-5.22(0.61)	10.75
120	18.51	-1.35	14.62
-120	21.55	-1.05(0.66)	14.92
150	21.99	-0.94	15.03
-150	20.95	-1.03	14.94
180	20.45	-1.22(0.52)	14.75

Table 6.3: Association free energies for ubiquitin with UIM at different orientations.

amphipathic, with a set of hydrophobic residues including L8, L9, A12, L13, L15 and L17 along one side. It is widely held that the hydrophobic residues form a principal binding patch, engaging the L8-I44-H68-V70 hydrophobic patch of ubiquitin upon association. Figure 6.13 shows the relative positions of the patches involved in binding at 0° orientation (Panel A) and -120° orientation (Panel B). Lower association free energy corresponds to the case where two hydrophobic patches face each other, and higher association free energy corresponds to the case that they are facing away from each other. This is an indication that a hydrophobic effect, intimate and specific interaction of currently-defined hydrophobic moieties, is important for the ubiquitin-UIM association.

To further explore the free energetic differences in association with various orientations, decomposition of the total free energy into the enthalpic and entropic components is required. The decomposition results are shown in Table 6.4. Where we observe the most negative (stabilizing) association free energies, as with the 0° and -30° orientations, the entropic component is the dominant contribution to the association PMF, with a value around -25 kcal/mol. In contrast, the enthalpic component makes an unfavorable contribution of about 10 kcal/mol. The trend is reversed for the cases

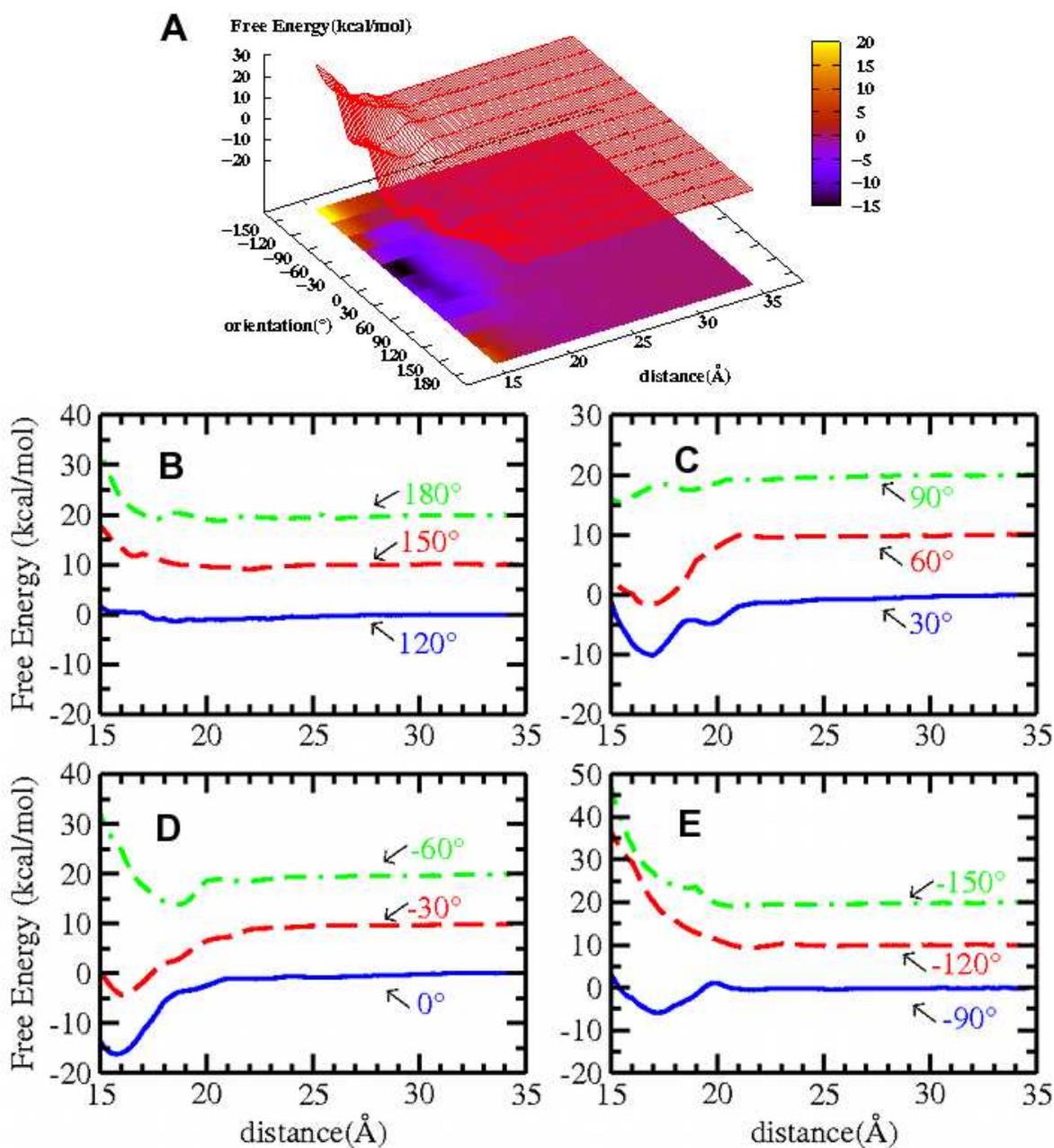


Figure 6.12: Free energies of restrained UIM with different orientations association with ubiquitin. (A) Mapping the free energies of association at various orientations and separations; (B),(C),(D),(E) Free energy profiles for all the orientations. For clarity, red lines are shifted by 10 kcal/mol; green lines are shifted by 20 kcal/mol.

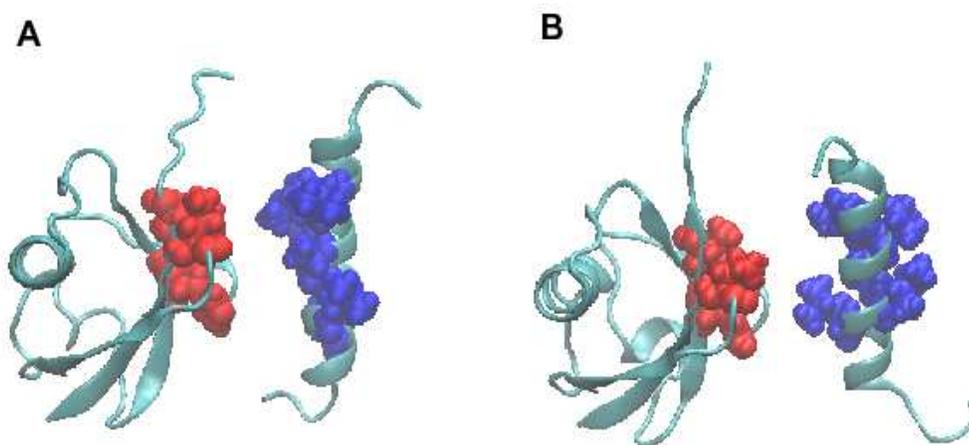


Figure 6.13: Representation of hydrophobic residues involved in ubiquitin-UIM binding (A) UIM restrained at 0° orientation (B) UIM restrained at -120° orientation. Color scheme: red, hydrophobic residues L8, I44, H68 and V70 on ubiquitin; blue, hydrophobic residues L8, I9, A12, I13, L15 on UIM.

with less free energetically favorable orientations such as -120° and 180° orientations. In these cases, the free energy advantage comes from the enthalpic contribution, which is slightly negative; correspondingly, the entropic contribution is slightly positive. One source of the change in thermodynamic character is related to differences in hydrophobicity of the interface of UIM at different orientations. When the hydrophobic face is toward the ubiquitin for association, we observe the signature of an entropically-driven process; when the hydrophilic face is toward the ubiquitin, the signature changes, with enthalpy having a major stabilization role. The preferred orientation for the UIM association is derived from the hydrophobic character of the interface; this is consistent with the idea ubiquitin association with its binding domains, at least in the case of UIM, is hydrophobically mediated.

<i>Orientation</i> (°)	ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)
0	-15.97(0.58)	10.72(2.80)	-26.69(3.38)
-30	-13.03(0.62)	10.88(2.60)	-23.91(3.22)
-60	-6.13(0.55)	10.08(2.78)	-16.21(3.33)
-90	-5.22(0.61)	18.78(3.64)	-24.00(4.25)
-120	-1.05(0.66)	-2.98(3.70)	1.93(4.36)
180	-1.22(0.52)	-3.80(2.44)	2.58(2.96)

Table 6.4: Enthalpic contribution and entropic contribution for different orientations.

6.3.6 Entropy Analysis of Protein

Entropy change, ΔS , associated with protein-protein binding in aqueous solution can be expressed as the sum of two terms, contribution from solvent and contribution from protein translational, vibrational, and conformational changes. Considering solvent, there will be a decrease of solvent accessible surface area upon protein association, resulting in release of confined solvent molecules to bulk. This contributes favorably to the total entropy of interaction. For the protein contribution, it is generally accepted that when two protein binding partners “merge” into one complex, conformational degrees of freedom of proteins are lost, giving rise to unfavorable entropy changes. However, it has also been mentioned that in some cases, one protein may increase the number of conformational degrees of freedom upon ligand binding. [264] Proteins conformational entropy change (ΔS_{conf}) can be derived by subtracting the free ubiquitin entropy (S^P) and free UIM domain entropy (S^D) from the protein complex (S^C). For each species, the entropy can be further decomposed into translational, rotational and vibrational contributions. Among them, the most complicated part is from the vibrational component. The vibrational entropy was computed from independent free simulations for protein, domain and complex based on the quasiharmonic approximation. [265] Quasiharmonic analysis calculates vibrational entropy based on snapshots from a simulation trajectory. One problem is the convergence of the vibrational entropy based on this approach, which has been pointed out by Harris et al.[266] One way to improve the convergence arising from sampling issue is to extrapolate the entropy at

infinite simulation time by using a function as has been shown in Equation 6.6.

$$S(t) = S_\infty - \frac{A}{t^n} \quad (6.6)$$

$S(t)$ represents the vibrational entropy at simulation time t , S_∞ denotes the extrapolated entropy at infinite sampling time by fitting the parameters A and n . $S(t)$ was obtained by analyzing the dynamics trajectory using the quasiharmonic method in the VIBRAN module of CHARMM. Figure 6.14 shows the evolution of vibrational entropy for the free proteins and complex structures. For clarity, the fitted curve was shifted up 2 units as shown by blue lines. The slopes of the fitted curves are close to zero after 15ns; we consider the S_∞ in each case as the final vibrational entropy involved in the process we are modeling. The fitted parameter values are shown in Table 6.5. Additionally, the translational and rotational entropies are computed using the following equations.[267]

$$S_{trans}^i = \frac{5}{2}k_B + k_B \ln\left[\frac{1}{C_0} \left(\frac{2\pi m^i k_B T}{h^2}\right)^{3/2}\right] \quad (6.7)$$

$$S_{rot}^i = \frac{3}{2}k_B + k_B \ln\left[\frac{(\pi I_x^i I_y^i I_z^i)^{1/2}}{\sigma^i} \left(\frac{8\pi^2 k_B T}{h^2}\right)^{3/2}\right] \quad (6.8)$$

where m is the mass and I_x , I_y and I_z are the moments of inertia and σ is the symmetry factor. The results of decoupled entropies from translational, rotational and vibrational contributions of each species are listed in Table 6.6. Here, we consider two complex states, one corresponds to UIM with 0° in the complex and another one corresponds to the UIM with -30° . Entropy changes mostly derive from translational and rotational parts, with a small (< 1 cal/mol/K) contribution from vibrational entropy change. The unfavorable protein configurational entropy changes (around -93 cal/mol/K) upon association suggest that the release of water molecules during the binding event provides a significant amount of favorable entropy contribution that is sufficient to compensate the "freezing" of protein degrees of freedom upon binding. We also note that $\Delta\Delta S_{conf}$, which is the difference between ΔS_{com-30} and ΔS_{com0} ,

<i>System</i>	S_∞ (cal/mol/K)	A	n
complex 0°	6.363(0.065)	3.235(0.052)	0.660(0.038)
complex -30°	6.637(0.103)	3.458(0.075)	0.564(0.042)
UIM	2.010(0.040)	0.895(0.031)	0.445(0.042)
ubiquitin	4.796(0.021)	2.225(0.023)	0.778(0.025)

Table 6.5: Fitting constants for entropy extrapolation. The numbers in the brackets represent asymptotic standard errors upon fitting.

<i>System</i>	S_{tran} (cal/mol/K)	S_{rot} (cal/mol/K)	S_{vib} (cal/mol/K)	S_{total} (cal/mol/K)
complex 0°	53.860	53.163(0.323)	6.363(0.065)	113.386(0.388)
complex -30°	53.860	53.129(0.475)	6.637(0.103)	113.626(0.578)
UIM	49.700	46.005(1.214)	2.010(0.040)	97.715(1.254)
ubiquitin	53.011	51.425(0.100)	4.796(0.021)	109.232(0.121)
ΔS_{com0}	-48.851	-44.267(1.637)	-0.443(0.126)	-93.561(1.763)
ΔS_{com-30}	-48.851	-44.301(1.789)	-0.169(0.164)	-93.321(1.953)

Table 6.6: Decoupled entropies from translational, rotational and vibrational contributions.

is around 0.24 (cal/mol/K). This results in a -0.072 (kcal/mol) of configurational entropic component difference between the two orientations. Since we use a rigid model of domain in the simulation to evaluate the free energy differences among various orientations in Table 6.3, the configurational entropic component difference is not included here. For the free energy difference between 0° and -30° orientation, -0.072 (kcal/mol) is a negligible value compared with 2.94 (kcal/mol) shown in Table 6.3. Thus, we have some confidence in comparing relative orientational free energetics based solely on the data of $\Delta\Delta G$ value as discussed in the previous section.

6.4 Conclusion

To investigate the thermodynamic signature of hydrophobic association within the context of a biochemical system defined by ubiquitin and one of its many binding partners, UIM, widely considered to associated through predominantly hydrophobic interactions, we present calculations of the potential of mean force for association in solution using ABF sampling coupled with atomistic molecular dynamics simulations.

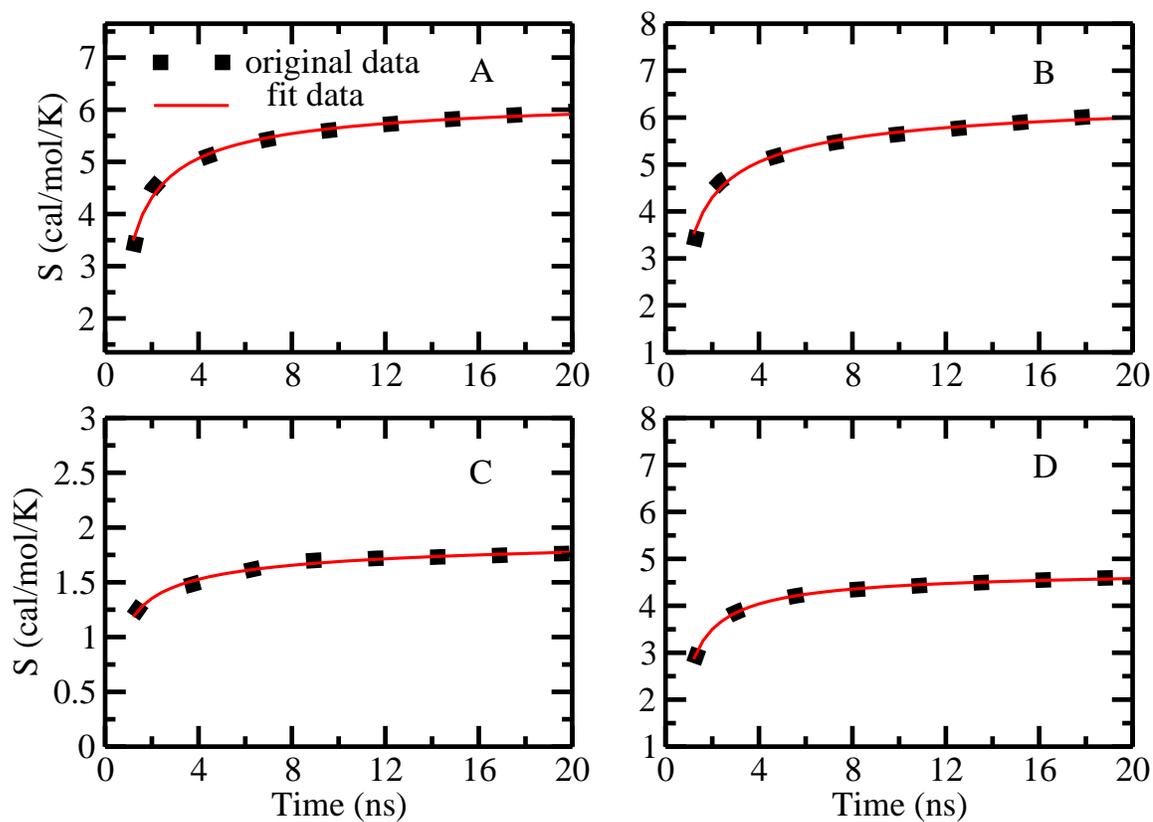


Figure 6.14: Evolution of vibrational entropy for (A) Ubiquitin-UIM complex with 0° orientation (B) Ubiquitin-UIM complex with -30° orientation (C) UIM (D) Ubiquitin.

We find that at a low resolution, the thermodynamics of association of this system, treated as the binding of two semi-rigid bodies, exhibits characteristics of the canonical hydrophobic interaction. There is a large entropic stabilization component to the overall association, with a concurrent destabilizing enthalpic contribution. The entropic component originates from the release of hydration water molecules around the proteins. The enthalpic contribution, though destabilizing in this system as in the case of purely hydrophobic solutes, is a result of the loss of specific protein-water interactions within the hydration shells. For purely hydrophobic solute systems, destabilizing enthalpic contributions to association free energy arise from the loss of strong water-water interactions that arise due to the need for solvent structuring in the vicinity of the hydrophobe. By computing the density profile of hydration waters as a function of protein-protein separation, we find that the onset of entropic stabilization coincides with the separation distance at which significant water release is observed. This is again consistent with a hydrophobic like process, but the origins of the effect are different. Our results indicate that association of ubiquitin and UIM along a reaction coordinate representing rigid-body translation along the axis between centers of mass of the two proteins is entropically dominated. This is consistent with the effective binding region (amino acids sequence Leu8-Ile44-Val70) being on the surface of ubiquitin as opposed to being a buried site. The release of degrees of freedom associated with solvating waters in the vicinity of the hydrophobic patch of ubiquitin (and the UIM) gives rise to the canonical entropic signature of hydrophobic association in this system. This is consistent with literature reports indicating that well-solvated binding sites/regions are involved with association processes with signatures of being entropically dominated. We stress that for other scenarios, this signature may be different. For this reason, broader studies are warranted and ongoing. We finally note that our results are somewhat in accord with recent experiments indicating that the association of SUMO-1 with binding partners (RanBP2/Nup358)[268] is both entropically and enthalpically driven. This is consistent with the present observations from simulation in that whereas the ubiquitin binding domain is predominantly located on the surface of

the protein, the SUMO-1 binding domain is somewhat more buried; the interaction domain of SUMO-1 forms a groove dotted with aromatic and hydrophobic residues including histidine, isoleucine, tyrosine, phenylalanine, leucine, and valine. This would suggest, in light of the experimental observations discussed in the Introduction that show how binding site hydration/solvation affects the dominant contribution to the binding process (enthalpy versus entropy), that SUMO-1 would present an association signature that is intermediate between a system with a severely buried binding region and a system with a purely surface (highly solvent accessible) binding region. Since our system, ubiquitin-UIM, is a surface binding region composed of a hydrophobic motif, that we obtain entropy as the dominant driving force for association is consistent with the several literature observations discussed[269].

We further consider the free energy of association as a function of the relative orientation of the UIM binding domain and find that the force field based approach used here is able to recapitulate the global free energy stability of the experimental NMR structure (within the constraints of the semi-rigid modeling approach applied). It is remarkable that using such a simple method allows us to observe that the NMR structure is indeed associated with a free energy minimum as computed by the force field.

Chapter 7

CONCLUSION

7.1 Summary of Key Results

This dissertation discussed the approach to identify the effective hydrophobic patch of single protein, which could serve as the indication of the possible binding sites for protein extensively involved in the hydrophobic association. With the identified effective hydrophobic patch, details were considered regarding the ion-specific effect around the hydrophobic surface region. From a simple case to more complicated case, my study further extended to the interactions between effective hydrophobic protein surface and molecular ion Gdm^+ , then interactions between effective hydrophobic protein surface and small peptide.

Initially, a method that exploits water network percolation behavior in the first solvation shell of small proteins in order to predict clusters of residues potentially involved in binding interactions. The select proteins in this study involving the hydrophobic characteristics of binding interfaces. The effective hydrophobicity of a residue that is dictated by the character of neighboring residues as well as local water were evaluated in this study at the critical hydration level. At the specific hydration level, water molecules could form the hydrogen-bonded water networks around the effective hydrophilic region of the protein while leaving the effective hydrophobic region to be uncovered. In order to identify such a critical hydration level, ubiquitin was applied as an example. A protocol was developed by finding the critical hydration level where a single distribution of the probability of largest cluster transitions to a bimodal distribution. This point gives rise to a networked, spanning water cluster which effectively seeks less hydrophobic regions of the protein surface in order to maximize favorable

interactions. In this case, water number density around various residues on the protein surface was computed. Finally, single linkage clustering of the low hydrated residues was applied to identify a structure continuous patch. Compared our predictions of binding patch residues to those from automated servers (SPPIDER, InterproSurf and meta-PPISP). We find that the current method is competitive in terms of the average accuracy 60% and the average coverage 75% across the series of proteins studied. Though simplistic in principle and spirit, this method is able to predict with significant accuracy and coverage the binding interaction residues for a series of small proteins. Besides, these results are consistent with previous studies that consider water density fluctuation based approaches for characterizing local hydrophobicity of protein surface regions.

Further computation of the free energetics association of single Cl^-/I^- with the effective hydrophobic patch region of protein reveals that I^- displays a larger extent of interfacial stability compared with Cl^- . This is consistent with the trend showing around the liquid-vapor interface region, which indicates an inherent connection between the ideally hydrophobic aqueous liquid-vapor interface and a more somewhat more realistic, and certainly more complex, aqueous protein hydrophobic interface. Furthermore, we find that the more surface stable I^- induces significantly larger interface fluctuations on approach to the interface compared to the smaller, more charge-dense chloride. This is again in keeping with observations at the aqueous liquid-vapor interface. These behaviors approaching the hydrophobic interface, are related to the coupling of local hydration water in the vicinity of the protein with the hydration water around the individual anions; specifically, the differential ability of the water environments to couple with one another in the case of Cl^- and I^- leads to the specific-ion behavior as it is related to induced interfacial fluctuations. Water molecules in the hydration shells of I^- are shown to be more dynamic and less persistent compared to those in proximity to Cl^- . When approaching the interfacial region, coupling of local solvent around anions with solvent near the interface leads to different perturbations of the interface by the two anions, and thus different contributions to interface height

fluctuations, and ultimately surface stability via contributions from interfacial entropy arising from surface fluctuations correlations. Therefore, building upon the insights gained from the previous study of specific ion behaviors at aqueous liquid-vapor interfaces and current study of specific ion behaviors at aqueous hydrophobic protein interface, we have presented here a discussion regarding the unique fluctuation inducing properties of two anions for which the degree of induced interfacial fluctuations correlates with stability at the interface. In another case, as anion approaching hydrophilic interface of the protein, we observe that both anions display similar behaviors in terms of surface stability and induced interface fluctuations. These differences offer a view of the anions as having different characters in different contexts. Where strong local interactions are not dominant, as in the case of hydrophobic surfaces that lead to higher fluctuations in general, the anions tend to differentiate themselves based on their hydrophobicity; the large, less charge-dense I^- has a higher propensity to associate with hydrophobic regions due to its inherent higher hydrophobicity. The smaller, more charge-dense, less hydrophobic Cl^- is not a stable at a hydrophobic interface.

Then we continue to explore and demonstrate a connection between interfacial stability and induced interfacial fluctuations as denaturant solutes including Gdm^+ and urea approach ostensibly hydrophobic interfaces of protein HFBII. Particularly, we considered the contributions to the total free energy arising from two relative orientations of the solute, the parallel and perpendicular orientations as defined relative to the surface of the protein. For both Gdm^+ and urea, the parallel configurations are associated with stronger free energy minima compared to configurations where the solutes approach in perpendicular configuration. Furthermore, there is a correlative behavior between solute orientation and free energy stability with surface stable parallel configurations inducing larger extent of interfacial height fluctuation compared with perpendicular configurations. This trends is also observed previously around the liquid-vapor interface. The relation between solute orientation and induced fluctuations is related to the nature of the solvation shells of the solute presented towards the interface upon approach of the solute. In the case of Gdm^+ approaching the interface

in a parallel orientation, the solvation shell presented is a more malleable one, where the solvent is more labile and free to rearrange like the character around I^- . This leads to greater solvent density fluctuations and hence, higher interfacial induced fluctuations. In the case of the perpendicular orientations of Gdm^+ and urea, the tighter hydrogen bonding patterns of water create a more rigid, well-defined solvation shell that is not easily disrupted similar to high charge-dense Cl^- . This translates to lower solvent density fluctuations, and hence lower induced fluctuations. The present results are thus consistent with single anion behavior and provides yet another example of the relation of hydrophobic effect, solvent fluctuations, and interfacial stability. This relationship appears to be common across a series of atomic and molecular species, as well as encompassing charged, polar, and non-polar characteristics of the solutes considered. These observations suggest that molecular ions, such as Gdm^+ , as well as polar molecules with heterogeneous charge distributions inherently have built into them regions of high and low charge density. The dependence of local solvation structure on this heterogeneous charge density is to a large extent involved in determining the propensities of the modalities involved in specific association of molecules with specific types of interfaces.

In the last part of the dissertation, we are attempted to investigate the thermodynamic signature of hydrophobic association within the context of a biochemical system defined by ubiquitin and one of its binding partners UIM. With application of ABF sampling approach, we computed the potential of mean force for association in solution between the two proteins. We find that the thermodynamics of association of this system, treated as the binding of two semi-rigid bodies, exhibits characteristics of the canonical hydrophobic interaction. There is a large entropic stabilization component to the overall association, with a concurrent destabilizing enthalpic contribution. The entropic component originates from the release of hydration water molecules around the proteins. The enthalpic contribution, though destabilizing in this system as in the case of purely hydrophobic solutes, is a result of the loss of specific protein-water interactions within the hydration shells. By computing the density profile of hydration

waters as a function of protein-protein separation, we find that the onset of entropic stabilization coincides with the separation distance at which significant water release is observed. This is again consistent with a hydrophobic like process. Our results indicate that association of ubiquitin and UIM along a reaction coordinate representing rigid-body translation along the axis between centers of mass of the two proteins is entropically dominated. This is consistent with the effective binding region involving residues Leu8-Ile44-Val70 being on the surface of ubiquitin. The release of degrees of freedom associated with solvating waters in the vicinity of the hydrophobic patch of ubiquitin and the UIM gives rise to the canonical entropic signature of hydrophobic association in this system. All these results are consistent with literature reports indicating that well-solvated binding sites/regions are involved with association processes with signatures of being entropically dominated. We further consider the free energy of association as a function of the relative orientation of the UIM binding domain and find that the force field based approach used here is able to recapitulate the global free energy stability of the experimental NMR structure. It is remarkable that using such a simple method allows us to observe that the NMR structure is indeed associated with a free energy minimum as computed by the force field.

7.2 Future Work

In this dissertation, an approach was developed to identify the effective hydrophobic patch around the protein surface based on the local water number density. In this process, a critical hydration level at which the water molecules display an obvious distinction in the coverage of effective hydrophobic region and effective hydrophilic region has to be determined first. In our current approach, in order to determine the critical hydration level for each protein, we actually explore the probability distribution of the size of the largest hydrogen-bonded water network in the system in a range of unfully solvated hydration levels. Based on the distribution characters, a critical hydration level can be located. In this process, comparison and test are required in order to determine the critical hydration level for each protein and several simulations

with protein solvated in different amount of water has to be performed. If an approach that would directly link the critical hydration level with some properties can be identified, then essentially only one simulation with single protein solvated at the critical hydration level was required.

Therefore, the next step in our research is to seek a property that could directly point into the critical hydration level. We would like to consider the chemical potential since the binding ability of water molecules around the protein surface essentially is determined by the chemical potential. With increasing of the hydration level around the protein surface, the average chemical potential around the protein surface should be perturbed. Initially in a limited solvated case, the effective hydrophilic region around protein surface can not be fully covered. Therefore, with increasing the hydration by adding the number of solvent molecules into the system, more effective hydrophilic region around the protein surface would be wetted. The favorable interaction between the water molecules and the effective hydrophilic region around the protein surface resulting in a negative value of chemical potential. Overall, the average chemical potential around the protein surface should decrease. As the hydration coverage go beyond the critical hydration level, water molecules would cover the effective hydrophobic region around protein surface, which results in an unfavorable interaction between water molecules and the effective hydrophobic region around protein surface and the average chemical potential around the protein surface should increase. In this sense, at the critical hydration level around the protein surface, the average chemical potential should display a minimum value. If this value could be normalized in a way that it is independent of the proteins we investigated, then such a normalized value for chemical potential can serve as the criterion for judging the critical hydration level for various types of proteins. Further, an analytical form between the normalized chemical potential and the number of water molecules in a particular solvated protein system should be established. With the criterion, the number of water molecules solvated the protein corresponding to the critical hydration level could be computed in this case. From such an approach, it avoids the simulations of hydrated proteins with different

number of water molecules in several systems. Besides, it could define the critical hydration level in a mathematic rigorous way. In our previous approach, the protein was total frozen during the whole simulation in order to obtain the hydrogen-bonded water network distribution. With this proposed normalized chemical potential approach, hydrogen-bonded water network analysis could be skipped and a total flexible protein simulation condition can be realized.

Besides using the water density to scale the effective hydrophobicity around the protein surface, it has been extensively discussed that fluctuations of water density around the effective hydrophobic region and effective hydrophilic region are quite different. Around the effective hydrophobic region, the interaction between water molecules and the protein surface is weaker. Therefore, the nature of hydration shells around hydrophobic surfaces are softer and more flickering than that of hydrophilic ones, showing a larger extent of density fluctuation. This enhanced fluctuations are reflected by the broad probability distributions of water number density around hydrophobic surfaces compared with the bulk solution and hydrophilic surfaces. Moreover, the enhanced density fluctuations around hydrophobic surfaces could further be characterized by more compressible hydration shells and increased cavity formation. Instead of considering water density fluctuation, in this dissertation, we also considered the protein-solvent interface height fluctuations. A coarse-grain water density field was first constructed in space, then the instaneous protein-solvent interface can be identified and the interface height fluctuations can be computed. Although conceptually different from water density fluctuations, these studies reflect the similar trend that around effective hydrophobic region, it displays a larger extent of fluctuation. From these studies, it suggest that in a fully solvated protein system, based on the water density fluctuation, it can show an obvious distinction around the effective hydrophobic region and effective hydrophilic region. Although it is possible to scale the effective hydrophobicity around the protein surface in such an approach, the fully hydrated condition makes this process computationally expensive. A mainly computational cost in this process

arising from the computation of the interactions between water and water in bulk regions where the water molecules are not relevant for the concerned problem. In this sense, I propose a way to reduce the computational cost by replacing these irrelevant all-atom water with the coarse-grained water. Such an adaptive resolution simulation approach has been previously applied in a simulation of pure bulk water system. [270] Usually, between the all-atom water region and the coarse-grain water region, there is a hybrid region where the interchange between all-atom water and coarse-grain water would happen. From such an approach, around the protein surface region, the water molecules are taken into account explicitly, which could give the detailed information about the density fluctuation around different regions on protein surface to characterize its effective hydrophobicity. In the bulk region, where water molecules are not relevant, coarse-grained water are applied to reduce the computational cost.

BIBLIOGRAPHY

- [1] H. Zhou and S. Qin. Interaction-site prediction for protein complexes: a critical assessment. *Bioinformatics*, 23:2203–2209, 2007.
- [2] A. Ben-Naim. Hydrophobic interactions. *Plenum Press: New York*, 1980.
- [3] C. Tanford. The hydrophobic effect and the organization of living matter. *Science*, 200:1012–1018, 1978.
- [4] L. Maibaum, A. Dinner, and D. Chandler. Micelle formation and the hydrophobic effect. *The Journal of Physical Chemistry B*, 108:6778–6781, 2004.
- [5] B. Berne, J. Weeks, and R. Zhou. Dewetting and hydrophobic interaction in physical and biological systems. *Annual Review of Physical Chemistry*, 60:85–103, 2009.
- [6] Y. Levy and J. Onuchic. Water mediation in protein folding and molecular recognition. *Annual Review of Biophysics and Biomolecular Structure*, 35:389–415, 2006.
- [7] M. Krone, L. Hua, P. Soto, R. Zhou, B. Berne, and J. Shea. Role of water in mediating the assembly of alzheimer amyloid- β a β 16-22 protofilaments. *Journal of the American Chemical Society*, 130:11066–11072, 2008.
- [8] H. Frank and M. Evans. Free volume and entropy in condensed systems iii. entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *The Journal of Chemical Physics*, 13:507–532, 1945.
- [9] R. Broadbent and G. Neilson. The interatomic structure of argon in water. *The Journal of Chemical Physics*, 100:7543–7547, 1994.
- [10] P. DeJong, G. Neilson, and A. Buckingham. Hydrophobic hydration of methane. *Molecular Physics*, 91:99–103, 1997.
- [11] A. Filipponi, C. Lobban, and J. Finney. Structural determination of the hydrophobic hydration shell of kr. *Physical Review Letters*, 79:1293–1296, 1997.
- [12] D. Bowron, A. Filipponi, M. Roberts, and J. Finney. Hydrophobic hydration and the formation of a clathrate hydrate. *Physical Review Letters*, 81:4164–4167, 1998.

- [13] F. Stillinger. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *Journal of Solution Chemistry*, 2:141–158, 1973.
- [14] R. Pierotti. Aqueous solutions of nonpolar gases. *The Journal of Physical Chemistry*, 69:281–288, 1965.
- [15] D. Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437:640–647, 2005.
- [16] K. Lum, D. Chandler, and J. Weeks. Hydrophobicity at small and large length scales. *The Journal of Physical Chemistry B*, 103:4570–4577, 1999.
- [17] Y. Sun, D. Spellmeyer, D. Pearlman, and P. Kollman. Simulation of the solvation free energies for methane, ethane, and propane and corresponding amino acid dipeptides: A critical test of the bond-pmf correction, a new set of hydrocarbon parameters, and the gas phase-water hydrophobicity scale. *Journal of the American Chemical Society*, 114:6798–6801, 1992.
- [18] L. Pratt. Hydrophobic solvation of nonspherical solutes. *The Journal of Chemical Physics*, 73:3430–3433, 1980.
- [19] D. Huang, P. Geissler, and D. Chandler. Scaling of hydrophobic free energies. *The Journal of Physical Chemistry B*, 105:6704–6709, 2001.
- [20] R. Wolfenden, L. Andersson, P. Cullis, and C. Southgate. Affinities of amino acid side chains for solvent water. *Biochemistry*, 20:849–855, 1981.
- [21] W. Wimley and S. White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology*, 3:842–848, 1996.
- [22] W. Wimley, T Creamer, and S. White. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry*, 1996:5109–5124, 1996.
- [23] W. Wimley and S. White. Hydrophobic interactions of peptides with membrane interfaces. *Biochimica et Biophysica Acta*, 1376:339–352, 1998.
- [24] W. Wimley and S. White. Membrane protein folding and stability: physical principles. *Annual Review of Biophysics and Biomolecular Structure*, 28:319–365, 1999.
- [25] J. Kyte and Doolittle R. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157:105–132, 1982.
- [26] N. Giovambattista, P. Debenedetti, and P. Rossky. Hydration behavior under confinement by nanoscale surfaces with patterned hydrophobicity and hydrophilicity. *The Journal of Physical Chemistry C*, 111:1323–1332, 2007.

- [27] H. Acharya, S. Vembanur, S. Jamadagni, and Garde S. Mapping hydrophobicity at the nanoscale: Applications to heterogeneous surfaces and proteins. *Faraday Discussions*, 146:353–365, 2010.
- [28] S. Jamadagni, R. Godawat, and S. Garde. Hydrophobicity of proteins and interfaces: Insights from density fluctuations. *Annual Review of Chemical and Biomolecular Engineering*, 2:147–171, 2011.
- [29] S. Garde. Hydrophobic interactions in context. *Nature*, 517:277–279, 2015.
- [30] T. Beuming, Y. Che, R. Abel, B. Kim, V. Shanmugasundaram, and W. Sherman. Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins: Structure, Function, and Bioinformatics*, 80:871–883, 2012.
- [31] T. Lazaridis. Inhomogeneous fluid approach to solvation thermodynamics. 1. theory. *The Journal of Physical Chemistry B*, 18:3531–3541, 1998.
- [32] T. Lazaridis. Inhomogeneous fluid approach to solvation thermodynamics. 2. applications to simple fluids. *The Journal of Physical Chemistry B*, 18:3542–3550, 1998.
- [33] R. Remsing and A. Patel. Water density fluctuations relevant to hydrophobic hydration are unaltered by attractions. *The Journal of Chemical Physics*, 142:024502(1–10), 2015.
- [34] A. Patel, Varilly P., and Chandler D. Fluctuations of water near extended hydrophobic and hydrophilic surfaces. *The Journal of Physical Chemistry B*, 114:1632–1637, 2010.
- [35] A. Patel and S. Garde. Efficient method to characterize the context-dependent hydrophobicity of proteins. *The Journal of Physical Chemistry B*, 118:1564–1573, 2014.
- [36] R. Godawat, S. Jamadagni, and S. Garde. Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proceedings of the National Academy of Sciences*, 106:15119–15124, 2009.
- [37] E. Chi, S. Krishnan, T. Randolph, and J. Carpenter. Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Pharmaceutical Research*, 20:1325–1336, 2003.
- [38] M. Record, C. Anderson, and T. Lohman. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Quarterly Reviews of Biophysics*, 11:103–178, 1978.

- [39] S. Timasheff and T. Arakawa. Mechanism of protein precipitation and stabilization by co-solvents. *Journal of Crystal Growth*, 90:1–3, 1988.
- [40] M. Athawale, S. Sarupria, and S. Garde. Enthalpy-entropy contributions to salt and osmolyte effects on molecular-scale hydrophobic hydration and interactions. *The Journal of Physical Chemistry B*, 112:5661–5670, 2008.
- [41] P. Lo Nostro and B. Ninham. Hofmeister phenomena: An update on ion specificity in biology. *Chemical Reviews*, 112:2286–2322, 2012.
- [42] 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. *Current Opinion in Colloid & Interface Science*, 9:19–37, 2004.
- [43] K. Collins and M. Washabaugh. The hofmeister effect and the behaviour of water at interfaces. *Quarterly Reviews of Biophysics*, 18:323–422, 1985.
- [44] Y. Zhang and P. Cremer. Interactions between macromolecules and ions: The hofmeister series. *Current Opinion in Chemical Biology*, 10:658–663, 2006.
- [45] L. Dang. Computational study of ion binding to the liquid interface of water. *The Journal of Physical Chemistry B*, 106:10388–10394, 2002.
- [46] T. Chang and L. Dang. Recent advances in molecular simulations of ion solvation at liquid interfaces. *Chemical Reviews*, 106:1305–1322, 2006.
- [47] P. Jungwirth and D. Tobias. Specific ion effects at the air/water interface. *Chemical Reviews*, 106:1259–1281, 2006.
- [48] P. Petersen and R. Saykally. On the nature of ions at the liquid water surface. *Annual Review of Physical Chemistry*, 57:333–364, 2006.
- [49] M. Baer and C. Mundy. Toward an understanding of the specific ion effect using density functional theory. *The Journal of Physical Chemistry Letters*, 2:1088–1093, 2011.
- [50] R. Netz and D. Horinek. Progress in modeling of ion effects at the vapor/water interface. *Annual Review of Physical Chemistry*, 63:401–418, 2012.
- [51] D. Horinek, A. Herz, L. Vrbka, F. Sedlmeier, S. Mamatkulov, and R. Netz. Specific ion adsorption at the air/water interface: The role of hydrophobic solvation. *Chemical Physics Letters*, 479:173–183, 2009.
- [52] L. Dang and T. Chang. Molecular mechanism of ion binding to the liquid/vapor interface of water. *The Journal of Physical Chemistry B*, 106:235–238, 2002.
- [53] D. Herce, L. Perera, T. Darden, and C. Sagui. Surface solvation for an ion in a water cluster. *The Journal of Chemical Physics*, 122:024513(1–10), 2004.

- [54] P. Jungwirth and D. Tobias. Ions at the air/water interface. *The Journal of Physical Chemistry B*, 106:6361–6373, 2002.
- [55] D. Otten, P. Shaffer, P. Geissler, and R. Saykally. Elucidating the mechanism of selective ion adsorption to the liquid water surface. *Proceedings of the National Academy of Sciences*, 109:701–705, 2012.
- [56] S. Ou, Y. Hu, H. Wan, and S. Patel. Spherical monovalent ions at aqueous liquid-vapor interfaces: Interfacial stability and induced interface fluctuations. *The Journal of Physical Chemistry B*, 117:11732–11742, 2013.
- [57] S. Ou and S. Patel. Temperature dependence and energetics of single ions at the aqueous liquid-vapor interface. *The Journal of Physical Chemistry B*, 117:6512–6523, 2013.
- [58] P. Jungwirth and B. Winter. Ions at aqueous interfaces: From water surface to hydrated proteins. *Annual Review of Physical Chemistry*, 59:343–366, 2008.
- [59] J. Heyda, J. Vincent, D. Tobias, J. Dzubiella, and P. Jungwirth. Ion specificity at the peptide bond: Molecular dynamics simulations of *n*-methylacetamide in aqueous salt solutions. *The Journal of Physical Chemistry B*, 114:1213–1220, 2010.
- [60] M. Lund, R. Vacha, and P. Jungwirth. Specific ion binding to macromolecules: Effects of hydrophobicity and ion pairing. *Langmuir*, 24:3387–3391, 2008.
- [61] P. Mason, C. Dempsey, L. Vrbka, J. Heyda, J. Brady, and P. Jungwirth. Specificity of ion-protein interactions: Complementary and competitive effects of tetrapropylammonium, guanidinium, sulfate, and chloride ions. *The Journal of Physical Chemistry B*, 113:3227–3234, 2009.
- [62] K. Rembert, J. Paterova, J. Heyda, C. Hilty, P. Jungwirth, and P. Cremer. Molecular mechanisms of ion-specific effects on proteins. *Journal of the American Chemical Society*, 134:10039–10046, 2012.
- [63] L. Vrbka, P. Jungwirth, P. Bauduin, D. Touraud, and W. Kunz. Specific ion effects at protein surfaces: A molecular dynamics study of bovine pancreatic trypsin inhibitor and horseradish peroxidase in selected salt solutions. *The Journal of Physical Chemistry B*, 110:7036–7043, 2006.
- [64] N. Schwierz, D. Horinek, and R. Netz. Reversed anionic hofmeister series: The interplay of surface charge and surface polarity. *Langmuir*, 26:7370–7379, 2010.
- [65] N. Schwierz, D. Horinek, and R. Netz. Anionic and cationic hofmeister effects on hydrophobic and hydrophilic surfaces. *Langmuir*, 29:2602–2614, 2013.

- [66] M. Lund, P. Jungwirth, and C. Woodward. Ion specific protein assembly and hydrophobic surface forces. *Physical Review Letters*, 100:258105(1–4), 2008.
- [67] M. Auton and B. Wayne. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proceedings of the National Academy of Sciences*, 102:15065–15068, 2005.
- [68] D. Bolen and I. Baskakov. The osmophobic effect: Natural selection of a thermodynamic force in protein folding. *Journal of Molecular Biology*, 310:955 – 963, 2001.
- [69] D. Canchi and A. García. Backbone and side-chain contributions in protein denaturation by urea. *Biophysical Journal*, 100:1526 – 1533, 2011.
- [70] L. Hua, R. Zhou, D. Thirumalai, and B. Berne. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proceedings of the National Academy of Sciences*, 105:16928–16933, 2008.
- [71] A. Ben-Naim. Theoretical aspects of pressure and solute denaturation of proteins: A kirkwood-buff-theory approach. *The Journal of Chemical Physics*, 137:235102(1–8), 2012.
- [72] D. Canchi, D. Paschek, and A. García. Equilibrium study of protein denaturation by urea. *Journal of the American Chemical Society*, 132:2338–2344, 2010.
- [73] P. Mason, G. Neilson, J. Enderby, M. Saboungi, C. Dempsey, A. MacKerell, and J. Brady. The structure of aqueous guanidinium chloride solutions. *Journal of the American Chemical Society*, 126:11462–11470, 2004.
- [74] T. Koishi, K. Yasuoka, S. Willow, S. Fujikawa, and X. Zeng. Molecular insight into different denaturing efficiency of urea, guanidinium, and methanol: A comparative simulation study. *Journal of Chemical Theory and Computation*, 9:2540–2551, 2013.
- [75] J. England, V. Pande, and G. Haran. Chemical denaturants inhibit the onset of dewetting. *Journal of the American Chemical Society*, 130:11854–11855, 2008.
- [76] M. Vazdar, J. Vymetal, J. Heydra, J. Vondrasek, and P. Jungwirth. Like-charge guanidinium pairing from molecular dynamics and ab initio calculations. *The Journal of Physical Chemistry A*, 115:11193–11201, 2011.
- [77] R. Godawat, S. Jamadagni, and S. Garde. Unfolding of hydrophobic polymers in guanidinium chloride solutions. *The Journal of Physical Chemistry B*, 114:2246–2254, 2010.

- [78] J. Soetens, C. Millot, C. Chipot, G. Jansen, J. Angyan, and B. Maigret. Effect of polarizability on the potential of mean force of two cations. the guanidinium-guanidinium ion pair in water. *The Journal of Physical Chemistry B*, 1997:10910–10917, 1997.
- [79] J. Rösgen, B. Pettitt, and D. Bolen. Protein folding, stability, and solvation structure in osmolyte solutions. *Biophysical Journal*, 89:2988 – 2997, 2005.
- [80] B. Moeser and D. Horinek. Unified description of urea denaturation: Backbone and side chains contribute equally in the transfer model. *The Journal of Physical Chemistry B*, 118:107–114, 2014.
- [81] M. Cacace, E. Landau, and J. Ramsden. The Hofmeister series: Salt and solvent effects on interfacial phenomena. *Quarterly Reviews of Biophysics*, 30:241–277, 1997.
- [82] C. Ma, C. Wang, C. Acevedo-Velez, S. Gellman, and N. Abbott. Modulation of hydrophobic interactions by proximally immobilized ions. *Nature*, 517:347–350, 2015.
- [83] A. Caffisch and M. Karplus. Structural details of urea binding to barnase: a molecular dynamics analysis. *Structure*, 7:477–488, 1999.
- [84] B. Bennion and V. Daggett. The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences*, 100:5142–5147, 2003.
- [85] P. Mason, G. Neilson, C. Dempsey, A. Barnes, and J. Cruickshank. The hydration structure of guanidinium and thiocyanate ions: Implications for protein stability in aqueous solution. *Proceedings of the National Academy of Sciences*, 100:4557–4561, 2003.
- [86] E. Wernersson, J. Heyda, M. Vazdar, M. Lund, P. Mason, and P. Jungwirth. Orientational dependence of the affinity of guanidinium ions to the water surface. *The Journal of Physical Chemistry B*, 115:12521–12526, 2011.
- [87] S. Ou, D. Cui, and S. Patel. Liquid-vapor interfacial properties of aqueous solutions of guanidinium and methyl guanidinium chloride: Influence of molecular orientation on interface fluctuations. *The Journal of Physical Chemistry B*, 117:11719–11731, 2013.
- [88] J. England and G. Haran. Role of solvation effects in protein denaturation: From thermodynamics to single molecules and back. *Annual Review of Physical Chemistry*, 62:257–277, 2011.

- [89] K. Dill, T. Truskett, V. Vlachy, and B. Hribar-Lee. Modeling water, the hydrophobic effect, and ion solvation. *Annual Review of Biophysics and Biomolecular Structure*, 34:173–199, 2005.
- [90] A. Daranas, H. Shimizu, and S. Homans. Thermodynamics of binding of d-galactose and deoxy derivatives thereof to the l-arabinose-binding protein. *Journal of the American Chemical Society*, 126:11870–11876, 2004.
- [91] J. Kyte. The basic of the hydrophobic effect. *Biophysical Chemistry*, 100:193–203, 2003.
- [92] A. Wallqvist and B. Berne. Computer simulation of hydrophobic hydration forces on stacked plates at short range. *The Journal of Physical Chemistry*, 99:2893–2899, 1995.
- [93] R. Zangi and B. Berne. Temperature dependence of dimerization and dewetting of large-scale hydrophobes: A molecular dynamics study. *The Journal of Physical Chemistry B*, 112:8634–8644, 2008.
- [94] N. Choudhury and B. Pettitt. Enthalpy-entropy contributions to the potential of mean force of nanoscopic hydrophobic solutes. *The Journal of Physical Chemistry B*, 110:8459–8463, 2006.
- [95] S. Ou, S. Patel, and B. Bauer. Free energetics of carbon nanotube association in pure and aqueous ionic solutions. *The Journal of Physical Chemistry B*, 116:8154–8168, 2012.
- [96] S. Sharrow, M. Novotny, and M. Stone. Thermodynamic analysis of binding between mouse major urinary protein-i and the pheromone 2-sec-butyl-4,5-dihydrothiazole. *Biochemistry*, 20:6302–6309, 2003.
- [97] R. Bingham, J. Findlay, S. Hsieh, A. Kalverda, A. Kjeliberg, C. Perazzolo, S. Phillips, K. Seshadri, C. Trinh, W. Turnbull, G. Bodenhausen, and S. Homans. Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein. *Journal of the American Chemical Society*, 126:1675–1681, 2004.
- [98] E. Barratt, R. Bingham, D. Warner, C. Laughton, S. Phillips, and S. Homans. Van der waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water. *Journal of the American Chemical Society*, 127:11827–11834, 2005.
- [99] R. Talhout, A. Villa, A. Mark, and J. Engberts. Understanding binding affinity: A combined isothermal titration calorimetry/molecular dynamics study of the binding of a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin. *Journal of the American Chemical Society*, 125:10570–10579, 2003.

- [100] P. Setny, R. Baron, and J. McCammon. How can hydrophobic association be enthalpy driven? *Journal of Chemical Theory and Computation*, 6:2866–2871, 2010.
- [101] A. MacKerell. Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry*, 25:1584–1604, 2004.
- [102] D. Yin and A. MacKerell. Combined ab initio/empirical approach for the optimization of lennard-jones parameters. *Journal of Computational Chemistry*, 19:334–348, 1998.
- [103] G. Warren and S. Patel. Comparison of the solvation structure of polarizable and nonpolarizable ions in bulk water and near the aqueous liquid-vapor interface. *The Journal of Physical Chemistry C*, 112:7455–7467, 2008.
- [104] J. Harris, J. Gryko, and S. Rice. Self-consistent monte carlo simulations of the electron and ion distributions of inhomogeneous liquid alkali metals. i. longitudinal and transverse density distributions in the liquid-vapor interface of a one-component system. *The Journal of Chemical Physics*, 87:3069–3081, 1987.
- [105] J. Huang and W. Webb. Diffuse interface in a critical fluid mixture. *The Journal of Chemical Physics*, 50:3677–3693, 1969.
- [106] D. Beysens and M. Robert. Thickness of fluid interfaces near the critical point from optical reflectivity measurements. *The Journal of Chemical Physics*, 87:3056–3061, 1987.
- [107] A. Ismail, G. Grest, and M. Stevens. Capillary waves at the liquid-vapor interface and the surface tension of water. *The Journal of Chemical Physics*, 125:014702(1–10), 2006.
- [108] J. Alejandre, D. Tildesley, and G. Chapela. Molecular dynamics simulation of the orthobaric densities and surface tension of water. *The Journal of Chemical Physics*, 102:4574–4583, 1995.
- [109] W. Kuo, C. Mundy, B. Eggimann, M. McGrath, I. Siepmann, B. Chen, J. Vieceli, and D. Tobias. Structure and dynamics of the aqueous liquid-vapor interface: A comprehensive particle-based simulation study. *The Journal of Physical Chemistry B*, 110:3738–3746, 2006.
- [110] E. Blokhuis, D. Bedeaux, C. Holcomb, and J. Zollweg. Tail corrections to the surface tension of a lennard-jones liquid-vapour interface. *Molecular Physics*, 85:665–669, 1995.
- [111] B. Widom. Potential-distribution theory and the statistical mechanics of fluids. *The Journal of Physical Chemistry*, 86:869–872, 1982.

- [112] J. Kirkwood and F. Buff. The statistical mechanical theory of surface tension. *The Journal of Chemical Physics*, 17:338–343, 1949.
- [113] C. Vega and E. Miguel. Surface tension of the most popular models of water by using the test-area simulation method. *The Journal of Chemical Physics*, 126:154707(1–10), 2007.
- [114] E. Guggenheim. The principle of corresponding states. *The Journal of Chemical Physics*, 13:253–261, 1945.
- [115] L. Vega, E. Miguel, L. Rull, G. Jackson, and I. McLure. Phase equilibria and critical behavior of square-well fluids of variable width by gibbs ensemble monte carlo simulation. *The Journal of Chemical Physics*, 96:2296–2305, 1992.
- [116] J. Alejandre, J. Rivera, M. Mora, and V. Garza. Force field of monoethanolamine. *The Journal of Physical Chemistry B*, 104:1332–1337, 2000.
- [117] S. Patel and C. Brooks. A nonadditive methanol force field: Bulk liquid and liquid-vapor interfacial properties via molecular dynamics simulations using a fluctuating charge model. *The Journal of Chemical Physics*, 122:024508(1–10), 2005.
- [118] J. Harris and K. Yung. Carbon dioxide’s liquid-vapor coexistence curve and critical properties as predicted by a simple molecular model. *The Journal of Physical Chemistry*, 99:12021–12024, 1995.
- [119] J. Nelder and R. Mead. A simplex algorithm for function minimization. *The Computer Journal*, 7:308–313, 1965.
- [120] B. Bauer and S. Patel. Properties of water along the liquid-vapor coexistence curve via molecular dynamics simulations using the polarizable tip4p-qdp-lj water model. *The Journal of Chemical Physics*, 131:084709(1–16), 2009.
- [121] A. Imre, G. Mayer, G. Hazi, R. Rozas, and T. Kraska. Estimation of the liquid-vapor spinodal from interfacial properties obtained from molecular dynamics and lattice boltzmann simulations. *J. Chem. Phys.*, 128:114708(1–11), 2008.
- [122] C. Ibergay, A. Ghoufi, F. Goujon, P. Ungerer, A. Boutin, B. Rousseau, and P. Malfreyt. Molecular simulations of the n-alkane liquid-vapor interface: Interfacial properties and their long range corrections. *Phys. Rev. E*, 75:051602(1–18), 2007.
- [123] C. Vega, J. Abascal, and I. Nezbeda. Vapor-liquid equilibria from the triple point up to the critical point for the new generation of tip4p-like models: Tip4p/ew, tip4p/2005, and tip4p/ice. *J. Chem. Phys.*, 125:034503(1–9), 2006.

- [124] W. Fawcett. Liquids, solutions, and interfaces: From macroscopic descriptions to modern microscopic details. *Oxford University Press*, 2004.
- [125] J. Koryta, J. Dvorak, and L. Kavan. Principles of electrochemistry. *Wiley*, 1993.
- [126] A. Bard and L. Faulkner. Electrochemical methods: Fundamentals and applications. *Wiley*, 1980.
- [127] S. Kathmann, I. Kuo, and C. Mundy. Electronic effects on the surface potential at the vapor-liquid interface of water. *Journal of the American Chemical Society*, 130:16556–16561, 2008.
- [128] A. Adamson. Physical chemistry of surfaces. *Wiley*, 1990.
- [129] D. Asthagiri, L. Pratt, and H. Ashbaugh. Absolute hydration free energies of ions, ion-water clusters, and quasichemical theory. *The Journal of Chemical Physics*, 119:2702–2708, 2003.
- [130] M. Kastenholtz and P. Hünenberger. Computation of methodology-independent ionic solvation free energies from molecular simulations. i. the electrostatic potential in molecular liquids. *The Journal of Chemical Physics*, 124:124106, 2006.
- [131] P. Hünenberger and J. McCammon. Ewald artifacts in computer simulations of ionic solvation and ion-ion interaction: A continuum electrostatics study. *The Journal of Chemical Physics*, 110:1856–1872, 1999.
- [132] S. Pandit, D. Bostick, and M. Berkowitz. Mixed bilayer containing dipalmitoylphosphatidylcholine and dipalmitoylphosphatidylserine: Lipid complexation, ion binding, and electrostatics. *Biophysical Journal*, 85:3120–3131, 2003.
- [133] S. Patel and C. Brooks. Structure, thermodynamics, and liquid-vapor equilibrium of ethanol from molecular-dynamics simulations using nonadditive interactions. *The Journal of Chemical Physics*, 123:164502, 2005.
- [134] M. Wilson, A. Pohorille, and L. Pratt. Comment on “study on the liquid-vapor interface of water. i. simulation results of thermodynamic properties and orientational structure”. *The Journal of Chemical Physics*, 90:5211–5213, 1989.
- [135] J. Jackson. Classical electrodynamics. *Wiley*, 1998.
- [136] V. Sokhan and D. Tildesley. The free surface of water: molecular orientation, surface potential and nonlinear susceptibility. *Molecular Physics*, 92:625–640, 1997.
- [137] J. Kirkwood. Statistical mechanics of fluid mixtures. *Journal of Computational Chemistry*, 3:300–313, 1935.

- [138] K. Wong and D. York. Exact relation between potential of mean force and free-energy profile. *Journal of Chemical Theory and Computation*, 8:3998–4003, 2012.
- [139] E. Darve, D. Rodriguez-Gomez, and A. Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of Chemical Physics*, 128:144120, 2008.
- [140] J. Henin, G. Fiorin, C. Chipot, and M. Klein. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *Journal of Chemical Theory and Computation*, 6:35–47, 2010.
- [141] E. Darve and A. Pohorille. Calculating free energies using average force. *The Journal of Chemical Physics*, 115:9169–9183, 2001.
- [142] C. Faller, K. Reilly, R. Hills, and O. Guvench. Peptide backbone sampling convergence with the adaptive biasing force algorithm. *The Journal of Physical Chemistry B*, 117:518–526, 2013.
- [143] D. Rodriguez-Gomez, E. Darve, and A. Pohorille. Assessing the efficiency of free energy calculation methods. *The Journal of Chemical Physics*, 120:3563–3578, 2004.
- [144] C. Chipot and J. Henin. Exploring the free-energy landscape of a short peptide using an average force. *The Journal of Chemical Physics*, 123:244906(1–6), 2005.
- [145] S. Kumar, J. Rosenberg, D. Bouzida, R. Swendsen, and P. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13:1011–1021, 1992.
- [146] G. Torrie and J. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Chemistry*, 23:187–199, 1977.
- [147] A. Willard and D. Chandler. Instantaneous liquid interfaces. *The Journal of Physical Chemistry B*, 114:1954–1958, 2010.
- [148] A. Bizzarri and S. Cannistraro. Molecular dynamics of water at the protein-solvent interface. *The Journal of Physical Chemistry B*, 106:6617–6633, 2002.
- [149] V. Makarov, B. Pettitt, and M. Feig. Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment. *Accounts of Chemical Research*, 35:376–384, 2002.
- [150] Y. Levy and J. Onuchic. Water and proteins: A love-hate relationship. *Proceedings of the National Academy of Sciences*, 101:3325–3326, 2004.

- [151] S. Pal and A. Zewail. Dynamics of water in biological recognition. *Chemical Reviews*, 104:2099–2123, 2004.
- [152] A. Geiger, F. Stillinger, and A. Rahman. Aspects of the percolation process for hydrogenbond networks in water. *The Journal of Chemical Physics*, 70:4185–4193, 1979.
- [153] I. Brovchenko, A. Geiger, and A. Oleinikova. Clustering of water molecules in aqueous solutions: Effect of water-solute interaction. *Physical Chemistry Chemical Physics*, 6:1982–1987, 2004.
- [154] D. Murakami and K. Yasuoka. Molecular dynamics simulation of quasi-two-dimensional water clusters on ice nucleation protein. *The Journal of Chemical Physics*, 137:054303(1–5), 2012.
- [155] A. Oleinikova, I. Brovchenko, N. Smolin, A. Krukau, A. Geiger, and R. Winter. Percolation transition of hydration water: from planar hydrophilic surfaces to proteins. *Physical Review Letters*, 95:247802(1–4), 2005.
- [156] A. Oleinikova, N. Smolin, I. Brovchenko, A. Geiger, and R. Winter. Formation of spanning water networks on protein surfaces via 2d percolation transition. *The Journal of Physical Chemistry B*, 109:1988–1998, 2005.
- [157] W. Doster, A. Bachleitner, R. Dunau, M. Hiebl, and E. Luscher. Thermal properties of water in myoglobin crystals and solutions at subzero temperatures. *Biophysical Journal*, 50:213–219, 1986.
- [158] A. Kuffel and J. Zielkiewicz. Why the solvation water around proteins is more dense than bulk water. *The Journal of Physical Chemistry B*, 116:12113–12124, 2012.
- [159] A. Willard and D. Chandler. Coarse-grained modeling of the interface between water and heterogeneous surfaces. *Faraday Discussions*, 141:209–220, 2009.
- [160] L. Young, R. Jernigan, and D. Covell. A role for surface hydrophobicity in protein-protein recognition. *Protein Science*, 3:717–729, 1994.
- [161] C. Tsai, S. Lin, H. Wolfson, and R. Nussinov. Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Science*, 6:53–64, 1997.
- [162] R. Patil, S. Das, A. Stanley, L. Yadav, A. Sudhakar, and A. Varma. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PLoS One*, 5:12029(1–10), 2010.
- [163] L. Pratt. Theory of hydrophobic effects. *Annual Review of Physical Chemistry*, 36:433–449, 1985.

- [164] I. Moreira, P. Fernandes, and M. Ramos. Hot spots - a review of the protein-protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics*, 68:803–812, 2007.
- [165] J. Vaynberg and J. Qin. Weak protein-protein interactions as probed by nmr spectroscopy. *Trends in Biotechnology*, 24:22–27, 2006.
- [166] I. Nooren and J. Thornton. Structural characterisation and functional significance of transient protein-protein interactions. *Journal of Molecular Biology*, 325:991–1018, 2003.
- [167] N. Giovambattista, P. Rossky, and P. Debenedetti. Effect of pressure on the phase behavior and structure of water confined between nanoscale hydrophobic and hydrophilic plates. *Physical Review E*, 73:041604(1–14), 2006.
- [168] Giovambattista N., Lopez C., Rossky P., and Debenedetti P. Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proceedings of the National Academy of Sciences*, 105:2274–2279, 2008.
- [169] D. Cui, S. Ou, E. Peters, and S. Patel. Ion-specific induced fluctuations and free energetics of aqueous protein hydrophobic interfaces: Toward connecting to specific-ion behaviors at aqueous liquid-vapor interfaces. *The Journal of Physical Chemistry B*, 118:4490–4504, 2014.
- [170] K. Swanson, R. Kang, S. Stamenova, L. Hicke, and I. Radhakrishnan. Solution structure of vps27 uim-ubiquitin complex important for endosomal sorting and receptor downregulation. *The EMBO Journal*, 22:4597–4606, 2003.
- [171] R. Kang, C. Daniels, S. Francis, S. Shih, W. Salerno, L. Hicke, and I. Radhakrishnan. Solution structure of a cue-ubiquitin complex reveals a conserved mode of ubiquitin binding. *Cell*, 113:621–630, 2003.
- [172] A. Ohno, J. Jee, K. Fujiwara, T. Tenno, N. Goda, H. Tochio, H. Kobayashi, H. Hiroaki, and M. Shirakawa. Structure of the uba domain of dsk2p in complex with ubiquitin: Molecular determinants for ubiquitin recognition. *Structure*, 13:521–532, 2005.
- [173] M. Kawasaki, T. Shiba, Y. Shiba, Y. Yamaguchi, N. Matsugaki, N. Igarashi, M. Suzuki, R. Kato, K. Kato, K. Nakayama, and S. Wakatsuki. Molecular mechanism of ubiquitin recognition by gga3 gat domain. *Genes Cells*, 10:639–654, 2005.
- [174] Y. Chang, A. Song, Y. Gao, Y. Shi, X. Lin, X. Cao, D. Lin, and H. Hu. Solution structure of the ubiquitin-associated domain of human bmsc-ubp and its complex with ubiquitin. *Protein Science*, 15:1248–1259, 2006.

- [175] T. Mueller and J. Feigon. Structural determinants for the binding of ubiquitin-like domains to the proteasome. *The EMBO Journal*, 22:4634–4645, 2003.
- [176] K. Fujiwara, T. Tenno, K. Sugasawa, J. Jee, I. Ohki, C. Kojima, H. Tochio, H. Hiroaki, F. Hanaoka, and M. Shirakawa. Structure of the ubiquitin-interacting motif of s5a bound to the ubiquitin-like domain of hr23b. *The Journal of Biological Chemistry*, 6:4760–4767, 2004.
- [177] D. Reverter, K. Wu, T. Erdene, Z. Pan, K. Wilkinson, and C. Lima. Structure of a complex between nedd8 and the ulp/senp protease family member den1. *Journal of Molecular Biology*, 345:141–151, 2005.
- [178] T. Pawson and J. Schlessingert. Sh2 and sh3 domains. *Current Biology*, 3:434–442, 1993.
- [179] Y. He, L. Hicke, and I. Radhakrishnan. Structural basis for ubiquitin recognition by sh3 domains. *Journal of Molecular Biology*, 373:190–196, 2007.
- [180] J. Vaynberg, T. Fukuda, K. Chen, O. Vinogradova, A. Velyvis, Y. Tu, L. Ng, C. Wu, and J. Qin. Structure of an ultraweak protein-protein complex and its crucial role in regulation of cell morphology and motility. *Molecular Cell*, 17:513–523, 2005.
- [181] I. Bezsonova, M. Bruce, S. Wiesner, H. Lin, D. Rotin, and J. Forman-Kay. Interactions between the three cin85 sh3 domains and ubiquitin: Implications for cin85 ubiquitination. *Biochemistry*, 47:8937–8949, 2008.
- [182] L. Donaldson, G. Gish, T. Pawson, L. Kay, and J. Forman-Kay. Structure of a regulatory complex involving the abl sh3 domain, the crk sh2 domain, and a crk-derived phosphopeptide. *Proceedings of the National Academy of Sciences*, 99:14053–14058, 2002.
- [183] J. Hakanpaa, A. Paananen, S. Askolin, T. Nakari-Setala, T. Parkkinen, M. Penttila, M. Linder, and J. Rouvinen. Atomic resolution structure of the hfbii hydrophobin, a self-assembling amphiphile. *The Journal of Biological Chemistry*, 279:534–539, 2004.
- [184] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, 151:283–312, 1999.
- [185] J. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with namd. *Journal of Computational Chemistry*, 26:1781–1802, 2005.

- [186] A. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102:3586–3616, 1998.
- [187] A. Mackerell, M. Feig, and C. Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*, 25:1400–1415, 1998.
- [188] N. Deshpande, K. Address, W. Bluhm, J. Merino-Ott, W. Townsend-Merino, Q. Zhang, C. Knezevich, L. Xie, L. Chen, Z. Feng, R. Green, J. Flippen-Anderson, J. Westbrook, H. Berman, and P. Bourne. The rcsb protein data bank: a redesigned query system and relational database based on the mmcif schema. *Nucleic Acids Research*, 33:233–237, 2005.
- [189] R. Kumar, V. Iyer, and W. Im. Charmm-gui: A graphical user interface for the charmm users. *Journal of Computational Chemistry*, 29:1859–1865, 2007.
- [190] W. Jorgensen and C. Jenson. Temperature dependence of tip3p, spc, and tip4p water from npt monte carlo simulations: Seeking temperatures of maximum density. *Journal of Computational Chemistry*, 19:1179–1186, 1998.
- [191] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An $n^*\log(n)$ method for ewald sums in large systems. *The Journal of Chemical Physics*, 98:10089–10092, 1993.
- [192] J. Ryckaert, G. Ciccotti, and H. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23:327–341, 1977.
- [193] I. Brovchenko and A. Oleinikova. Which properties of a spanning network of hydration water enable biological functions? *ChemPhysChem*, 9:2695–2702, 2008.
- [194] A. Patel, P. Varilly, S. Jamadagni, M. Hagan, D. Chandler, and S. Garde. Sitting at the edge: How biomolecules use hydrophobicity to tune their interactions and function. *The Journal of Physical Chemistry B*, 116:2498–2503, 2012.
- [195] A. Oleinikova and I. Brovchenko. Percolation transition of hydration water in biosystems. *Molecular Physics*, 104:3841–3855, 2006.
- [196] N. Smolin, A. Oleinikova, I. Brovchenko, A. Geiger, and R. Winter. Properties of spanning water networks at protein surfaces. *The Journal of Physical Chemistry B*, 109:10995–11005, 2005.

- [197] A. Oleinikova, I. Brovchenko, A. Geiger, and B. Guillot. Percolation of water in aqueous solution and liquid-liquid immiscibility. *The Journal of Chemical Physics*, 117:3296–3304, 2002.
- [198] D. Stauffer. *Introduction to Percolation Theory*. Taylor and Francis: London and Philadelphia, 1985.
- [199] B. Lee and F. Richards. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, 55:379–400, 1971.
- [200] F. Richards. Areas, volumes, packing, and protein structure. *Annual Review of Biophysics and Biomolecular Structure*, 6:151–176, 1977.
- [201] L. Wesson and D. Eisenberg. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Science*, 1:227–235, 1992.
- [202] R. Sibson. Slink: An optimally efficient algorithm for the single-link cluster method. *The Computer Journal*, 16:30–34, 1973.
- [203] A. Oleinikova and I. Brovchenko. Thermodynamic properties of hydration water around solutes: Effect of solute size and water-solute interaction. *The Journal of Physical Chemistry B*, 116:14650–14659, 2012.
- [204] D. Cui, S. Ou, and S. Patel. Free energetics of rigid body association of ubiquitin binding domains: A biochemical model for binding mediated by hydrophobic interaction. *Proteins: Structure, Function, and Bioinformatics*, 82:1453–1468, 2014.
- [205] J. Winget and T. Mayor. The diversity of ubiquitin recognition: Hot spots and varied specificity. *Molecular Cell*, 38:627–635, 2010.
- [206] O. Keskin and R. Nussinov. Similar binding sites and different partners: Implications to shared proteins in cellular pathways. *Structure*, 15:341–354, 2007.
- [207] G. Cornilescu, B. Lee, C. Cornilescu, G. Wang, A. Peterkofsky, and G. Clore. Solution structure of the phosphoryl transfer complex between the cytoplasmic a domain of the mannitol transporter iimannitol and hpr of the escherichia coli phosphotransferase system. *The Journal of Biological Chemistry*, 277:42289–42298, 2002.
- [208] P. Petersen and R. Saykally. Confirmation of enhanced anion concentration at the liquid water surface. *Chemical Physics Letters*, 397:51–55, 2004.
- [209] P. Nandi and D. Robinson. Effects of salts on the free energy of the peptide group. *Journal of the American Chemical Society*, 94:1299–1308, 1972.
- [210] D. Horinek and R. Netz. Specific ion adsorption at hydrophobic solid surfaces. *Physical Review Letters*, 99:226104, 2007.

- [211] R. Friedman, E. Nachliel, and M. Gutman. Molecular dynamics of a protein surface: Ion-residues interactions. *Biophysical Journal*, 89:768–781, 2005.
- [212] M. Lund, L. Vrbka, and P. Jungwirth. Specific ion binding to nonpolar surface patches of proteins. *Journal of the American Chemical Society*, 130:11582, 2008.
- [213] M. Lund and P. Jungwirth. Patchy proteins, anions and hofmeister series. *Journal of Physics: Condensed Matter*, 20:494218, 2008.
- [214] B. Brooks, C. Brooks, A. MacKerell, L. Nilsson, R. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. Dinner, M. Feig, S. Fisher, J. Gao, M. Hodoscek, W. IM, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. Pastor, C. Post, J. Pu, M. Schaefer, B. Tidor, R. Venable, H. Woodcock, X. Wu, W. Yang, D. York, and M. Karplus. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 30:1545–1614, 2009.
- [215] S. Nosé. A molecular dynamics methods for simulations in the canonical ensemble. *Molecular Physics*, 52:255–268, 1984.
- [216] I. Joung and T. Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The Journal of Physical Chemistry B*, 112:9020–9041, 2008.
- [217] C. Wick and L. Dang. Recent advances in understanding transfer ions across aqueous interfaces. *Chemical Physics Letters*, 458:1–5, 2008.
- [218] J. Hakanpää, M. Linder, A. Popov, A. Schmidt, and J. Rouvinen. Hydrophobin hfbii in detail: Ultrahigh-resolution structure at 0.75Å. *Acta Crystallographica*, 62:356–367, 2006.
- [219] G. Zhu, F. and Hummer. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry*, 33:453–465, 2012.
- [220] H. Flyvbjerg and H. Petersen. Error estimates on averages of correlated data. *The Journal of Chemical Physics*, 91:461–466, 1989.
- [221] H. Yu, T. Whitfield, E. Harder, G. Lamoureux, I. Vorobyov, V. Anisimov, A. MacKerel, and B. Roux. Simulating monovalent and divalent ions in aqueous solution using a drude polarizable force field. *Journal of Chemical Theory and Computation*, 6:774–786, 2010.
- [222] J. Neyt, A. Wender, V. Lachet, A. Ghoufi, and P. Malfreyt. Prediction of the concentration dependence of the surface tension and density of salt solutions: Atomistic simulations using drude oscillator polarizable and nonpolarizable models. *Physical Chemistry Chemical Physics*, 15:11679–11690, 2013.

- [223] G. Lamoureux and B. Roux. Absolute hydration free energy scale for alkali and halide ions established from simulations with a polarizable force field. *The Journal of Physical Chemistry B*, 110:3308–3322, 2006.
- [224] M. Auton, L. Holthauzen, and D. Bolen. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proceedings of the National Academy of Sciences*, 104:15317–15322, 2007.
- [225] M. Auton, A. Ferreon, and D. Bolen. Metrics that differentiate the origins of osmolyte effects on protein stability: A test of the surface tension proposal. *Journal of Molecular Biology*, 361:983–992, 2006.
- [226] C. Hu, H. Kokubo, G. Lynch, D. Bolen, and B. Pettitt. Backbone additivity in the transfer model of protein solvation. *Protein Science*, 19:1011–1022, 2010.
- [227] M. Auton, J. Rosgen, M. Sinev, L. Holthauen, and D. Bolen. Osmolyte effects on protein stability and solubility: A balancing act between backbone and side-chains. *Biophysical Chemistry*, 159:90–99, 2011.
- [228] A. Kubícková, T. Krížek, P. Coufal, E. Wernersson, J. Heyda, and P. Jungwirth. Guanidinium cations pair with positively charged arginine side chains in water. *The Journal of Physical Chemistry Letters*, 2:1387–1389, 2011.
- [229] A. Patel, P. Varilly, S. Jamadagni, H. Acharya, S. Garde, and D. Chandler. Extended surfaces modulate hydrophobic interactions of neighboring solutes. *Proceedings of the National Academy of Sciences*, 108:17678–17683, 2011.
- [230] M. Linder, G. Szilvay, T. Nakari-Setälä, and M. Penttilä. Hydrophobins: The protein-amphiphiles of filamentous fungi. *FEMS Microbiology Reviews*, 29:877–896, 2005.
- [231] M. Linder. Hydrophobins: Proteins that self assemble at interfaces. *Current Opinion in Colloid & Interface Science*, 14:356–363, 2009.
- [232] J. Kallio, M. Linder, and J. Rouvinen. Crystal structures of hydrophobin HFBII in the presence of detergent implicate the formation of fibrils and monolayer films. *The Journal of Biochemistry*, 282:28733–28739, 2007.
- [233] R. Cooper, S. Heiles, M. DiTucci, and E. Williams. Hydration of guanidinium: Second shell formation at small cluster size. *The Journal of Physical Chemistry A*, 118:5657–5666, 2014.
- [234] M. Stumpe and H. Grubmüller. Aqueous urea solutions: Structure, energetics, and urea aggregation. *The Journal of Physical Chemistry B*, 111:6220–6228, 2007.

- [235] P. Mason, J. Brady, G. Neilson, and C. Dempsey. The interaction of guanidinium ions with a model peptide. *Biophysical Journal*, 93:L04–L06, 2007.
- [236] A. Patel, P. Varilly, S. Jamadagni, M. Hagan, D. Chandler, and S. Garde. Sitting at the edge: How biomolecules use hydrophobicity to tune their interactions and function. *The Journal of Physical Chemistry B*, 116:2498–2503, 2012.
- [237] S. Jamadagni, R. Godawat, J. Dordick, and S. Garde. How interfaces affect hydrophobically driven polymer folding. *The Journal of Physical Chemistry B*, 113:4093–4101, 2009.
- [238] M. Baer and C. Mundy. An ab initio approach to understanding the specific ion effect. *Faraday Discussions*, 160:89–101, 2013.
- [239] H. Bohm and G. Schneider. Protein-ligand interactions: From molecular recognition to drug design. *Wiley-VCH*, 2003.
- [240] K. Muller-Dethlefs and P. Hobza. Noncovalent interactions: A challenge for experiment and theory. *Chemical Reviews*, 100:143–167, 2000.
- [241] A. Fersht, J. Shi, J. Knill-Jones, D. Lowe, A. Wilkinson, D. Blow, P. Brick, P. Carter, M. Waye, and G. Winter. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, 314:235–238, 1985.
- [242] W. Blokzijl and J. Engberts. Hydrophobic effects - opinions and facts. *Angewandte Chemie International Edition*, 32:1545–1579, 1993.
- [243] A. Davis and S. Teague. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angewandte Chemie International Edition*, 38:737–749, 1999.
- [244] L. Hicke, H. Schubert, and C. Hill. Ubiquitin-binding domains. *Nature Reviews Molecular Cell Biology*, 6:610–621, 2005.
- [245] J. Hurley, S. Lee, and G. Prag. Ubiquitin-binding domains. *Biochemical Journal*, 399:361–372, 2006.
- [246] I. Dikic, S. Wakatsuki, and K. Walters. Ubiquitin-binding domains - from structures to functions. *Nature Reviews Molecular Cell Biology*, 10:659–671, 2009.
- [247] R. Fisher, B. Wang, S. Alam, D. Higginson, H. Robinson, W. Sundquist, and C. Hill. Structure and ubiquitin binding of the ubiquitin-interacting motif. *The Journal of Biological Chemistry*, 278:28976–28984, 2003.
- [248] J. Lim, W. Son, J. Park, E. Kim, B. Lee, and H. Ahn. Solution structure of uim and interaction of tandem ubiquitin binding domains in stam1 with ubiquitin. *Biochemical and Biophysical Research Communications*, 405:24–30, 2011.

- [249] X. Periole, A. Knepp, T. Sakmar, S. Marrink, and T. Huber. Structural determinants of the supramolecular organization of g protein-coupled receptors in bilayers. *Journal of the American Chemical Society*, 134:10959–10965, 2012.
- [250] J. Chen, W. Im, and C. Brooks. Balancing solvation and intramolecular interactions. *Journal of the American Chemical Society*, 128:3728–3736, 2006.
- [251] M. Bhandarkar, A. Bhatele, E. Bohm, R. Brunner, F. Buelens, C. Chipot, A. Dalke, S. Dixit, G. Fiorin, P. Freddolino, P. Grayson, J. Gullingsrud, A. Gursoy, D. Hardy, C. Harrison, J. Henin, W. Humphrey, D. Hurwitz, N. Krawetz, S. Kumar, D. Kunzman, J. Lai, C. Lee, R. McGreevy, C. Mei, M. Nelson, J. Phillips, O. Sarood, A. Shinozaki, D. Tanner, D. Wells, G. Zheng, and F. Zhu. *NAMD User’s Guide Version 2.9*, pages 128–129, 2012.
- [252] M. Bhandarkar, A. Bhatele, E. Bohm, R. Brunner, F. Buelens, C. Chipot, A. Dalke, S. Dixit, G. Fiorin, P. Freddolino, P. Grayson, J. Gullingsrud, A. Gursoy, D. Hardy, C. Harrison, J. Henin, W. Humphrey, D. Hurwitz, N. Krawetz, S. Kumar, D. Kunzman, J. Lai, C. Lee, R. McGreevy, C. Mei, M. Nelson, J. Phillips, O. Sarood, A. Shinozaki, D. Tanner, D. Wells, G. Zheng, and F. Zhu. *Namd user’s guide version 2.9*. page 106, 2012.
- [253] L. Li, D. Bedrov, and G. Smith. Water-induced interactions between carbon nanoparticles. *The Journal of Physical Chemistry B*, 110:10509–10513, 2006.
- [254] J. Dunitz. Win some, lose some-enthalpy-entropy compensation in weak intermolecular interactions. *Chemistry & Biology*, 2:709–712, 1995.
- [255] S. Homans. Dynamics and thermodynamics of ligand-protein interactions. *Topics in Current Chemistry*, 272:51–82, 2007.
- [256] A. Ben-Naim. The rise and fall of the hydrophobic effect in protein folding and protein-protein association, and molecular recognition. *Open Journal of Biophysics*, 1:1–7, 2011.
- [257] A. Ben-Naim. Inversion of the hydrophobic/hydrophilic paradigm demystifies the protein folding and self-assembly of problems. *International Journal of Physics*, 1:66–71, 2013.
- [258] L. Li, I. Vorobyov, and T. Allen. Potential of mean force and pK_a profile calculation for the lipid membrane-exposed arginine side chain. *The Journal of Physical Chemistry B*, 112:9574–9587, 2008.
- [259] L. Li, D. Bedrov, and G. Smith. A molecular-dynamics simulation study of solvent-induced repulsion between c-60 fullerenes in water. *The Journal of Chemical Physics*, 123:204504(1–7), 2005.
- [260] R. Gregory. Protein-solvent interactions. *Marcel Dekker: New York*, 1995.

- [261] J. Finney and A. Soper. Solvent structure and perturbations in solutions of chemical and biological importance. *Chemical Society Reviews*, 23:1–10, 1994.
- [262] B Bagchi. Water dynamics in the hydration layer around proteins and micelles. *Chemical Reviews*, 105:3197–3219, 2005.
- [263] S. Ebbinghaus, S. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. Leitner, and M. Havenith. An extended dynamical hydration shell around proteins. *Proceedings of the National Academy of Sciences*, 104:20749–20752, 2007.
- [264] C. MacRaild, A. Daranas, A. Bronowska, and S. Homans. Global changes in local protein dynamics reduce the entropic cost of carbohydrate binding in the arabinose-binding protein. *Journal of Molecular Biology*, 368:822–832, 2007.
- [265] M. Karplus and J. Kushick. Method for estimating the configurational entropy of macromolecules. *Macromolecules*, 14:325–332, 1981.
- [266] S. Harris and C. Laughton. A simple physical description of dna dynamics: Quasi-harmonic analysis as a route to the configurational entropy. *Journal of Physics: Condensed Matter*, 19:076103(1–14), 2007.
- [267] D. McQuarrie. Statistical thermodynamics. *University Science Books: Mill Valley, CA*, 1973.
- [268] J. Song, L. Durrin, T. Wilkinson, T. Krontiris, and Y. Chen. Identification of a sumo-binding motif that recognizes sumo-modified proteins. *Proceedings of the National Academy of Sciences*, 101:14373–14378, 2004.
- [269] J. Song, Z. Zhang, W. Hu, and Y. Chen. Small ubiquitin-like modifier (sumo) recognition of a sumo binding motif. *The Journal of Biological Chemistry*, 280:40122–40129, 2005.
- [270] M. Praprotnik, S. Matysiak, L. Site, K. Kremer, and C. Clementi. Adaptive resolution simulation of liquid water. *J. Phys.: Condens. Matter*, 19:292201, 2007.

Appendix A

CLUSTER ANALYSIS OF LOW-HYDRATED RESIDUES OF VARIOUS
OF PROTEINS

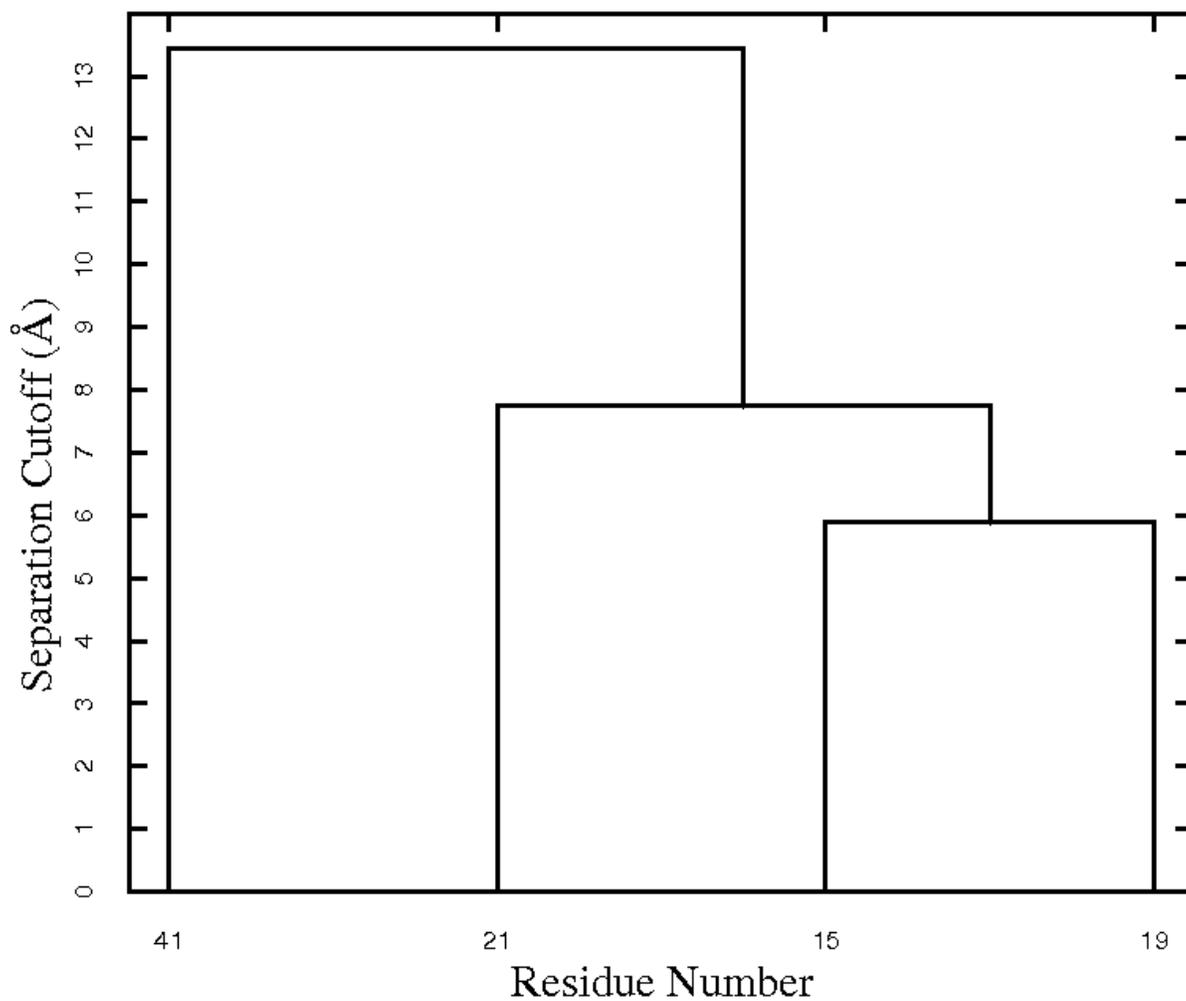


Figure A.1: Cluster analysis of low-hydrated residues of CUE domain

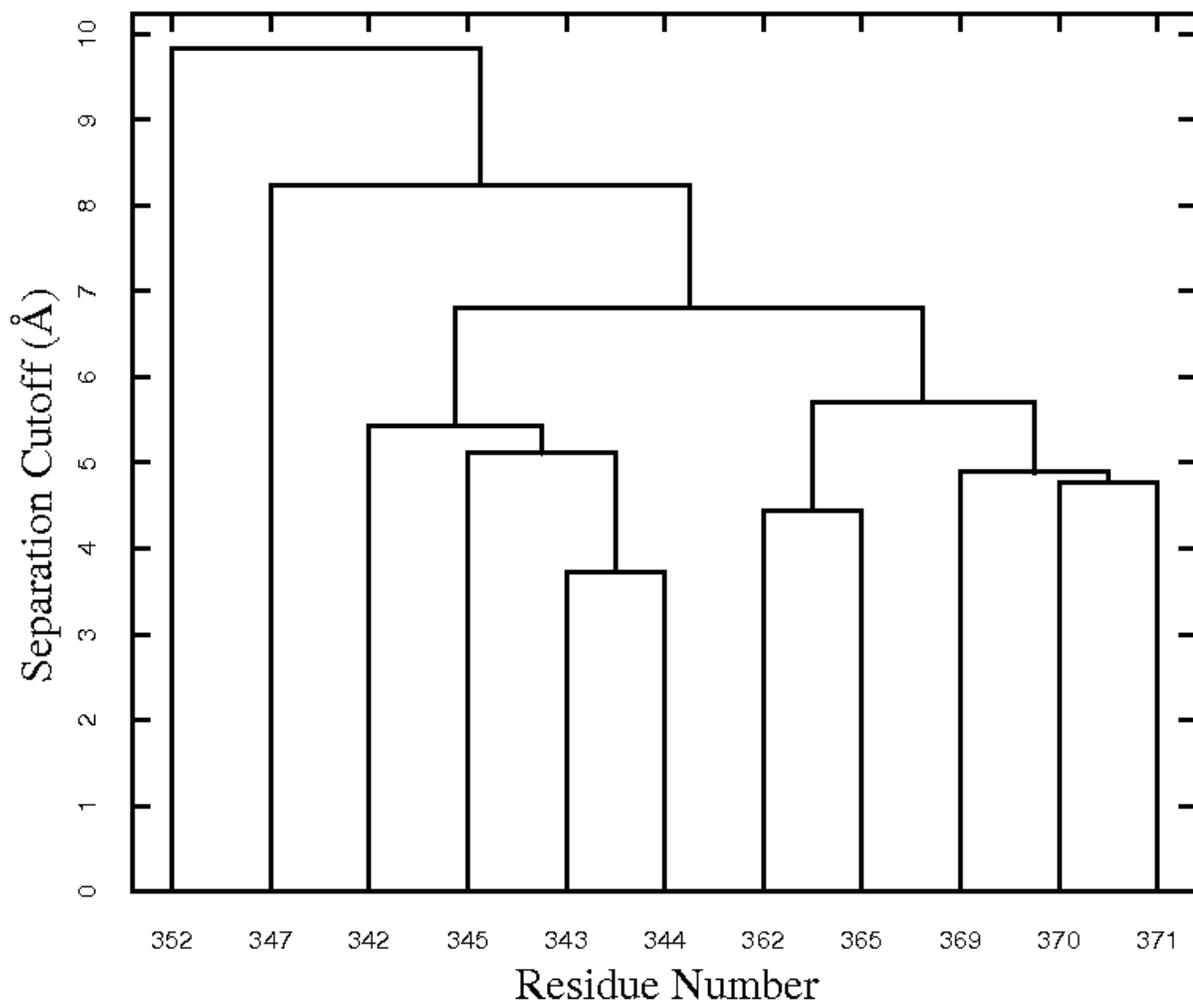


Figure A.2: Cluster analysis of low-hydrated residues of UBA of DSK2

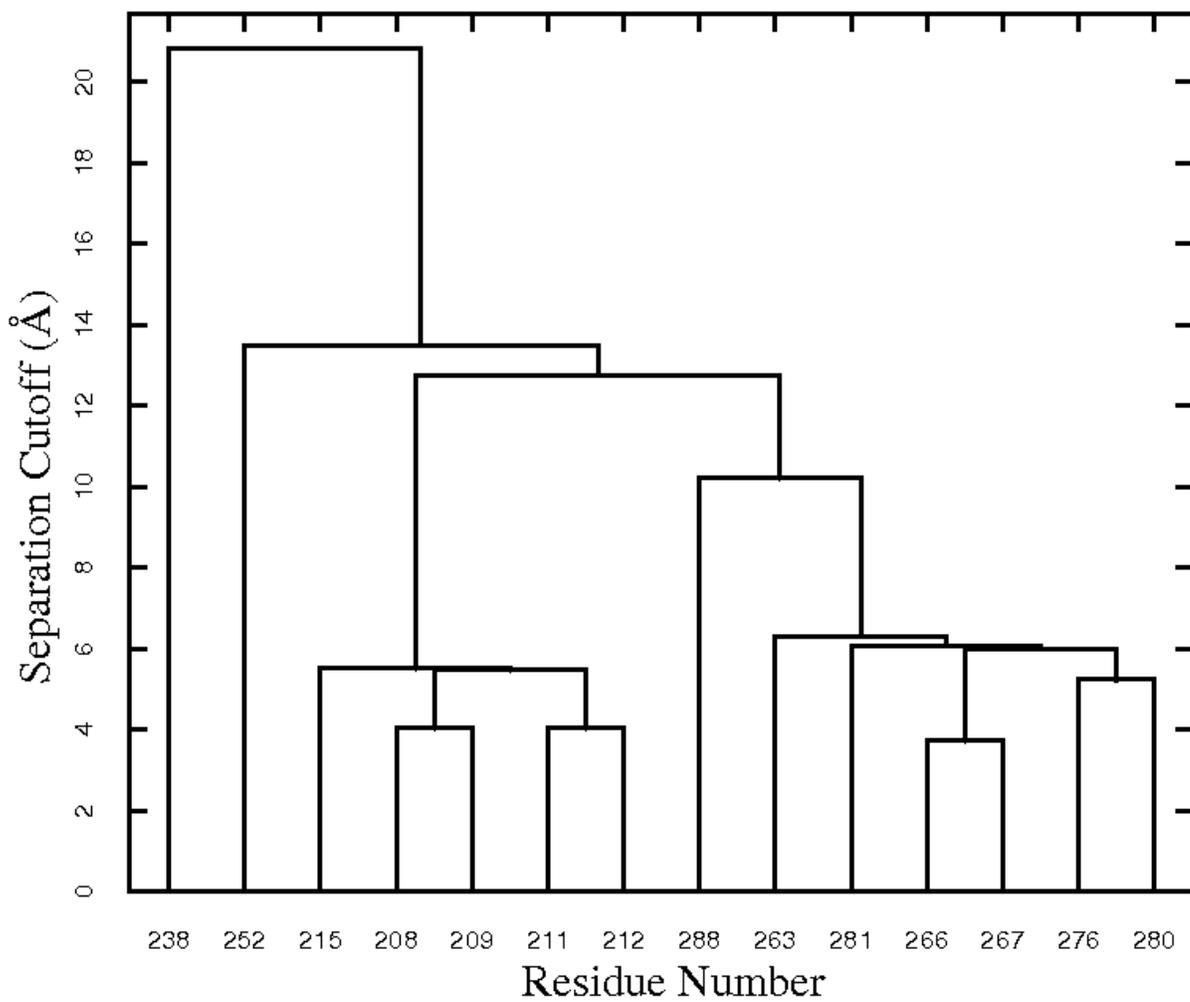


Figure A.3: Cluster analysis of low-hydrated residues of GGA3 GAT domain

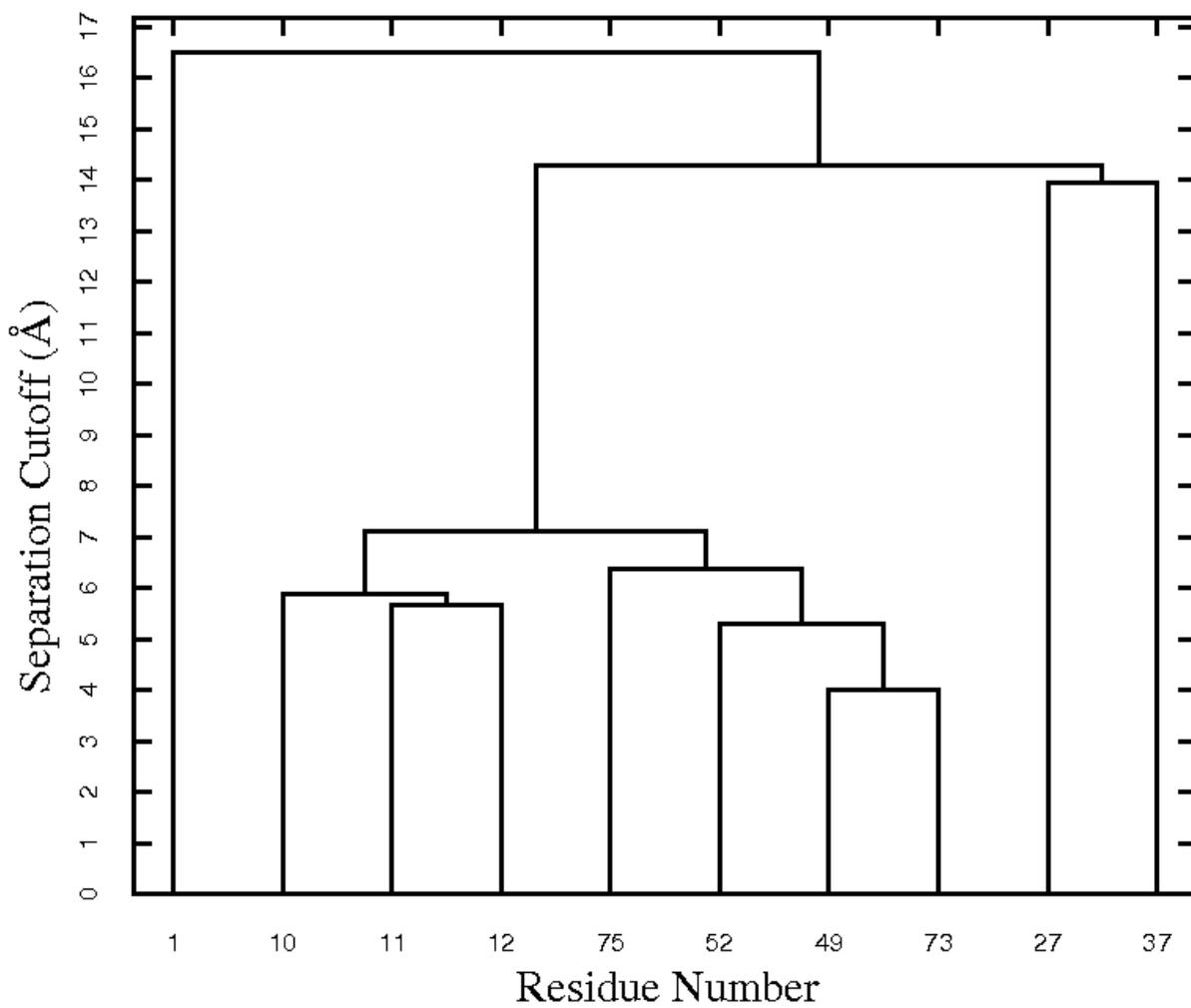


Figure A.5: Cluster analysis of low-hydrated residues of Ubl-domain of HHR23A

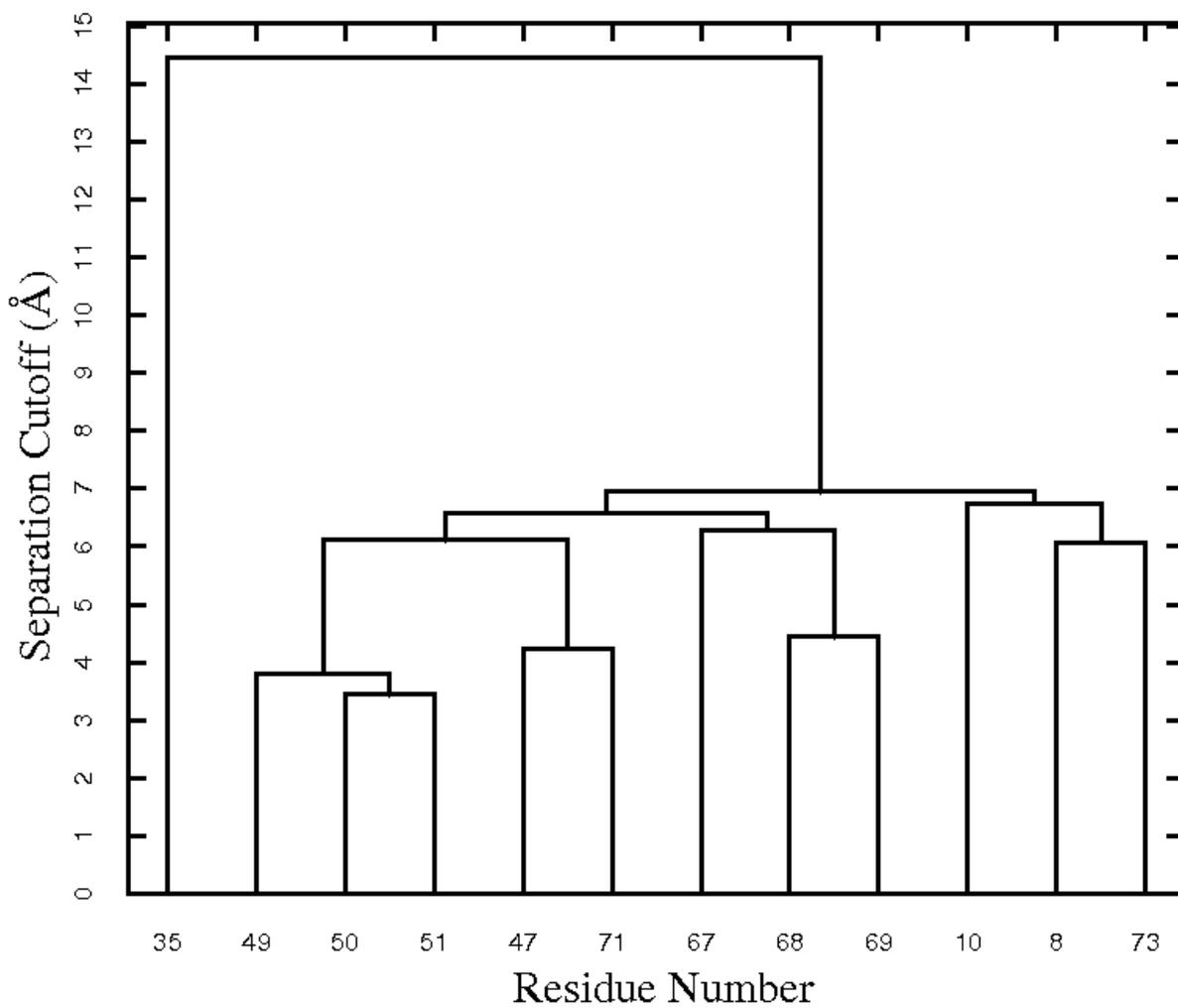


Figure A.6: Cluster analysis of low-hydrated residues of Ubl-domain of HHR23B

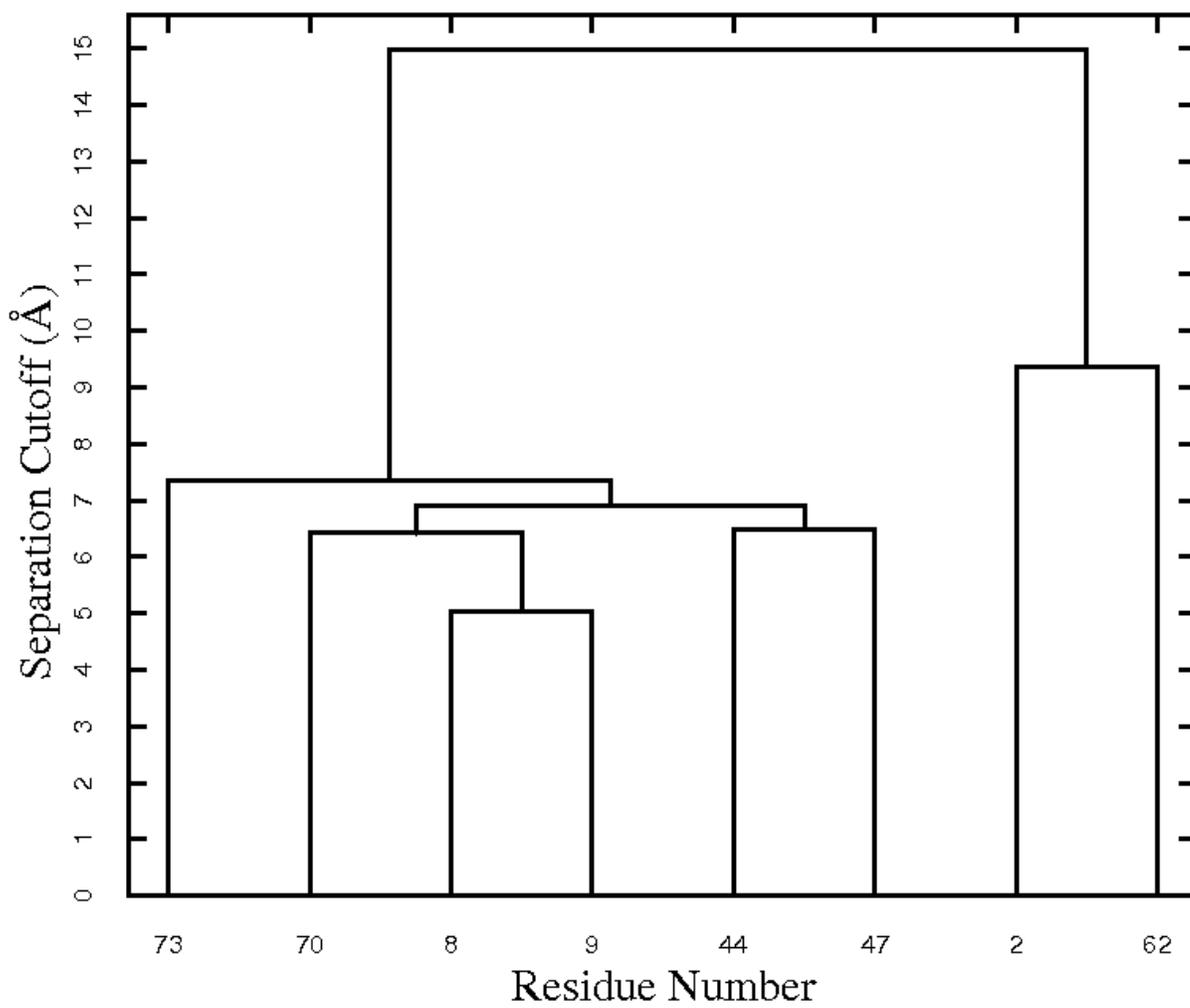


Figure A.7: Cluster analysis of low-hydrated residues of NEDD8

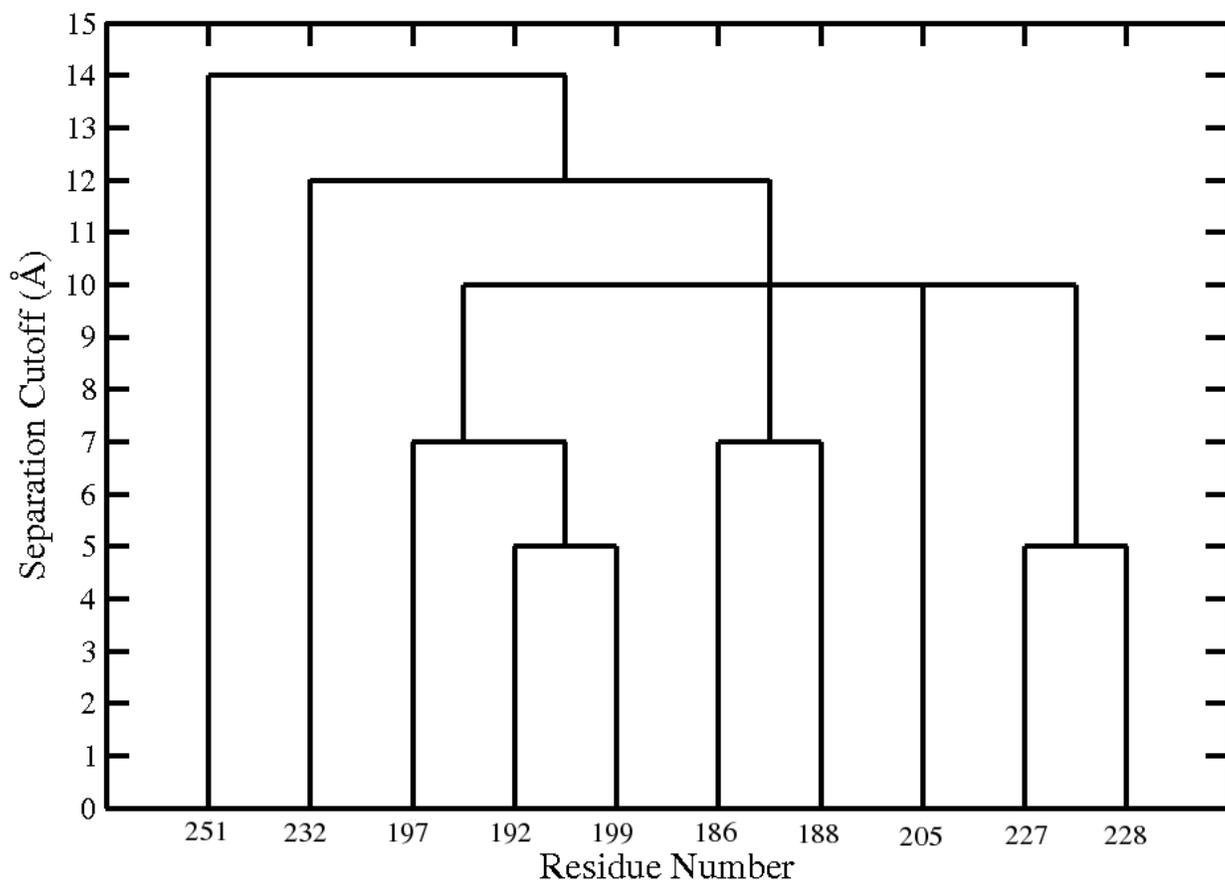


Figure A.8: Cluster analysis of low-hydrated residues of Pinch-1 LIM4 domain

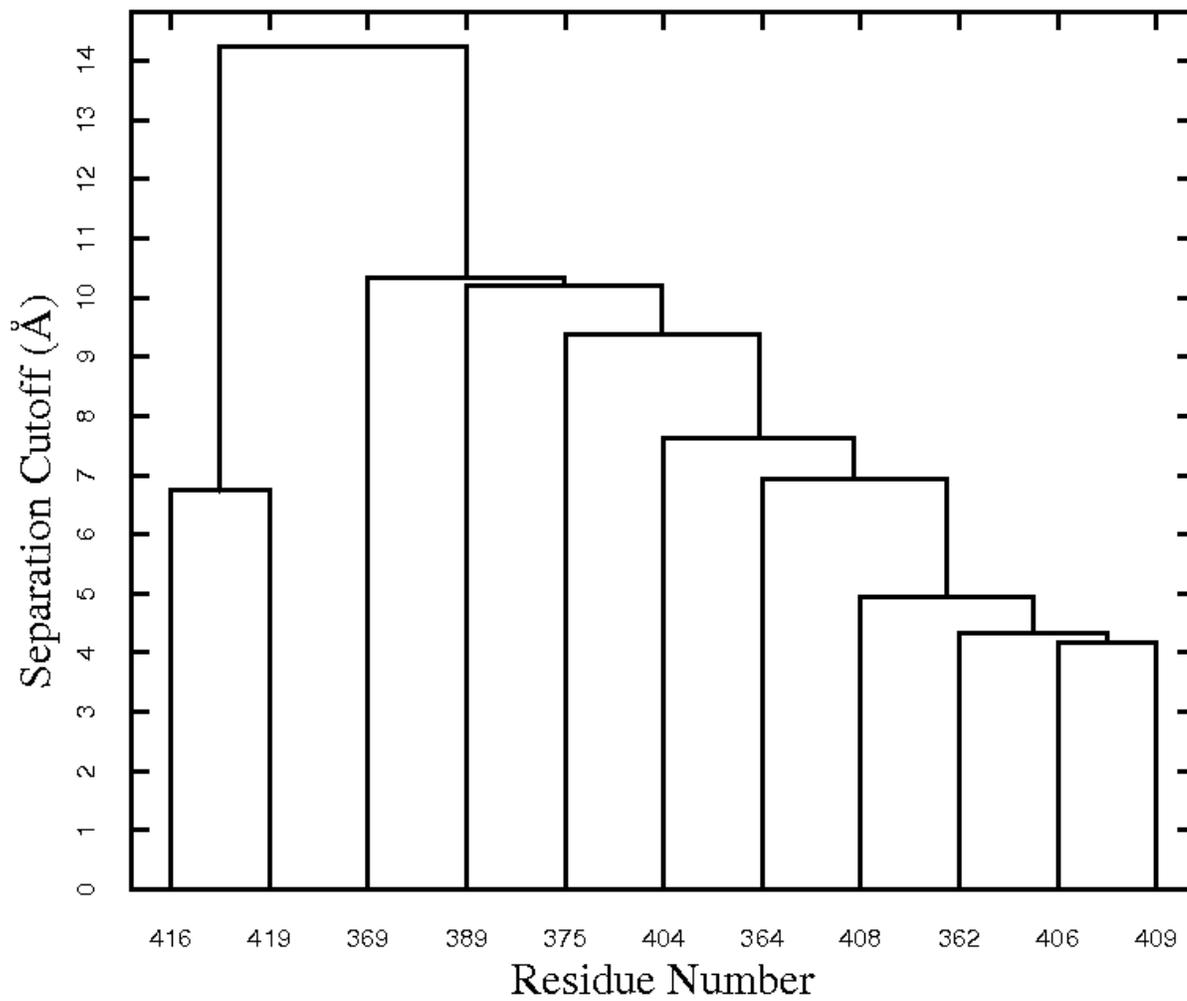


Figure A.9: Cluster analysis of low-hydrated residues of Sla1 SH3-3 domain

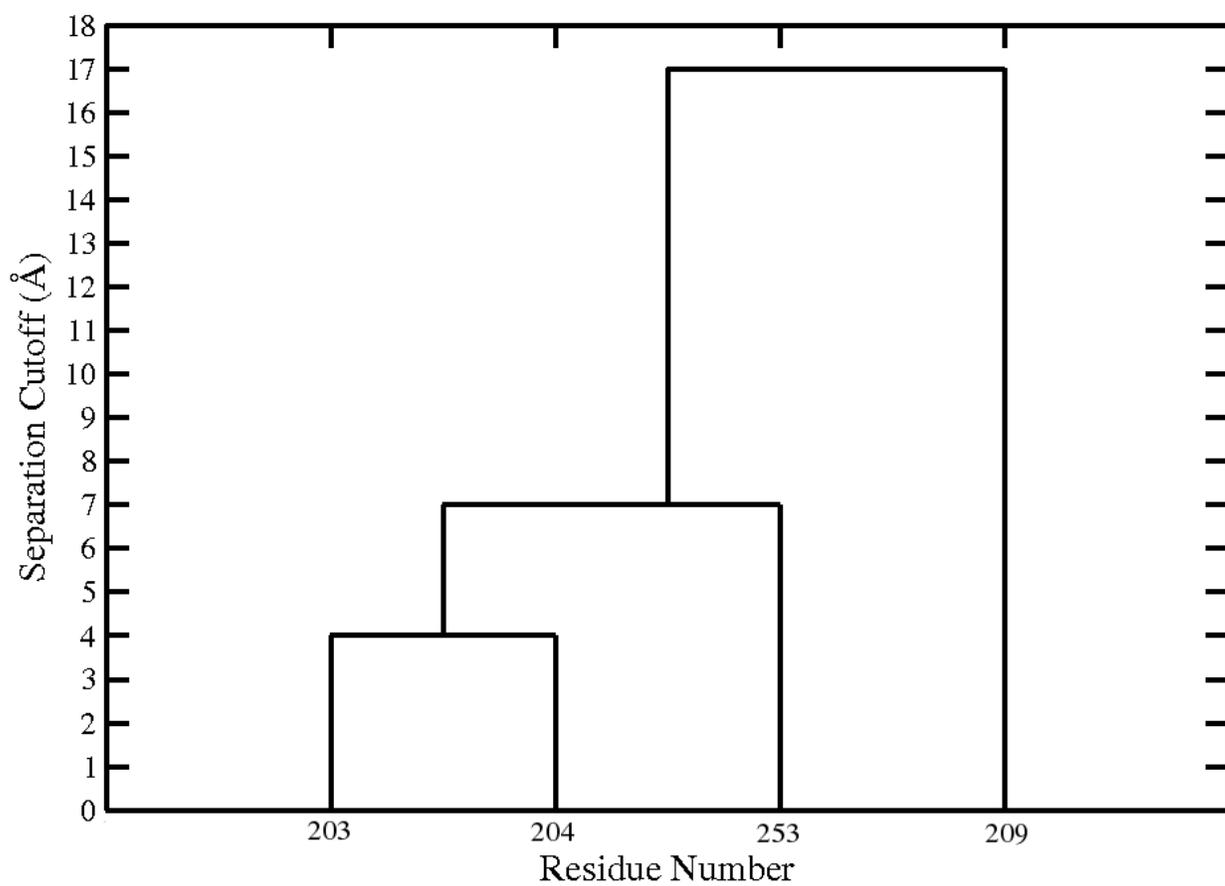


Figure A.10: Cluster analysis of low-hydrated residues of Nck-2 SH3 domain

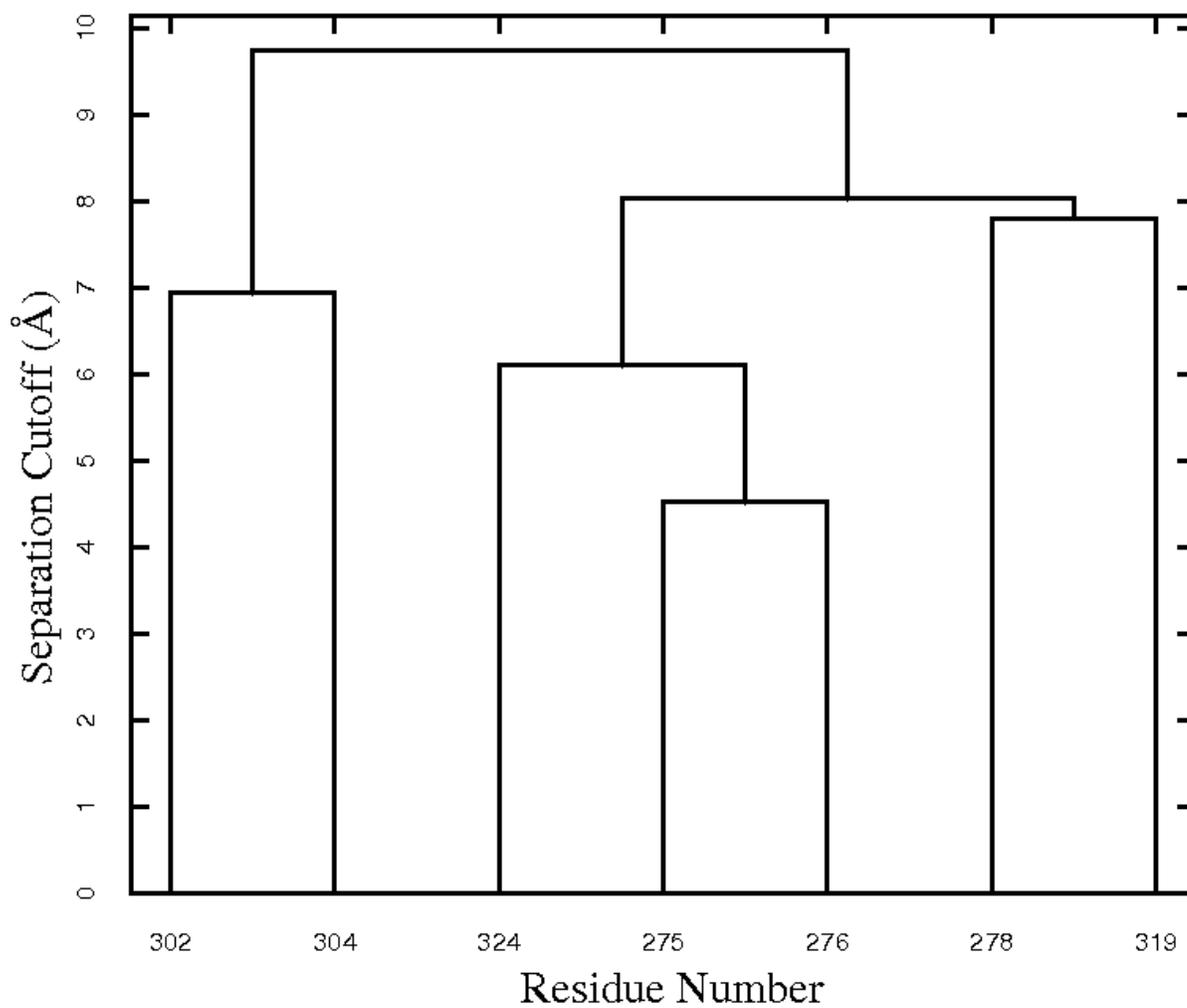


Figure A.11: Cluster analysis of low-hydrated residues of CIN85 SH3-3 domain

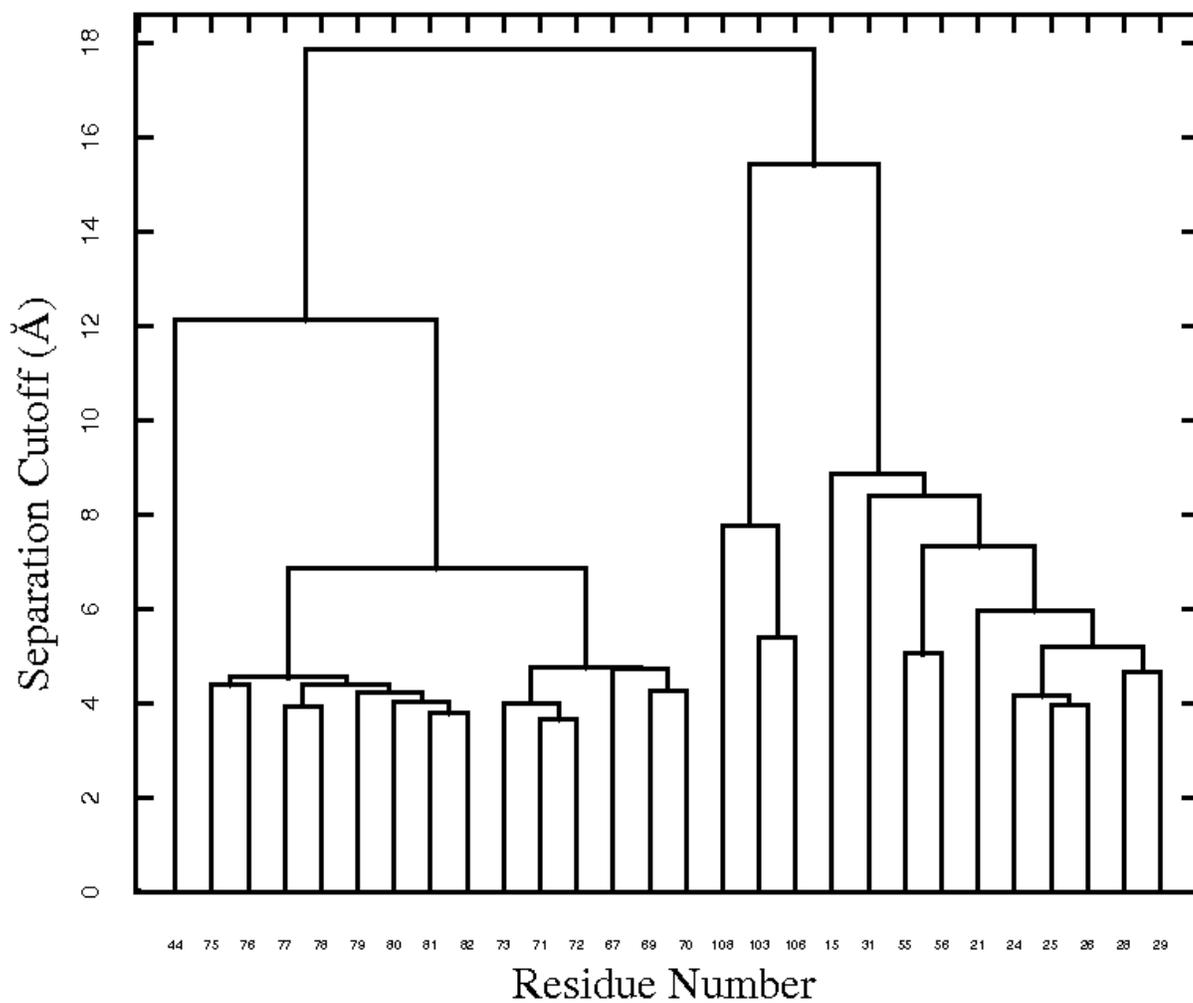


Figure A.12: Cluster analysis of low-hydrated residues of Crk SH2 domain

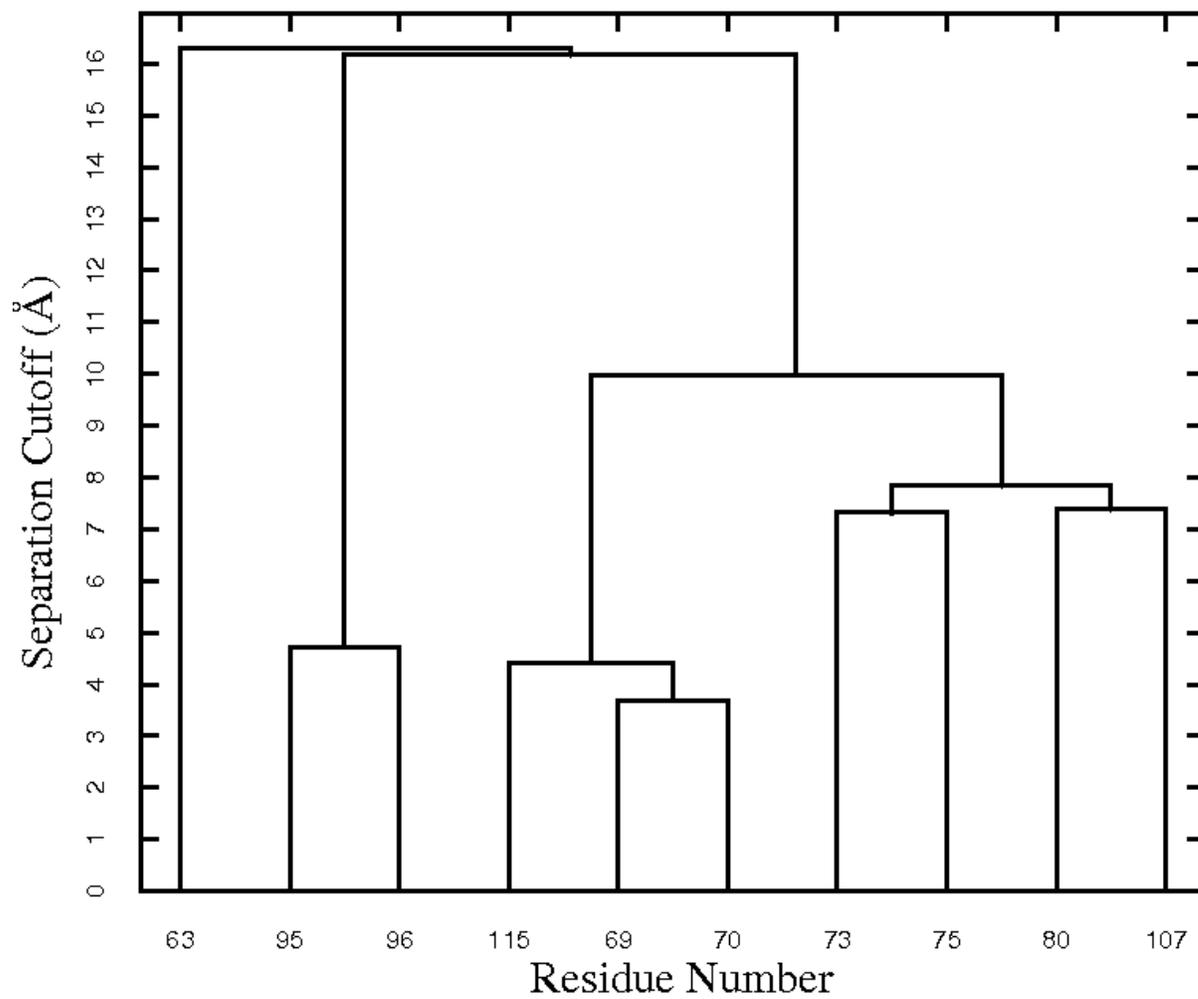


Figure A.13: Cluster analysis of low-hydrated residues of Abl SH3 domain

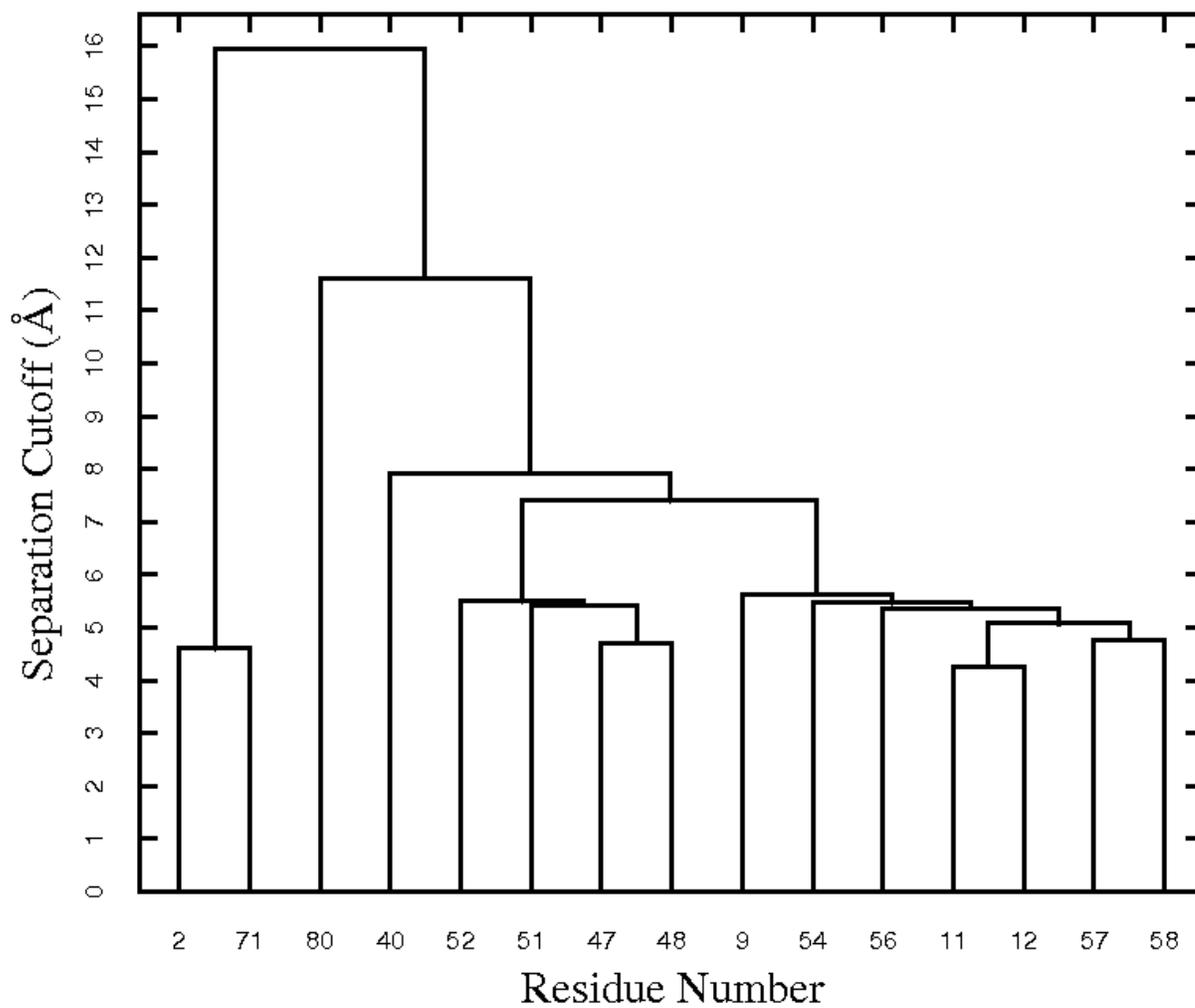


Figure A.14: Cluster analysis of low-hydrated residues of HPR

Appendix B

PERMISSIONS

Portions of this thesis are reproduced with permission from the following original research articles:

Di Cui, Shuching Ou, Sandeep Patel. "Protein-Spanning Water Networks and Implications for Prediction of Protein-Protein Interactions Mediated through Hydrophobic Effects." *Proteins: Structure, Function, and Bioinformatics*. **2014**, *82* (12), 3312-3326. Copyright 2014, Wiley Periodicals Inc.

Di Cui, Shuching Ou, Eric Peters, Sandeep Patel. "Ion-Specific Induced Fluctuations and Free Energetics of Aqueous Protein Hydrophobic Interfaces: Towards Connecting to Specific-Ion Behaviors at Aqueous Liquid-Vapor Interfaces." *Journal of Physical Chemistry B*. **2014**, *118* (17), 4490-4504. Copyright 2014, American Chemical Society

Di Cui, Shuching Ou, Sandeep Patel. "Protein Denaturants at Aqueous-Hydrophobic Interfaces: Self-Consistent Correlation between Induced Interfacial Fluctuations and Denaturant Stability at the Interface." *Journal of Physical Chemistry B*. **2015**, *119* (1), 164-178. Copyright 2014, American Chemical Society

Di Cui, Shuching Ou, Sandeep Patel. "Free Energetics of Rigid Body Association of Ubiquitin Binding Domains: A Biochemical Model for Binding Mediated by Hydrophobic Interaction." *Proteins: Structure, Function, and Bioinformatics*. **2014**, *82* (7), 1453-1468. Copyright 2014, Wiley Periodicals Inc.