

# UC Riverside

## UC Riverside Electronic Theses and Dissertations

### Title

Modeling Molecular Recognition From Thermodynamics to Kinetics: Affinity, Selectivity, Binding Rates and Binding Pathways

### Permalink

<https://escholarship.org/uc/item/1tc7k8sc>

### Author

Tang, Zhiye

### Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA  
RIVERSIDE

Modeling Molecular Recognition From Thermodynamics to Kinetics: Affinity,  
Selectivity, Binding Rates and Binding Pathways

A Dissertation submitted in partial satisfaction  
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Zhiye Tang

August 2016

Dissertation Committee:

Dr. Chia-en A. Chang, Chairperson

Dr. Gregory J. O. Beran

Dr. Jingsong Zhang

Copyright by  
Zhiye Tang  
2016

The Dissertation of Zhiye Tang is approved:

---

---

---

Committee Chairperson

University of California, Riverside

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation to my advisor, Dr. Chia-en Chang. The research environment in her lab encouraged free exploration on different possibilities and new ideas relying on the solid fundamental training. Her personal persistent characteristic also inspired me throughout the five year research and supported my progression as a graduate student. Many thanks also go to my committee members, Dr. Beran and Dr. Zhang. The suggestive and enlightening questions during my qualify examinations and oral presentations were invaluable for my research improvements and presentation skills. I would also like to thank Dr. Yadong Yin and Dr. Qiao Zhang, for sharing interesting research areas in nano science. My thanks also go to other collaborators. Finally, I would like to thank my group members who helped me with my research through the time we shared. I would like to especially thank Dr. Christopher Roberts in my group who inspired me by many of his works and discussions we had.

Published materials are used as follows:

Chapter 2: Zhiye Tang; Qiao Zhang; Yadong Yin; Chia-en Chang. The Facet Selectivity of Ligands on Silver Nanoplates: A Molecular Mechanics Study. *J. Phys. Chem. C*, **2014**, 118 (37), 21589–21598

## ABSTRACT OF THE DISSERTATION

Modeling Molecular Recognition From Thermodynamics to Kinetics: Affinity, Selectivity, Binding Rates and Binding Pathways

by

Zhiye Tang

Doctor of Philosophy, Graduate Program in Chemistry  
University of California, Riverside, August 2016  
Dr. Chia-en Chang, Chairperson

The overall picture of molecular recognition covers the thermodynamic and kinetic properties of molecular systems. Regarding thermodynamics, binding affinity or free energy, which can be decomposed into enthalpy and entropy, determines the strength of binding. Binding kinetics, the association and dissociation rate constants, describe the rates for molecular binding and unbinding. In this thesis, I carried out various molecular mechanics modeling tools to understand binding thermodynamics and kinetics of host-guest systems and protein-ligand systems.

The work includes both novel method development and applications using existing tools. Here we computed the binding free energy of the surfactant-silver nano-plate complexes using the M2 method and to understand how a capping ligand helps grow shaped nanomaterials. Several factors were determined to be crucial, and our findings allow rational design of capping surfactants in addition to the magic compound citrate. In

addition, a molecular dynamics (MD) and docking were applied to find the inhibitor binding site on protein TIM44 to assist cell biology studies. I examined the inhibitor selectivity of wildtype TIM44 and its mutants using energy calculations.

The binding/unbinding processes are carefully investigated for  $\beta$ -cyclodextrin and its guests to thoroughly study the thermodynamics and kinetics of binding. I applied MD simulations, post-analyzed the MD trajectories and developed and implemented methods such as the cell method to compute binding enthalpy, solute entropy and solvent entropy. Notably, the latter two entropy components are known to be very challenging to obtain computationally. I also computed the kinetic rate constants and investigated the influence of force fields used. It is the first work that illustrates all entropy and enthalpy components of binding in great details.

Realizing how crucial conformational sampling can be in answering fundamental scientific questions and valuable applications, a novel method to accelerate searching molecular conformations and ligand dissociation pathways was developed. The method uses a multi-layer internal coordinate and an internal PCA search algorithm to post-analyze a given MD trajectory and then guide motions of a molecular system. This algorithm is able to find possible dissociation pathways with high efficiency. Five examples are discussed to illustrate the functionality and capacity of this new method.

## Table of Contents

<b>CHAPTER 1 INTRODUCTION</b> .....	<b>1</b>
1.1 OVERVIEW.....	1
1.2 MODELING WITH MOLECULAR MECHANICS .....	3
1.2.1 Force Fields .....	3
1.2.2 Solvent Models.....	8
1.3 TECHNIQUES IN MOLECULAR MODELING .....	12
1.3.1 Minimization.....	12
1.3.2 Docking .....	14
1.3.3 Molecular Dynamics.....	17
1.3.3.1 Model Preparation.....	17
1.3.3.2 Molecular Dynamics .....	19
1.3.3.3 Acceleration Techniques.....	20
1.3.4 Post Analysis.....	23
1.3.4.1 Structural Analysis .....	23
1.3.4.2 Energetic Analysis .....	25
1.3.4.3 Principal Component Analysis.....	25
1.3.5 Free Energy and Entropy.....	27
1.3.6 Kinetics.....	29
1.4 REFERENCES .....	32
<b>CHAPTER 2 THE FACET SELECTIVITY OF LIGANDS ON SILVER NANOPLATES: A MOLECULAR MECHANICS STUDY</b> .....	<b>47</b>
2.1 INTRODUCTION.....	47
2.2 METHODS .....	51
2.2.1 Description of Systems .....	51
2.2.2 Force-Field Parameterization .....	51
2.2.3 Computational Details.....	53
2.2.4 Conformation Filtering.....	58

2.2.5	Explicit Solvent Validation of the Distance Dependent Dielectric Constant	58
2.2.6	Synthesis Method.....	59
2.3	RESULTS AND DISCUSSION .....	60
2.3.1	Overview of Ligands on the Ag Surface.....	60
2.3.2	Binding Preference of Ligand to the (111) and (100) Facets.....	65
2.3.3	Relative Binding Affinities of Various Ligands .....	67
2.3.4	Partial Charge Investigation .....	79
2.3.5	Design of Novel Surfactants.....	85
2.4	CONCLUSIONS .....	86
2.5	REFERENCES .....	88

<b>CHAPTER 3 A MODELING INVESTIGATION ON INHIBITOR MITOBLOCK-10 BINDING IN WILDTYPE AND MUTANTS OF TIM44.....</b>	<b>100</b>	
3.1	INTRODUCTION.....	100
3.2	METHODS .....	102
3.2.1	Overview .....	102
3.2.2	System Preparation.....	103
3.2.3	Molecular Dynamics.....	103
3.2.4	Docking .....	104
3.2.5	Mutations.....	105
3.2.6	MMPB/SA .....	105
3.3	RESULTS AND DISCUSSION .....	105
3.3.1	MB-10 Binding Site .....	107
3.3.2	Wildtype vs Mutants.....	108
3.3.3	The Shallow Binding Site.....	112
3.3.4	Investigation of Deeper Binding .....	115
3.4	CONCLUSIONS .....	116
3.5	REFERENCES .....	117

<b>CHAPTER 4 ACCURATE SOLVENT ENTROPY ESTIMATION: BEYOND THE IMPLICIT SOLVENT MODEL .....</b>	<b>122</b>
4.1 INTRODUCTION.....	122
4.2 ENTROPY FROM PAIR CORRELATION FUNCTION .....	124
4.2.1 Entropy of Fluids.....	124
4.2.2 Inhomogeneous Fluid Solvation Theory.....	126
4.3 TWO-PHASE THERMODYNAMICS MODEL .....	128
4.4 APPLICATION OF 2PT TO BULK WATER .....	131
4.4.1 Method .....	131
4.4.2 Results and Discussion .....	131
4.4.3 Summaries.....	135
4.5 CONCLUSIONS .....	135
4.6 REFERENCES.....	137
<b>CHAPTER 5 CALCULATION OF BETA-CYCLODEXTRIN AND GUEST BINDING ENTHALPY AND ENTROPY: FROM THERMODYNAMICS TO KINETICS .....</b>	<b>143</b>
5.1 INTRODUCTION.....	143
5.2 METHOD.....	147
5.2.1 System Selection and Parameterization .....	147
5.2.2 Molecular Dynamics Simulation .....	148
5.2.3 Enthalpy Calculation.....	148
5.2.4 Solute Entropy Calculation.....	149
5.2.5 Solvent entropy Calculation.....	150
5.2.6 Association (kon) and Dissociation (koff) Rate Calculation.....	153
5.3 RESULTS AND DISCUSSION .....	154
5.3.1 Thermodynamics .....	157
5.3.2 Kinetics .....	178
5.3.3 Results from q4md-CD Force Field.....	183
5.4 CONCLUSION .....	190

5.5 REFERENCES.....	191
<b>CHAPTER 6 PSIM: PATHWAY SEARCH GUIDED BY INTERNAL MOTIONS</b>	
.....	<b>205</b>
6.1 INTRODUCTION.....	205
6.2 METHOD .....	210
6.2.1 Method Overview .....	210
6.2.2 Computational Details.....	214
6.2.2.1 Multi-Layer Internal Coordinate Definition.....	214
6.2.2.2 Periodicity of Dihedral .....	218
6.2.2.3 Internal Coordinate Element Selection.....	219
6.2.2.4 Calculation of Eigenvalues and Eigenvectors .....	220
6.2.2.5 Conformational Search Algorithm.....	220
6.2.3 Molecular Systems .....	224
6.2.3.1 Alanine Dipeptide .....	224
6.2.3.2 HIV Protease .....	228
6.2.3.3 Beta-Cyclodextrin-2-Naphthyl Ethanol.....	228
6.2.3.4 Cryptophane-Tetramethyl Ammonium .....	229
6.2.3.5 Tryptophan Synthase .....	231
6.3 RESULTS AND DISCUSSION .....	232
6.3.1 Performance of the Multi-Layer-Internal Coordinate Definition .....	233
6.3.2 Pathway of Conformation Transitions for Alanine Dipeptide.....	236
6.3.3 Comparison with MD Conformations .....	238
6.3.4 Tetramethyl Ammonium Unbinding from Cryptophane-E. Hist: Finding the Pathways and the Corresponding Free Energy Barriers .....	242
6.3.5 Application to Protein-Ligand System .....	245
6.4 GENERAL COMMENTS .....	247
6.5 CONCLUSIONS .....	248
6.6 REFERENCES .....	249

<b>CHAPTER 7 CONCLUSION AND FUTURE DIRECTIONS IN PROTEIN- LIGAND DISSOCIATION RESEARCH.....</b>	<b>263</b>
7.1 CONCLUSION .....	263
7.2 FUTURE WORKS.....	265
7.2.1 Cyclin-Dependent Kinase 8 (CDK8) Dissociation Pathway.....	265
7.2.2 Improvement of Internal PCA Search Method .....	269
7.3 REFERENCES .....	271

## List of Tables

Table 2.1. Cambridge Crystallographic Data Centre (CCDC) IDs for small molecules. The CCDC IDs of the crystal structures used to modify the bond angles of the ligands. ....	52
Table 2.2. Binding free-energy calculations and experimental data for nine ligands. Data were grouped by number of carboxylate groups. Experimental yields <sup>29</sup> were obtained from the experimentalists. $\Delta G_{\text{comp}}$ : binding affinity of ligands on (111) facet; $\Delta\Delta G_{\text{comp}}$ : relative binding affinity of ligands on (111) facet within each group. (For $\Delta\Delta G$ , the reference for Group 2 is succinate and for Group 3 citrate.) .....	62
Table 2.3. Binding energy components of ligands. The unit of this table is kcal/mol. The last column, $\Delta G_{111}-\Delta G_{100}$ , indicates the ability of the ligands to distinguish the two Ag facets, where $\Delta G_{111}$ and $\Delta G_{100}$ represent the binding free energy of the same ligand to the (111) and (100) facet, respectively. ....	64
Table 2.4. Computed distances between oxygen and Ag atoms. The average oxygen–Ag distance was calculated by averaging the distances between the oxygen atoms of the ligands and the three Ag atoms that are closest to each of the oxygen atoms. The unit is Å. The distances were consistently shorter on the (111) than (100) facet. On the (111) facet, the distances decrease from oxalate to pimelate. This finding can be explained by more flexible molecules fitting the periodic surface better than less flexible ones. As well, because of this factor, oxalate and malonate are unfavorable to the (111) facet. ....	67
Table 2.5. Numbers of water in the region that is the first layer (here cut off at 3.2 Å from the surface) water that is within 10 Å from the projection of center of geometry of the ligand on silver surface of silver plate-ligand complexes.....	71

Table 2.6. The partial charges (Vcharge model) on oxygen atoms in the ligands. The values of partial charges decrease because as a molecule becomes larger, the -2.0 formal charges must be distributed on more atoms, and thus each of the oxygen atoms carries less negative charges.....	76
Table 2.7. Partial charges from other calculation methods. The ccpvDZ-b3lyp is <i>ab initio</i> calculated, whereas the remainder are semi-empirical methods. Multiple values indicate that the partial charges for oxygen atoms slightly differed from one another.....	77
Table 2.8. Computed configuration entropy loss. The entropy loss increases as the ligands become more flexible. The table lists numbers of free ligand conformations and the bound complexes within 3 kcal/mol of the global energy minimum.....	78
Table 2.9. Binding energy components of ligands with +0.2 partial charge set. The unit of binding energy terms and entropy loss are kcal/mol. ....	80
Table 2.10. Binding energy components of ligands with +0.3 partial charge set. The unit of binding energy terms and entropy loss are kcal/mol. ....	81
Table 2.11. Binding energy components of ligands with +0.4 partial charge set. The unit of binding energy terms and entropy loss are kcal/mol. ....	82
Table 3.1. The binding energy decomposition from MMPB/SA of MB-10 to wildtype TIM44 and its four mutants. The units are in kcal/mol. ....	108
Table 3.2. The MMPB/SA results of the simulations starting from the shallow binding site. Frame 82, 112, 136 and 138 in the accelerated MD were used in the first round docking, and the lowest energy conformations were used for conventional MD simulation. MMPB/SA calculations were done on the corresponding trajectories. The units are kcal/mol, and the standard deviations are indicated in the parenthesis.....	114

Table 4.1. Calculated molar entropies from 12 ns equilibrium results. The deviation is from the experimental data (69.95 J/mol/K). The deviation of 1000 water is calculated by multiplying the previous column by 1000. ....	134
Table 5.1. Calculated and experimental free energies, enthalpies and entropies of binding. The standard deviations of mean of the enthalpy changes are shown in parentheses. All values are in kcal/mol. ....	155
Table 5.2. Calculated and experimental $k_{on}$ , $k_{off}$ and $K$ . The units of $k_{on}$ is in $s^{-1}M^{-1}$ , $k_{off}$ in $s^{-1}$ and $K$ in $M^{-1}$ . ....	156
Table 5.3. The decompositions of binding enthalpies of the complexes. All values are in kcal/mol. ....	158
Table 5.4. The calculated and experimental values of the thermodynamic properties with excluded volumes for each complex. Thermodynamic values are in kcal/mol. Excluded volumes are in $\text{\AA}^3$ . ....	168
Table 5.5. The entropy decompositions of the complexes. All values are in kcal/mol. .	171
Table 5.6. The bound and free period lengths of the complexes. All values are in nanosecond. ....	179
Table 5.7. Calculated and experimental free energies, enthalpies and entropies of binding in q4md-CD force field. The standard deviations of mean of the enthalpy changes are shown in parentheses. All values are in kcal/mol. ....	184
Table 5.8. The decompositions of binding enthalpies of the complexes in q4md-CD force field. All values are in kcal/mol. ....	185
Table 5.9. The entropy decompositions of the complexes in q4md-CD force field. All values are in kcal/mol. ....	186

Table 5.10. Calculated and experimental $k_{on}$ , $k_{off}$ and $K$ from q4md-CD force field. The units of $k_{on}$ is in $s^{-1}M^{-1}$ , $k_{off}$ in $s^{-1}$ and $K$ in $M^{-1}$ . $\beta$ -cyclodextrin-2-naphthyl ethanol complex was in bound state through the trajectory and therefore no kinetic properties were obtained. ....	187
Table 6.1. The details of parameters used in the PSIM search for all systems.....	224
Table 6.2. The details of transition pathways. Searching directions are listed in order from top to bottom. Mode in use, its direction and number of steps are listed. ....	237

## List of Figures

Figure 1.1 Dihedral torsion angle $\varphi$ defined by a set of four sequentially bonded atoms i, j, k and l.....	6
Figure 1.2 Improper torsion angle $\varphi$ defined by a set of three atoms j, k and l bonded to the same atom i. ....	6
Figure 2.1. Flow chart of applying M2 method to silver plate-ligand systems. ....	55
Figure 2.2. Top view of the bound geometry of the ligands on two silver (Ag) facets (top for [111] facet, bottom for [100] facet). a) acetate, b) oxalate, c) malonate, d) succinate, e) glutarate, f) pimelate, g) citrate, h) isocitrate, i) trimesic acid, j) designed ligand #1, k) designed ligand #2. The oxygen atoms are color mapped to represent the average distance between the oxygen atom and the closest three Ag atoms according to the color bar (in Å). ....	61
Figure 2.3. Top-down view of the (111) (left) and (100) facet (right). The atom density is greater for the (111) than (100) facet. ....	66
Figure 2.4. Plots of RMSD (in Å) computed by all atoms, including hydrogen atoms, for succinate from MD simulations. Each plot used one local energy minima found by M2 as a reference structure. . The frame was saved every 1 ps, and the MD simulation length is 5 ns. ....	72
Figure 2.5. Binding pose of glutarate on silver plate in +0.2, +0.3 and +0.4 partial charge model.....	84
Figure 2.6. Binding pose of trimesic acid on silver plate in +0.2 partial charge model. ..	84
Figure 3.1. Structure of MitoBloCK -10 (MB-10). ....	103
Figure 3.2. The MB-10 binding site in TIM44 wildtype structure, which is located differently from the PEG binding site.....	106

Figure 3.3. The deep binding pose of MB-10 in the wildtype TIM44. The residues within 9 Å of the ligand is rendered in orange. The ligands which blocks the binding site are shown with licorice representation. The ligand is shown in vdW representation. ....	108
Figure 3.4. The convergence of MMPB/SA calculation. The average accumulative binding affinity over number of conformations was plotted against the number of conformations. ....	109
Figure 3.5. The binding pose of the wildtype of TIM44 and the mutants. For comparison, the wildtype is shown for each mutant. The mutation site residues are rendered in red, and other relevant residues are in gray. ....	110
Figure 3.6. The shallow MB-10 binding site. It is on the surface of the deep MB-10 binding. ....	113
Figure 4.1. The calculated molar water entropy from 2PT plotted against number of water molecules in the water box. The black rectangle indicates the results from 6 ns equilibration, and red dots for the 12 ns equilibration. The blue triangle marks the experimental molar water entropy (69.95 J/mol/K). The second order fit curve for the 2PT results is only shown for the data from 6 ns equilibrium. ....	132
Figure 4.2. The molar water entropy (6 ns) from each trajectory plotted against the H-O-H angle RMSD in each trajectory. Noticeably, the correlation between these two quantities is 0.9885. ....	133
Figure 5.1. The dihedral used to define the macrocyclic ring conformationa of $\beta$ -cyclodextrin (Blue). The $\beta$ -cyclodextrin is a macrocyclic compound consisting of 7 glucopyranose units. The glucopyranose unit is more rigid than the single bonds linking them so the 14 dihedral rotations were selected to characterize the macrocyclic ring conformation. ....	150

Figure 5.2. The definition of a loose hydrogen bond. It is based on normal hydrogen bond definition except that the complementary angle of X-H...Y is 80° and the distance between H and Y is 2.65Å. ....	151
Figure 5.3. The conformational entropy definition of water. It is based on the Pauling's model, where the donor waters (DW) and acceptor waters (AW) are defined with the loose hydrogen bond definition and any water molecule that is within 3.2 Å from the oxygen atom of the target water molecule but cannot form a loose hydrogen bond is considered as a surrounding water (SW). ....	153
Figure 5.4. The convergence of the trajectories represented by average potential energy plotted against the number of frames. ....	160
Figure 5.5. The plot of average volume and average potential energy against number of frames in TIP3P solvent model. After convergence, which takes roughly 4000-5000 frame, the average volume and average potential show same trend. Both of them converge at a higher value, and then drop to a lower value. ....	163
Figure 5.6. The two binding poses of β-cyclodextrin-1-butanol complex. In Pose 1, the ligand stays in the middle of the cavity of β-cyclodextrin and the space between the two molecules is too small for any water molecule to sneak into and thus becomes vacuum. On the other hand in Pose 2, the ligand sticks to the surface of the β-cyclodextrin and the cavity of β-cyclodextrin is closed up by glucopyranose unit flipping. ....	164
Figure 5.7. The plot of average volume and average potential energy against number of frames of β-cyclodextrin-1-butanol complex in TIP4P solvent model. ....	165
Figure 5.8. The RMSDs in reference with the crystal structure plotted against the frame number. The RMSD is in Å. The representative conformations are shown above the plots and circled on the plots. ....	170

- Figure 5.9. The water entropy decompositions near the region of  $\beta$ -cyclodextrin. From top to bottom are pure water, free  $\beta$ -cyclodextrin, free 1-butanol, and  $\beta$ -cyclodextrin-1-butanol complex. From left to right are total entropy, vibrational entropy (translational + vibrational), translational entropy, rotational entropy and conformational entropy. Generally brighter means higher water entropy and vice versa. For each decomposed term the reference is the corresponding pure water entropy. The red areas are regions filled by no water molecules due to the presence of the solute molecules..... 174
- Figure 5.10. The translational and rotational entropies near the region of free 2-naphthylbutyrate. The solute is positioned as the formula below in the two entropy images. The red spots in the images correspond to the hydroxyl group in the molecule. .... 176
- Figure 5.11. The RMSDs in reference with the crystal structure plotted against the frame number in q4md-CD force field. The RMSD is in Å. The representative conformations are shown above the plots and circled on the plots..... 189
- Figure 6.1. The flowchart of internal Principal Component analysis (internal PCA). ... 210
- Figure 6.2. A representative systematic search tree of 5 PC modes. Disabled PC modes are rendered in gray dashed lines, while all other PC modes in use are in black solid lines. Only representative sub branches are shown. The blue arrowed curves show a representative flow of the search on the tree diagram. .... 211
- Figure 6.3. The flowchart of each search branch..... 212
- Figure 6.4. Classic internal coordinate representation (Z-matrix) for a small molecule. Atom 1 is at the origin; atom 2 is defined by a bond length ( $b_2$ ) and placed on the z-axis; and atom 3 is defined by a bond length ( $b_3$ ), bond angle ( $a_3$ ) and placed on the x-z plane. Six external coordinates are eliminated, and a molecule of N atoms has  $3N-6$  internal coordinates. Atoms 1-3 are termed root atoms for this

molecule. For atoms  $i > 3$  are defined by  $(b_i, a_i, t_i)$ , where  $t$  are torsion angles. For example, atom 4 is presented by  $(b_4, a_4, t_4)$ . The dashed light gray shows the molecule after rotating the bond between Atom 2 and 3 ( $t_4$ ) for 180 degrees. .. 214

Figure 6.5. Two peptide chains divided into three fragments, in blue, green and yellow. The atoms, N, CA, C, N, in root dihedrals in the fragments are circled. Pseudo bonds connecting the fragments are rendered in purple. The peptide bond (C-N) connecting the blue and green fragments is rendered in red. In the internal coordinates, the peptide bond (in red) connecting blue and green fragments are removed. For this reason, artifacts may arise from the absence of the chemical bond between groups. Instead, the pseudo bonds are used to connect the fragments. For example, between blue and yellow fragments, the N-N (4-5) pseudo bond connects the two fragments, and along the pseudo bond, two pseudo angles, 3-4-5 and 4-5-6, and three pseudo dihedrals, 2-3-4-5, 3-4-5-6 and 4-5-6-7 are also included. These six pseudo degrees of freedom are used as the linker between fragments. .... 216

Figure 6.6. An example of fragmentation of HIVp. each monomer was divided into 21 fragments. The primary fragment,  $1_{\text{fragment}}$ , on monomer 1 is colored in blue and on monomer 2 is in red. The rest fragments on monomer 1 are in light and dark green and on monomer 2 are in light and dark magenta. The greenish fragments on monomer 1 are defined to the blue primary fragment, instead of mutually connected. Similarly the magentaish fragments are defined to the red primary fragment. Finally, the blue fragment on monomer 1 is chosen as the starting point of the internal coordinate, and the red fragment on monomer 2 is defined to it. 217

Figure 6.7. The four transition path found by PSIM. The search started from Minimum A, and leded the transitions to different minima starting from four directions including the positive and negative of PC mode 1 and 2. The left views of each path show the conformational change of the original path using PC mode 1 (Green) and 2 (Yellow). The middle views of each path are the original path

(Black bordered dots) and the minimized path (White bordered dots) on PES of alanine dipeptide. The right views show the search path on PES with PC mode 1 in green and PC mode 2 in yellow. .... 226

Figure 6.8. Structure of the  $\alpha$ -subunit of tryptophan synthase. Loop 6 and Loop 2 are shown in orange and red. Bound state G3P is shown in green. .... 232

Figure 6.9. A comparison between a)  $\alpha$ -carbon PCA scaled by 2.67, b) dihedral PCA by using multi-level internal coordinate definition scaled by 1.0 and c) dihedral PCA by using standard linear type internal coordinate definition of HIVp scaled by 0.2. In a) and b) the motions of the HIVp have very high similarity, which the two monomers have scissoring motion with the two flaps opening and closing and the bottom part of the HIVp mostly is stationary. However, in c), the left monomer is stationary because the primary root of the internal coordinate is on the left monomer while the right monomer has overall translational and rotational motion because it is defined through the mobile loops of left monomer. The overall asymmetric behavior is the artifact from the accumulated dependence of dihedrals in the internal coordinate. .... 234

Figure 6.10. The RMSF of first dihedral PCA mode from multi-level internal coordinate definition (black) and standard internal coordinate definition (gray) in reference to the first standard  $\alpha$ -carbon PCA mode. Residues with RMSF greater than 0.5 Å are in blue, and that greater than 1.5 is in red, in the top structure and circled in the plot. .... 235

Figure 6.11. The comparison between MD dissociation events and dissociation events from PSIM. Trajectories are classified by number of glucose ring flipping in  $\beta$ -cyclodextrin, where a) is no significant ring flipping, b) is one ring flipping and c) is multiple ring flipping. The bound and free conformations of  $\beta$ -cyclodextrin are shown on the left and in the middle of the figure. The plots of 7  $\beta$ -cyclodextrin conformation fingerprint angles along trajectories are shown on the right.

Corresponding glucose ring are labeled on the conformations using the same color as in the plots. The bound state conformation (rectangle), free state conformation (triangle) and moment of dissociation (hexagon) are indicated on the plots. .... 239

Figure 6.12. The original dissociation pathways found by PSIM, Hopping Minima, and simply pulling the ligand from one direction. The dissociation path is shown in yellow..... 243

Figure 6.13. The post processed snapshots of crptophane-tetramethyl ammonium in the bound state (a) and the moment when the ligand cross the dissociation window (b). The ligand is shown in vdW representation. .... 243

Figure 6.14. This is the free energy plot of the dissociation pathways constructed based on the trajectory from PSIM. .... 244

Figure 6.15. Tryptophan  $\alpha$ -subunit-G3P dissociation pathway found by the Internal PCA Search method. The protein is shown in gray with Loop 6 highlighted in orange. The ligand is shown in green. Important residues shown in Figure 6.8 are also shown in this figure. In the process of the dissociation path found by the search, the protein maintains its native conformation except that the Loop 6 moves away from the native state and opens the dissociation pathway for the G3P. The G3P escapes from the space originally hindered by Loop 6..... 245

Figure 7.1. The structure of CDK8. The N-lobe is rendered in orange, the C-lobe is rendered in green, the activation loop is rendered in purple, and the DMG motif is rendered in blue. The ATP binding site and the allosteric binding site are on the left and right of the DMG motif respectively. .... 266

Figure 7.2. The dissociation pathway from internal PCA search. a) is the bound state, b) is the moment when the ligand cross the activation loop, and c) is the final state where the ligand stays in the allosteric binding site. The N-lobe is rendered in orange, the C-lobe is rendered in green, and the ligand is rendered in cyan. .... 267

Figure 7.3. The snapshots from MD resampled from the internal PCA search pathway. a) and b) are the intermediate state from two perspectives of view. c) and d) are the breathing conformation of the CDK8 found by conventional MD from two perspectives of view. a) and c) are aligned, and b) and d) are aligned. The color pattern is the same as in Figure 7.2..... 268

Figure 7.4. The comparison between MD (a) and internal PCA search (b) conformations. Both conformations are for the breathing motion of CDK8. The N-lobe of CDK8 is rendered in orange, and C-lobe in green. The hinge between the two lobes is more stretched in the internal PCA search result but more natural in the MD simulation..... 270

## **Chapter 1 Introduction**

### 1.1 Overview

Molecular modeling of ligand-receptor systems is the study which focuses on the microscopic structures and dynamics that determine the macroscopic properties including selective binding affinity, association/dissociation rates and binding pathways. Its research combines the concepts and techniques from chemistry, biology, computer science and pharmacology. It originated from the early models of ball and stick model and evolved into modern molecular modeling with the appearance and development of computer science, especially since the 21th century <sup>1-3</sup>. Nowadays it is an important component in fundamental studies in chemistry and biology, playing both explanatory and predictive roles. The objective of this work is to understand the binding process of bimolecular systems from the aspects of thermodynamic and kinetic properties such as enthalpy and entropy, association/dissociation rates and binding pathways using molecular modeling. Smaller host-guest systems containing only a few hundred atoms were studied to develop theoretical and practical benchmarks for the methods and protocols. These methods and protocols were applied in the larger protein-ligand systems that contain tens of thousands of atoms. Subsequent sections in this chapter describe the theories and techniques of molecular modeling. The chapters that follow present several studies of the binding/unbinding of bimolecular systems and a novel algorithm which specifically models the dissociation of bimolecular systems.

The thermodynamics of ligand-receptor binding has been intensively studied for decades. However, a fundamental understanding of kinetics is still absent, because kinetic studies require static snapshots of the initial and final states and a description of the entire transition process between the two states. To start with the relatively simple case, the study of facet selectivity of ligands on silver nanoplates is presented in Chapter 2. This work focused on the selective binding of several surfactants on the two silver facets. It was done by using a conformational search method which is aggressive in finding global conformation minimum in implicit solvent model with high efficiency.

Moving from small host-guest systems to protein-ligand systems, the increased scale in system size and number of atoms results in challenging technical issues, such as the exponential increase in computational cost. To overcome this challenge, a compensation between accuracy and computational cost has to be made. Chapter 3 demonstrates a molecular dynamic and docking study on inhibitor binding to TIM44 wildtype and its mutants.

Since the implicit solvent model is unable to result in a satisfactory description on solvent effects, especially on the estimation of the entropic contribution, an explicit solvent model needs to be utilized to obtain more accurate descriptions of solvent entropy via molecular dynamics simulations. In Chapter 4, solvent entropy evaluation methods are discussed and compared for the performance and limitations.

Among several entropy calculation methods available, the cell method stands out for its capacity to evaluate the entropy of single water molecule in addition to its ease of

implementation. Therefore this method is chosen for an enthalpy-entropy calculation of a host-guest system,  $\beta$ -cyclodextrin complexes, in explicit solvent model. The fast kinetics of  $\beta$ -cyclodextrin complexes allows for a direct evaluation of the association/dissociation rates via molecular dynamics. The investigation of this host-guest system provides useful insights into the thermodynamic-kinetic relationships for larger protein-ligand systems. Chapter 5 presents a study on the evaluation of thermodynamics and kinetics of ligands binding to  $\beta$ -cyclodextrin in explicit solvent.

Lastly, after obtaining accurate methods for evaluating the enthalpy and entropy of different size systems, another major challenge in molecular modeling is conformational sampling of systems in great size. Current sampling methods either suffer from absence of atomistic description of the system or the lack of temporal efficiency. To contribute to this issue, Chapter 6 introduces a novel conformational sampling method based on a new technique named internal Principle Component analysis or internal PCA.

## 1.2 Modeling with Molecular Mechanics

### 1.2.1 Force Fields

Molecular mechanics differs from quantum mechanics because the former models the structure and behavior of the molecular systems based on classical mechanics and empirical parameters. In molecular mechanics, the atoms are usually described by spheres with either hard potential or soft potential on the surface while they are connected by certain topology which governs the behavior of the entire molecule<sup>4,5</sup>. Starting from the all-atom molecular model, different levels of coarse graining can be achieved for

efficiency reasons <sup>6,7</sup>, but in this work, an all-atom non-polarizable description of the system is utilized.

To describe the molecular model in a mathematical framework, the system under study is assigned with a force field. A typical atomistic force field has a potential energy function defined as follows:

$$E = (E_{bond} + E_{angle} + E_{torsion}) + (E_{electrostatic} + E_{van\ der\ Waals}) \quad \text{eq. 1.1}$$

where the first parenthesis denotes the bonded terms and the second denotes the non-bonded terms. Although the interactions between two molecules involved in binding is usually non-bonded interaction, the bonded terms also play important roles because they determine the internal behaviors of the two interacting molecules.

Consider a force field widely used for protein simulations, the AMBER force field <sup>4</sup>. Its potential energy function obeys the equation below:

$$\begin{aligned} V(r) &= \sum_{\text{bonds}} k_b (r - r_0)^2 + \sum_{\text{angles}} k_a (a - a_0)^2 \\ &+ \sum_{\text{torsions}} \sum_p \frac{1}{2} V_n [1 + \cos(p\theta - \varphi)] \\ &+ \sum_{j=1}^{N-1} \sum_{i=j+1}^N f_{ij} \left\{ \epsilon_{ij} \left[ \left( \frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \end{aligned} \quad \text{eq. 1.2}$$

The chemical bonds are modeled as harmonic springs with a reference value equal to the equilibrium length of the bond, and a force constant describing the strength of the bond.

In the first term of the functional form,  $k_b$  is the force constant,  $r$  is the measured distance between the two atoms involved in the bond, and  $r_0$  is the equilibrium distance between the two atoms.

The angle term is introduced to further set up the internal structure of molecules. This term is similar to the bond term, where a harmonic potential is used to describe the angle between the two bonds sequentially connecting the three atoms. In the second term of the function, similarly,  $k_a$  is the force constant describing the rigidity of the angle,  $a$  is the measured position of the angle, and  $a_0$  is the equilibrium position of the angle.

The torsion term is the pivotal term which further defines the structure and differentiates the conformations of the molecule. In the third term of the function,  $V_n$  is the force constant which governs the strength of the torsion,  $p$  is the periodicity of the rotation of the torsion,  $\theta$  is the measured value of the torsion and  $\varphi$  is the phase angle determining the minima and maxima of the potential.

This torsion term consists of two major types of torsions: dihedral angle, and improper torsion angle. The dihedral angle directly defines the rotation of chemical bonds (Figure 1.1). For example, a methyl rotation in ethane will be defined as  $p = 3$  and  $\varphi = 180^\circ$ , where the three minima of the methyl rotation fall to  $0^\circ$ ,  $120^\circ$ ,  $-120^\circ$ .

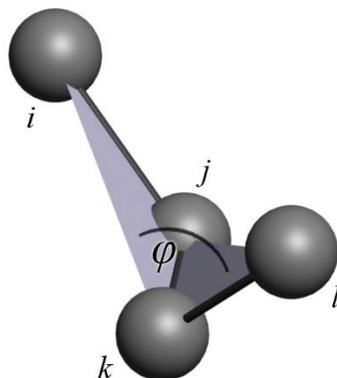


Figure 1.1 Dihedral torsion angle  $\varphi$  defined by a set of four sequentially bonded atoms i, j, k and l

The improper torsion angle has the same functional form as dihedral angle in AMBER force field, and it defines the angle at which the fourth atom deviates from the plane defined by the first three atoms (Figure 1.2). The improper torsion is the least frequent term in the potential energy function and is only used to enforce the planar geometry, for example, an  $sp^2$  hybridized atom.

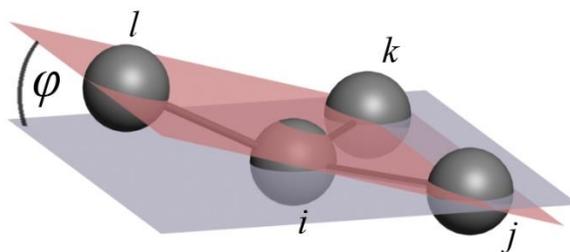


Figure 1.2 Improper torsion angle  $\varphi$  defined by a set of three atoms j, k and l bonded to the same atom i.

The non-bonded term are included to model electrostatic and van der Waals interactions. In molecular dynamics, the electrostatic interaction is modeled in a partial charge based monopole approach. The partial charge of the atom is usually assigned by empirical <sup>8,9</sup>, semi-empirical <sup>10</sup>, or quantum <sup>11</sup> calculations. Both electrostatic and van der Waals interaction are truncated at two body interactions and higher order of interactions are captured by empirically tuning the two body interaction parameters. The behavior of any pair of atoms in a system is under the government of non-bonded forces unless the two atoms are involved in a bond or angle term.

The last term in the potential energy functional describes the non-bonded interactions. The double summation loops over all pairs of atoms.  $f_{ij}$  is 1 is a scaling factor if the two atoms are the two ends of a torsion; a value of one means that the two atoms are not in the same bond, angle or dihedral, or is 0 if the two atoms are in the same bond or angle. In the outer parenthesis, the left term is the van der Waals force in the form of a 12-6 Lennard-Jones potential, where  $\epsilon_{ij}$  is the potential depth,  $r_{0ij}$  is the equilibrium position of the two atoms, and  $r_{ij}$  is the measured distance between the two atoms. The right term is the electrostatic force derived from the Coulomb's law, where  $q_i q_j$  is the product of the partial charges of the two atoms,  $\epsilon_0$  is the electric permittivity of the free space, and again,  $r_{ij}$  is the measured distance between the two atoms.

The functional form of the potential energy may vary in different force field. For example, in CHARMM force field <sup>5</sup>, the improper torsion angle is a harmonic function instead of a cosine function, and there are several other additional terms describing

certain specific geometric interactions. Hydrogen bond used to be included in the force field explicitly but now it is described implicitly. Overall, the trend of force field development is towards more accurate but mathematically simpler models. However some force fields focus more on accuracy while others focus more on efficiency.

### 1.2.2 Solvent Models

Many chemical reactions occur in the liquid phase and for the majority of biologically relevant processes, an aqueous environment is necessary. Therefore an appropriate description of the solvent is crucial for accurate evaluation of the liquid phase processes. In this work, the solvent is only limited to water because it is a common solvent used in chemistry, and more importantly, in biology. Existing solvent model ranges from computationally less expensive but also less accurate implicit solvent model, like generalized Born model (GB)<sup>12</sup> and Poisson Boltzmann (PB)<sup>13, 14</sup> solvent model, to more straightforward but more expensive explicit solvent model. Some examples of explicit solvent models which are commonly used are TIP3P<sup>15</sup>, TIP4P<sup>15, 16</sup> and TIP5P<sup>17</sup>. The implicit solvent models describes the solvent as a continuum, based on the electrostatic properties of the solute. They are usually computationally efficient and suitable for fast evaluation of the solvent effect. However they fail to capture the local orientation of the solvation shell. On the other hand, the explicit solvent models describes the solvent via presence of water molecules. Thus this type of model is capable of obtaining the anisotropic details of the solvent shell around the solute, but the computational costs are much higher than the implicit solvent model. A third type of solvent model is the RISM family of methods<sup>18-21</sup>. These methods enable a local

description of the orientation of the solvation shells with an implicit solvent model and thus serves as a bridging method between implicit solvent model and explicit solvent model. The following paragraphs discuss about the various solvent models in greater details.

An implicit solvent model uses the generalized Born (GB) equation <sup>12</sup> shown below to approximate the solvation free energy,

$$\Delta G_{pol} = -166 \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^n \sum_{j=1}^n \frac{q_i q_j}{f_{GB}} \quad \text{eq. 1.3}$$

where  $i$  and  $j$  denote the indices of a pair of atoms,  $\epsilon$  is the solvent dielectric constant,  $q$  is the partial charge on the atoms. The function of  $f_{GB}$  is typically in the following form,

$$f_{GB} = \sqrt{r_{ij}^2 + \alpha_i \alpha_j e^{\left(\frac{r_{ij}^2}{4\alpha_i \alpha_j}\right)}} \quad \text{eq. 1.4}$$

where  $r_{ij}$  is the distance between the atoms in a pair,  $\alpha_i$  and  $\alpha_j$  are the effective Born radii of the two atoms. For an isolated particle, the Born radius is equal to the van der Waals radius. However for an atom in a typical molecule, the effective Born radius is smaller than its van der Waals radius. The values of the Born radii effect the accuracy of GB calculation significantly. For example, an atomic anion has a greater radius than the neutral atom, so the GB free energy will deviate from experimental value significantly if the radius of the neutral atom is used.

The Poisson Boltzmann (PB) theory assumes that the dimensionless electrostatic potential  $\phi(x)$  can be solved from the non-linear partial differential equation (eq 1.5) because the charged solutes in the solution follow a Boltzmann distribution in the mean field approximation <sup>22</sup>,

$$-\nabla \cdot \varepsilon(x) \nabla \phi(x) - \sum_{i=1}^m c_i q_i e^{-q_i \phi(x) - V_i(x)} = \frac{4\pi e_c^2}{k_b T} \rho(x) \quad \text{eq. 1.5}$$

where  $\varepsilon(x)$  is the dielectric coefficient of the solvent,  $c_i$  is the number concentration of ion species  $i$ ,  $q_i$  is the charge of ion species  $i$ ,  $V_i(x)$  describes steric interactions between the solute and ions of species  $i$ ,  $e_c$  is the electron charge,  $k_b$  is the Boltzmann constant,  $T$  is the absolute temperature and  $\rho(x)$  is the fixed charge distribution of the solvent in units of electrons. The solution to the linear form of PB equation is usually used instead of the non-linear form in real applications of solvent free energy calculations, but the linear solution of PB equation is known to overestimate the ionic strength near the charged surface <sup>23, 24</sup>. The PB solvent model is used in static solvent free energy calculation due to its computational expense, rather than in the molecular dynamics simulations.

The reference interaction site model (RISM) theory is an integral equation theory of molecular liquids <sup>18</sup>. It has been successfully applied to many dynamics in aqueous solutions while it failed to describe the excluded volume effect correctly <sup>21</sup>. Its variant, 3D-RISM theory <sup>19, 20, 25</sup>, corrected the unphysical access to buried solute surface, and

thus witnessed wider applications in the past twenty years. It solves the 3D-RISM equation,

$$h_\gamma(r) = \sum_{\gamma'} c_{\gamma'}(r) * \left( w_{\gamma'\gamma}^{vv}(r) + \rho h_{\gamma'\gamma}^{vv}(r) \right) \quad \text{eq. 1.6}$$

where  $h_\gamma(r)$  is the solvent 3D total correlation function of site  $\gamma$ ,  $c_{\gamma'}(r)$  is the 3D direct correlation function,  $w_{\gamma'\gamma}^{vv}(r)$  is the intramolecular matrix,  $\rho$  is the number density of solvent species,  $h_{\gamma'\gamma}^{vv}(r)$  is the site-site total correlation function. The solution is coupled with the 3D-HNC closure including corrections ( $\Delta Q$ ) for the supercell periodicity artifact to the 3D correlation functions,

$$h_\gamma(r) = \exp(-\beta u_\gamma(r) + h_\gamma(r) - c_\gamma(r) - \Delta Q_\gamma) + \Delta Q_\gamma - 1 \quad \text{eq. 1.7}$$

where  $u_\gamma(r)$  is the interaction potential between solvent sites,  $\beta = (k_b T)^{-1}$  the Boltzmann factor, with the  $k_b$  the Boltzmann constant and  $T$  the temperature. The 3D-RISM method is an implicit solvent model which describes the local structure of the solvent shell rather than pure continuum, and therefore provides more accurate results than the continuum methods like PB or GB.

Despite the relatively expensive computational cost, an explicit solvent model is capable of accurately modeling the solvent effect especially for the first few solvation shells of the solutes. This is born from the facts that the orientation, local density change and hydrogen bond of water molecules are sometimes the determinant factor in dynamics. For example, a molecular dynamic study on the ligand binding to HIV-protease revealed that the strength of breaking of solvation shell hydrogen bonding contributes to the slow and

fast kinetics in different ligands <sup>26</sup>. Therefore, it is also desirable to model the solvent using explicit solvent models depending on the specific application. The most commonly used explicit water model is the TIP3P model, which depicts the water molecule as three monopoles centered on the three atoms and one Lennard-Jones sphere centered on the oxygen. This model is able to successfully reproduce water properties including hydrogen bond, density, volume, heat capacity and the like <sup>15</sup>, but it is also well known to fail in the correct description of the vibration of local density from the second water layer. This is due to the insufficient strength of the tetrahedral iceberg like orientation provided in the force field. TIP4P with the negative charge moved from the oxygen to a fourth site, and TIP5P with two additional sites of negative charges were introduced to enhance the tetrahedral geometry so that better radial distribution function can be reproduced from those later models. However to model the behavior or dynamics of the solutes, TIP3P is capable of capturing the effect from the first solvation shell and thus, more widely used in biological simulations. TIP4P and TIP5P, due to their higher computational cost compared to TIP3P, are often used in study of water properties.

### 1.3 Techniques in Molecular Modeling

#### 1.3.1 Minimization

The conformation of a molecular system consists of a set of coordinates positioning the atoms in the Cartesian space. Under the government of molecular mechanics, where the forces on each atom are modeled by an assigned force field, the atoms may feel extraordinarily strong force from their neighbors and undergo abnormal motions if the initial set of coordinates is not reasonable. Conformational optimization, or potential

energy minimization is a technique that removes unphysical contacts and any remaining thermal energy from the initial conformation. This is accomplished by iteratively optimizing the coordinates with respect to the forces in the system. The minimization algorithms may utilize the gradient or hessian to minimize the potential energy. Mathematically, at a minimized conformation, the gradient, which contains the forces on the atoms, is 0, and the hessian, which contains the changes of forces, has a positive determinant. Therefore a straightforward idea of minimization is to find a set of solution to the linear equations obtained by setting the gradient to zero. However, the gradient of the potential energy function of molecular systems does not result in an analytical or linear form. Thus more sophisticated mathematical methods are employed to solve for the minimized conformation iteratively.

The first class of minimization methods includes the second order Newton-Raphson algorithm<sup>27</sup> and the quasi-Newton method<sup>28</sup>. The Newton-Raphson method calculates the exact gradient and hessian at each cycle of iteration. Therefore this method converges to the local minimum with high efficiency. However the downside of this methods comes from the computational cost of evaluating the exact hessian at every step and the intrinsic problem of being only capable of finding local minimum starting from the vicinity of it. To solve the problem of computational cost, the quasi-Newton method was introduced in which the exact hessian is only calculated at the initial step while the hessian is approximated in sequential steps.

The second class of minimization methods includes the first order methods which only utilize the gradient of the potential energy function. These methods are gradient descent, conjugate gradient<sup>29</sup> and BFGS methods<sup>30-33</sup>. The gradient descent searches for a local minimum by moving the system towards the direction of gradient in every step but practically this often results in zigzag behavior of the search route and inefficient convergence. Similarly, the conjugate gradient also searches for a local minimum by using the gradient, but it updates the search direction based on the past history of search directions and thus avoids the zigzag behavior for higher efficiency. BFGS method, or Broyden-Fletcher-Goldfarb-Shanno algorithm, is a quasi-Newton method in which the hessian is never calculated exactly. It approximates the hessian at each iteration from the exactly calculated gradient. The BFGS method has a slightly better efficiency than the conjugate gradient method.

The first order minimization methods are not as efficient as the second order methods, but the advantage is that the starting point does not have to be near the minimum. This allows the combination of first order and second order methods to be used in the same application. With a random or coarse initial conformation of the molecular system, the first order methods are used to bring the conformation closer to the minimum and then the second order methods are used to fully minimize the system with high efficiency.

### 1.3.2 Docking

At the level of molecular dynamics simulation and MMPB/SA calculation, a high accuracy of the protein-ligand binding affinity can be achieved. However in the field of

drug discovery, efficiency is also an important factor due to the need to screen hundreds of thousands of ligands for a lead discovery. For this purpose, molecular docking was invented for fast estimation of the protein-ligand binding affinity. A docking method typically consists of a conformational search algorithm and a conformation scoring function.

The goal of the conformational search algorithm for docking is to generate reasonable conformations efficiently. There are three levels of conformational search algorithms in current docking methods: rigid body docking, flexible ligand docking and flexible protein docking<sup>34</sup>. In rigid body docking, the internal degrees of freedom of both ligand and protein are not considered, and thus the efficiency is favored over the accuracy. In the flexible ligand docking, every degree of freedom in the ligand is considered while the protein is still kept fixed in the conformational search. Flexible ligand docking can be further broken down to three categories of algorithms: systematic search, random or stochastic search and simulation search. Flexible ligand docking is successful in half of its applications, leaving the other half for more accurate flexible protein docking. The flexible protein docking allows the ligand and some portions of the protein to be flexible to account for the conformational change of the protein induced by the ligand binding. This is accomplished by MD or MC methods, rotamer libraries, protein ensemble grids, or soft-receptor modeling. Among the three types of conformational search algorithms, the most commonly used are flexible ligand docking and flexible protein sidechain docking. Multiple protein conformations are usually used in docking runs for better conformational search result.

With the conformations found by the search algorithm, the scoring function determines the ranking of them. This is not necessarily done by exact free energy calculation due to the efficiency concerns. The first category of scoring function is the force field based methods. These methods calculate the binding interaction energy based on assigned force field while normally neglecting the solvent contribution or estimating the solvent effect by implicit solvent model. The second category is the empirical scoring functions. These methods use several uncorrelated empirical terms to evaluate the ranking of the conformations. They are generally faster than the force field based scoring functions and include contributions that are beyond the description of molecular force fields. The third category of scoring functions is the knowledge-based scoring functions. These methods are designed based on the successful binding from databases following the rules and general principles statistically derived that aim to reproduce experimentally determined structures. Consensus scoring is the last category of scoring functions. They combine the advantages from various scoring functions for improvement of the probability of finding the correct binding pose.

In addition to the drug discovery applications, docking is also an important tool in molecular modeling. It is often used to generate the initial conformations of MD simulations in the situation when a crystal structure of ligand bound protein is unavailable or perturbation to the tightly bound structure is desirable. It has the advantage of fast binding conformational search and at the same time eliminating the bias from manual placement of the ligand to the protein-binding site. However it is not used for free energy evaluation due to the low accuracy of scoring functions, and this task is executed

by more sophisticated methods, like Mining Minima 2 (M2) method which will be discussed in later sections.

### 1.3.3 Molecular Dynamics

#### 1.3.3.1 Model Preparation

An all atom molecular dynamics simulation requires careful preparation of initial coordinates, protonation state, force field, solvent model and proper equilibrium because the potential energy function is very sensitive to the details of the model.

Generally there are two ways to build the initial coordinates of the model, either from the crystal structure database or built from scratch. The crystal structures can be directly used as the initial structure if it exists in a database. For protein structures, the Protein Data Bank <sup>35</sup> is a publicly available database which stores the experimentally determined apo protein and protein-ligand crystal structures. For generic molecules, the Cambridge Crystallographic Data Centre <sup>36</sup> is an academically available database. If the crystal structure is not available, careful work is necessary to build the structure. For proteins, homology modeling <sup>37</sup> is a tool which builds the undetermined protein structure based on a template of similar protein and its sequence. In the case of small molecules, various visualization and editing programs such as VEGA ZZ <sup>38</sup> can build the structure manually. If the protein and ligand structures are obtained by either means while the bound state structure has not been determined, docking <sup>34</sup> is the tool to search for the best reasonable bound conformations.

With the initial structure, the next step is to determine the protonation state of the molecules. For small molecules with known pKa, and pH of the desired simulation, one can directly edit the protonation state by adding or removing hydrogen atoms. For a protein molecule, the pKa of relevant residues must be calculated in a specific chemical environment for the correct protonation state with programs like MCCE<sup>39</sup>. The protonation state affects the dynamics and protein-ligand interactions such as hydrogen bonding and electrostatic interaction, and therefore must be assigned correctly for meaningful simulations.

The next step is to assign a proper force field to the molecules. A general force field applies to the small molecules, but carefully parameterized force fields must be assigned to the protein for proper secondary and tertiary structure behaviors. Specialized force fields also exist for lipids or other macromolecules.

The solvent model is usually built after the force field assignment. In the case of the implicit solvent model, no additional molecule is added to the system, but a flag in the simulation is turned on for specification of GB, PB, or even 3D-RISM solvent model. In the case of explicit solvent model, water molecules for the chosen solvent model, for example, TIP3P, are added to solvate the solutes. Collectively the shape of the added solvent is usually in a box shape with three dimensions measured for the periodic conditions in the MD simulation.

The last step before production run is to equilibrate the system. This includes a minimization stage, a water equilibrium stage, and a heating stage if applicable. The

minimization removes any possible clashes and brings the system to a local energy minimum. In the water equilibrium stage, the positions of the solutes are constrained while the water molecules are allowed to move freely to fully solvate the solutes and fill the empty space near the solutes. In the heating stage, the entire system is heated up from low temperature to the temperature of the production run, which is usually the room temperature. The purpose of gradual heating is to switch the velocity of the atoms gradually to avoid local overheating and the resulting unphysical movements. Once the system reaches equilibrium, production run can be performed.

### 1.3.3.2 Molecular Dynamics

Molecular dynamics (MD) is a standard and robust technique that is well established. It evolves the studied system under the government of Newton's equations of motion based on the assigned force field,

$$\frac{F_{x_i}}{m_i} = \frac{d^2x_i}{dt^2} \quad \text{eq. 1.8}$$

where  $F$  is the force on the particles (atoms),  $m$  is the mass,  $x$  is the position and  $t$  denotes the time. At each step, the force on each atom is calculated by an analytical method <sup>40</sup>.

Integration over time is accomplished by finite difference numerical method, such as Verlet <sup>41</sup>, velocity-Verlet <sup>42</sup>, and Leapfrog algorithms <sup>43,44</sup>. In this way, the coordinates of atoms in the system are updated correspondingly so that the motion of the system is produced. The temperature of the isolated system is maintained by enforcing an external heat bath environment with Langevin thermostat <sup>45,46</sup>, Andersen thermostat <sup>47</sup> or Nose-

Hoover thermostat<sup>48</sup>. In isothermal-isobaric ensemble (constant temperature constant pressure, or NPT) simulations, a constant pressure is maintained by similar algorithms. The time step in an all-atom simulations is usually limited to 1 or 2 femtoseconds due to the fast motions of hydrogen atoms. A SHAKE algorithm<sup>49</sup> is often employed to fix the motion of hydrogen atoms, allowing utilization of 2-femtosecond time step in protein simulations increasing efficiency. With such a small elemental time step, the concurrent limitation of timescale in molecular dynamics simulation is up to microseconds even with modern parallelization and acceleration techniques.

#### 1.3.3.3 Acceleration Techniques

The three major acceleration techniques in conventional MD are parallelization<sup>50,51</sup>, graphic processing unit (GPU) acceleration<sup>52</sup> and specialized machine (ANTON machine) acceleration<sup>53,54</sup>. The parallelization calculation was developed for MD simulation from the fact that the force felt by each atom and the motion of each atom are independent on one another. At each step, the calculation of forces and update of coordinates are distributed on the available CPU threads without significant mutual communications. The GPU accelerated MD appeared within a decade with the development of programmable graphic card. The idea behind GPU acceleration is similar to CPU parallelization, but for GPU acceleration the parallelization is realized by thousands of processors on the GPU card. The specialized ANTON machine approaches the acceleration by molecular dynamics oriented hardware design. With these acceleration techniques, the upper limit of current conventional MD reaches to

microsecond level, but the gap to second level relevant biology still hinders the direct application of conventional MD to many biomolecular problems.

To overcome this concern, variants of molecular dynamics were developed over the past twenty years. Among them the first category relies on the fact that the conformational changes happen slowly due to higher free energy barriers, and thus aims to lower the energy barrier or smoother the potential energy surface itself. The most obvious approach is to coarse grain the atomistic model into a residue bead model<sup>6, 55, 56</sup>, but this method loses the atomistic details and can only provide a reference about the overall motions of the system. Other early works used the locally enhanced sampling technique that focused on residues important to binding to accelerate the dissociation pathways<sup>57, 58</sup>.

Hyperdynamics was the first implementation of an algorithm to shoal the potential energy well of the dihedral torsions so that the conformational changes happens faster<sup>59</sup>.

Accelerated MD<sup>60</sup> was built on similar principals. Similarly its natural variant RaMD-db<sup>61</sup> aims at accelerating the most relevant portion of the system. Regarding the force field, a softcore force field was developed to allow easier motion at the region of interest<sup>62</sup>. A scaled potential was used to estimate the kinetics of ligand dissociation<sup>63</sup>. Recently, a Gaussian accelerated MD (GaMD) method was developed by using Gaussian functions to boost the conformational transition in protein systems, in a way similar to accelerated MD<sup>64</sup>, so that the Gaussian functions elevate the potential energy wells.

A second group of MD variants enhance sampling by employing additional forces that pull a ligand from its binding site. This group includes steered MD<sup>65, 66</sup> and target MD<sup>67</sup>

methods, which drive the conformational change to or away from some predefined states, respectively. A popular method in the past decade for enhanced sampling is metadynamics<sup>59, 68, 69</sup>. Instead of using Gaussian functions to systematically alleviate the potential energy surface of the system, it utilizes the Gaussian functions to raise the potential energy surface of the readily sampled regions so that the sampling evolves to unvisited, originally high energy portion of the potential energy surface, such as an energy barrier. However, collective variables must faithfully reproduce the reaction coordinates and distinguish the states of interests efficiently, which limits the applicability of metadynamics. Another method related to metadynamics is known as self-guided Langevin dynamics (SGLD)<sup>70, 71</sup>. It distinguishes from metadynamics in that it needs no pre-knowledge of the system. It accumulates the random force on the system for a directional accumulative force and enhances the low frequency motion of the system by applying that additional force back to the system.

Another class of methods build upon the replica exchange strategy. Replica exchange molecular dynamics (REMD)<sup>72, 73</sup> was developed on the idea that in high temperature or low interaction ensembles the system evolves with a faster speed than room temperature ensemble. Since its appearance, similar methods<sup>74-80</sup> were developed building upon the replica exchange principals, and allowed for many applications with this type of methods in, for example, helix formation<sup>81-84</sup>, ligand binding<sup>85-89</sup>, accurate free energy evaluation<sup>90-92</sup>.

### 1.3.4 Post Analysis

Creating data is straightforward, but post analyzing it is not trivial. Post analysis is carried out on the specific conformations or the continuous trajectory from either the processed results or the raw data.

#### 1.3.4.1 Structural Analysis

The root mean square deviation (RMSD) is a standard analysis used to calculate the deviation of the measured quantities, like the coordinates, the distance of a shift, or whatever is meaningful, from the average value or a designated reference, with the equation,

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x_0)^2} \quad \text{eq. 1.9}$$

where,  $x$  is the measured quantity,  $0$  denotes the reference and  $N$  is the number of counts. The alignment of molecules is performed based on the minimization of RMSD of coordinates.

Similar to RMSD, root mean square fluctuation (RMSF) is defined as the deviation of one atom or one group of atoms over a period of time from its average position, with the equation,

$$RMSF = \sqrt{\frac{1}{T} \sum_{t=1}^T (x_{i,t} - x_{i,0})^2} \quad \text{eq. 1.10}$$

where  $x$ , again, is the coordinate of the target atom or group of atoms,  $t$  denotes the time, and  $0$  denotes the reference average and  $T$  is the total time. This technique is especially useful in characterizing the flexibility of residues in a protein system in a trajectory <sup>93</sup>.

Hydrogen bond analysis is usually performed to characterize the stabilizing factors of ligand binding or the internal structure of a protein. Typically a hydrogen bond is defined as X-H...Y, where X-H is the hydrogen bond donor and the Y is the acceptor. A hydrogen bond is considered to be formed when the complementary angle of X-H...Y is less than  $20^\circ$  and the distance between H and Y is smaller than  $2.50 \text{ \AA}$  <sup>94</sup> although various cutoffs may be used in different applications. The energy of a hydrogen bond can be defined mathematically <sup>95</sup>.

A general structural analysis relies on the collective variables which refer to any measurable quantities from the trajectory, for instance, the distance between certain atoms, and the angle of rotation between any user-defined centers of masses of groups of atoms. A general definition of collective variables includes the RMSD of certain groups of atoms. The collective variables can serve as a projection of the intrinsic reaction coordinates depending on the efficiency of the choice of them and thus could be used as the dimensions of free energy surface. In certain MD acceleration methods, for instance, metadynamics, the collective variable is a set of fundamental parameters that must be defined prior to the application of the method.

#### 1.3.4.2 Energetic Analysis

MMPBSA (Molecular Mechanics Poisson Boltzmann/Surface Area)<sup>96</sup> is a technique for evaluating specific interactions and binding energy of a ligand-receptor system from an MD simulation. For energetic evaluation, MMPBSA removes the explicit solvent, and implements an implicit PB solvent model. This enables the program to separate the complex system into the receptor, the ligand and the original complex. A similar but less expensive version is MMGBSA (Molecular Mechanics Generalized Born/Surface Area) in which the PB solvent model is replaced by more efficient but less accurate generalized Born solvent model. In addition to reporting the different interaction energies for the ligand-receptor complex, it can also break the interaction energy down to residue-residue interactions.

#### 1.3.4.3 Principal Component Analysis

Principal component analysis (PCA)<sup>97</sup> is another common tool to extract useful information from the random motions of the systems in a long MD trajectory. It is similar to quasi-Harmonic approximation, in which the eigenvectors of the covariance matrix are calculated for intrinsic motion creation. Mathematically, it calculates the covariance matrix from the atom coordinates in the trajectory aligned to the structure under study, diagonalizes it for eigenvectors, and applies the eigenvectors on the average atom positions to create the motion of each principal component mode. Usually only the  $\alpha$ -carbon atoms in each residue of the protein are considered in the covariance matrix construction due to the minimal data required for convergence. Therefore the commonly

used PCA is termed as  $\alpha$ -carbon PCA. The principal component of each conformation along the PC mode can be obtained by,

$$q_i = R^T(x_i - x_a) \quad \text{eq. 1.11}$$

where the  $x$  denotes the conformation along the trajectory,  $i$  denotes the conformation number,  $a$  denotes the average conformation of the trajectory,  $q$  denotes the principal component on one PCA mode, and  $R^T$  is the transpose of the eigenvector matrix of the covariance matrix. It is able to produce the free energy surface on by projecting the trajectory onto the space of  $N$  principal component modes and calculating the free energy from population on the  $N$  dimension space<sup>98</sup>.

Another PCA algorithm constructs the covariance matrix using the  $\psi$  and  $\phi$  angles of protein backbone instead of the  $\alpha$ -carbon atoms and this technique is usually termed as dihedral PCA (dPCA)<sup>99</sup>. The advantage of dPCA over  $\alpha$ -carbon PCA is that it further eliminates any possible contribution from external degrees of freedom remaining in the alignment process and express the PCs in the dihedral subspace of internal coordinate space. Various works have illustrated the advantage of the dPCA technique<sup>99-103</sup>.

In Chapter 6 of this work, a variant of the dPCA, which is termed as internal PCA, developed by the author, is introduced. The internal PCA first constructs the internal coordinate by a novel mapping from Cartesian coordinate developed by the author. Then it constructs the covariance matrix from selected dihedral angles and bonds and angles in limited numbers. The purpose of internal PCA is to reproduce atomistic PCA motions. It

does not aim to illustrate any minima or create a free energy surface diagram. Details will be discussed in Chapter 6.

### 1.3.5 Free Energy and Entropy

From the classical view of statistical mechanics, the macroscopic quantities, which includes free energy, enthalpy and entropy can be expressed by the partition function of the microstates of the system in equilibrium. The classical partition function is defined by quantities like temperature and volume. In molecular modeling, such calculations of thermodynamic quantities reveals important information on the binding or relative binding of the ligand-receptor systems, such as binding affinity, selectivity and entropy-enthalpy balance.

One widely used ensemble in molecular dynamics is the canonical ensemble, where the number of particles is fixed, and the temperature and the volume are specified to certain values. The classical canonical ensemble partition function  $Z$  is given by,

$$Z = \int e^{-\beta E(x)} dx \quad \text{eq. 1.12}$$

where,  $x$  is the condition of the system that defines the microstate, or specifically, the momentum and position.  $\beta$  is the Boltzmann factor that equals to  $(k_b T)^{-1}$ , with the  $k_b$  the Boltzmann constant and  $T$  the temperature.  $E$  denotes the total energy of the system. The probability of the system in a microstate defined by  $x$  is given by,

$$P(x) = \frac{e^{-\beta E(x)}}{Z} = \frac{e^{-\beta E(x)}}{\int e^{-\beta E(x)} dx} \quad \text{eq. 1.13}$$

The probability function can be used to calculate the ensemble averages, or expectation values of an observable, such as the total energy,  $E$  of a system,

$$\langle E \rangle = \int E(x)P(x)dx = \frac{\int E(x)e^{-\beta E(x)}dx}{\int e^{-\beta E(x)}dx} \quad \text{eq. 1.14}$$

In the canonical ensemble, the free energy  $A$  (Helmholtz free energy) is given by,

$$A = -k_b T \ln Z \quad \text{eq. 1.15}$$

The entropy is defined according to the probability function,

$$S = -k_b \int P(x) \ln P(x) \quad \text{eq. 1.16}$$

or from the free energy and average energy,

$$S = \frac{-A + \langle E \rangle}{T} \quad \text{eq. 1.17}$$

Because of the difficulty in complete sampling of the phase space of complex systems, approximations are necessary in the calculation of the thermodynamic values. For example, the entropy a protein, which usually has more than a thousand atoms, can be estimated by calculating the entropy of individual dihedral angles from probability distributions and adding a correction from correlation map of the dihedral to account for the concerted motions <sup>104</sup>.

Another common approximation is to build the partition function of a molecular system via contributions from the minimum energy conformational states,

$$Z = \sum Z_i \quad \text{eq. 1.18}$$

By integrating the local partition function  $Z_i$  the total partition function can be obtained.

One method that utilizes this approximation is the Mining Minima generation 2 (M2) method<sup>105</sup>. It calculates the configurational integral, which is effectively the partition function, of the solute in implicit solvent resulting in the entropy of the solute. First it searches for minima including the global minimum aggressively by normal mode based algorithm and filters out the repeated conformations. Afterwards it calculates the partition function for each minimum by harmonic approximation and integrating the probability of conformations along each normal mode. Finally it calculates the total partition function of the system. Different from other similar methods, M2 does the calculation in internal coordinate instead of Cartesian coordinate, resulting in more physical normal modes.

For solvent entropy evaluation, other methods exist like distribution correlation based methods<sup>106</sup>, velocity correlation function<sup>107</sup> based method and harmonic approximation method. A detailed discussion on this topic is given in Chapter 4.

### 1.3.6 Kinetics

Compared with thermodynamics, kinetics is less well-studied with molecular modeling. Ligand-receptor binding/unbinding usually involves conformational changes that occur on the timescale of microseconds<sup>108</sup> or even longer. High energy barriers to binding and rough free energy surfaces naturally contribute to slow association/dissociation rates, which sometimes may be on the scale of seconds or even days. Both of these assert

challenges to contemporary computational power in sampling the dynamics of ligand-receptor binding, which can only simulate conformational changes up to microseconds<sup>54, 109-111</sup>. Furthermore, even if the conformational changes are sampled, the extraction of kinetics, or probability of state transitions, remains as a second obstacle before fully understanding the kinetics of the system.

Currently, the most important and successful method that tackles this topic is the Markov State model (MSM)<sup>112</sup>. This method does not require the sampling of a complete binding/unbinding event, but instead, it only requires a thorough sampling of the relevant states along the transition pathway, even from discrete trajectories. It clusters the reservoir of conformations by certain criteria and the transition rate between states can be assessed naturally from the trajectories. Finally these transition rates between states are stored into a matrix for diagonalization yielding the probabilities of the individual states, and the macroscopic transition rate, or the association/dissociation rates. This method is successful not only because it defines the calculation of transition rates, but also because it can be easily combined with distributed computing<sup>113</sup>, which is a powerful way to improve the computational power.

Another method that addresses the challenges of kinetics is the transition path sampling method<sup>114</sup>. In contrast to the MSM, this method provides a clear way to enhance the sampling of the transition path while providing a recipe to calculate the transition rate. Theoretically this method requires only an initial guess of the intermediate on the transition path, and thousands of sampling can be started stochastically from the

intermediate with randomized initial velocities. The transition path itself can be adjusted from the trial sampling of the intermediate. Some of the sampling will fall into the initial state of the transition, while others will fall into the final state of the transition. Through this method, the transition rate can be directly estimated from the probabilities of the two ending states. The downside of this method is that a large amount of sampling needs to be done. Also, the transition path can only consist of one major energy barrier. These limit the application of this method to the studies of transitions between one well-defined free energy barrier.

#### 1.4 References

1. Verlet, L. (1967) Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98-+
2. Bohacek, R. S., McMartin, C., and Guida, W. C. (1996) The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective. *Medicinal Research Reviews* **16**, 3-50
3. Stone, J. E., Phillips, J. C., Freddolino, P. L., Hardy, D. J., Trabuco, L. G., and Schulten, K. (2007) Accelerating Molecular Modeling Applications with Graphics Processors. *Journal of Computational Chemistry* **28**, 2618-2640
4. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **25**, 1157-1174
5. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **4**, 187-217
6. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007) The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *Journal of Physical Chemistry B* **111**, 7812-7824
7. Takada, S. (2012) Coarse-Grained Molecular Simulations of Large Biomolecules. *Current Opinion in Structural Biology* **22**, 130-137
8. Gasteiger, J., and Marsili, M. (1980) Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges. *Tetrahedron* **36**, 3219-3228

9. Gilson, M. K., Gilson, H. S. R., and Potter, M. J. (2003) Fast Assignment of Accurate Partial Atomic Charges: An Electronegativity Equalization Method That Accounts for Alternate Resonance Forms. *Journal of Chemical Information and Computer Sciences* **43**, 1982-1997
10. Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I. (2000) Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *Journal of Computational Chemistry* **21**, 132-146
11. Mulliken, R. S. (1955) Electronic Population Analysis on LCAO-MO Molecular Wave Functions .1. *Journal of Chemical Physics* **23**, 1833-1840
12. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical Treatment of Solvation for Molecular Mechanics And Dynamics. *Journal of the American Chemical Society* **112**, 6127-6129
13. Borukhov, I., Andelman, D., and Orland, H. (1997) Steric Effects in Electrolytes: A Modified Poisson-Boltzmann Equation. *Physical Review Letters* **79**, 435-438
14. Fogolari, F., Zuccato, P., Esposito, G., and Viglino, P. (1999) Biomolecular Electrostatics with the Linearized Poisson-Boltzmann Equation. *Biophysical Journal* **76**, 1-16
15. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935
16. Horn, H. W., Swope, W. C., Pitara, J. W., Madura, J. D., Dick, T. J., Hura, G. L., and Head-Gordon, T. (2004) Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *Journal of Chemical Physics* **120**, 9665-9678

17. Mahoney, M. W., and Jorgensen, W. L. (2000) A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly By Rigid, Nonpolarizable Potential Functions. *Journal of Chemical Physics* **112**, 8910-8922
18. Chandler, D., and Andersen, H. C. (1972) Optimized Cluster Expansions for Classical Fluids .2. Theory of Molecular Liquids. *Journal of Chemical Physics* **57**, 1930-+
19. Kovalenko, A., and Hirata, F. (1998) Three-Dimensional Density Profiles of Water in Contact with a Solute of Arbitrary Shape: A RISM Approach. *Chemical Physics Letters* **290**, 237-244
20. Kovalenko, A., and Hirata, F. (2000) Potentials of Mean Force of Simple Ions in Ambient Aqueous Solution. I. Three-Dimensional Reference Interaction Site Model Approach. *Journal of Chemical Physics* **112**, 10391-10402
21. Imai, T., Kovalenko, A., and Hirata, F. (2004) Solvation Thermodynamics of Protein Studied by The 3D-RISM Theory. *Chemical Physics Letters* **395**, 1-6
22. Baker, N. A. (2005) Improving Implicit Solvent Simulations: A Poisson-Centric View. *Current Opinion in Structural Biology* **15**, 137-143
23. Fenley, M. O., Harris, R. C., Jayaram, B., and Boschitsch, A. H. (2010) Revisiting the Association of Cationic Groove-Binding Drugs to DNA Using a Poisson-Boltzmann Approach. *Biophysical Journal* **99**, 879-886
24. Boschitsch, A. H., and Fenley, M. O. (2004) Hybrid Boundary Element and Finite Difference Method for Solving the Nonlinear Poisson-Boltzmann Equation. *Journal of Computational Chemistry* **25**, 935-955
25. Beglov, D., and Roux, B. (1997) An Integral Equation to Describe the Solvation of Polar Molecules in Liquid Water. *Journal of Physical Chemistry B* **101**, 7821-7826

26. Huang, Y.-m. M., Kang, M., and Chang, C.-e. A. (2014) Switches of Hydrogen Bonds During Ligand-Protein Association Processes Determine Binding Kinetics. *Journal of Molecular Recognition* **27**, 537-548
27. Hilderbrandt, R. L. (1977) Application of Newton-Raphson Optimization Techniques in Molecular Mechanics Calculations. *Computers & Chemistry* **1**, 179-186
28. Dennis, J. E., and More, J. J. (1977) Quasi-Newton Methods, Motivation and Theory. *Siam Review* **19**, 46-89
29. Hestenes, M. R., and Stiefel, E. (1952) Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards* **49**, 409-436
30. Broyden, C. G. (1970) The Convergence of a Class of Double-Rank Minimization Algorithms. II. The New Algorithm. *Journal of the Institute of Mathematics and Its Applications* **6**, 222-231
31. Fletcher, R. (1970) A New Approach to Variable Metric Algorithms. *Computer Journal* **13**, 317-&
32. Goldfarb, D. (1970) A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation* **24**, 23-&
33. Shanno, D. F. (1970) Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation* **24**, 647-&
34. Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2006) Protein-Ligand Docking: Current Status and Future Challenges. *Proteins-Structure Function and Bioinformatics* **65**, 15-26

35. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) Protein Data Bank - Computer-Based Archival File for Macromolecular Structures. *Journal of Molecular Biology* **112**, 535-542
36. Groom, C. R., Bruno, I. J., Lightfoot, M. P., and Ward, S. C. (2016) The Cambridge Structural Database. *Acta Crystallographica*
37. Cavasotto, C. N., and Phatak, S. S. (2009) Homology Modeling in Drug Discovery: Current Trends and Applications. *Drug Discovery Today* **14**, 676-683
38. Pedretti, A., Villa, L., and Vistoli, G. (2004) VEGA - An Open Platform to Develop Chemo-Bio-Informatics Applications, Using Plug-In Architecture and Script Programming. *Journal of Computer-Aided Molecular Design* **18**, 167-173
39. Georgescu, R. E., Alexov, E. G., and Gunner, M. R. (2002) Combining Conformational Flexibility and Continuum Electrostatics for Calculating pK(a)s in Proteins. *Biophysical Journal* **83**, 1731-1748
40. Blondel, A., and Karplus, M. (1996) New Formulation for Derivatives of Torsion Angles and Improper Torsion Angles in Molecular Mechanics: Elimination of Singularities. *Journal of Computational Chemistry* **17**, 1132-1141
41. Grubmueller, H., Heller, H., Windemuth, A., and Schulten, K. (1991) Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-Range Interactions. *Molecular Simulation* **6**, 121-142
42. Martys, N. S., and Mountain, R. D. (1999) Velocity Verlet Algorithm for Dissipative-Particle-Dynamics-Based Models of Suspensions. *Physical Review E* **59**, 3733-3736
43. Fincham, D. (1992) Leapfrog Rotational Algorithms. *Molecular Simulation* **8**, 165-178

44. Fincham, D. (1993) Leapfrog Rotational Algorithms for Linear-Molecules. *Molecular Simulation* **11**, 79-89
45. Doll, J. D., and Dion, D. R. (1976) Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - Numerical Techniques for Gaussian Generalized Langevin Dynamics. *Journal of Chemical Physics* **65**, 3762-3766
46. Adelman, S. A. (1979) Generalized Langevin Theory for Many-Body Problems in Chemical-Dynamics - General Formulation and the Equivalent Harmonic Chain Representation. *Journal of Chemical Physics* **71**, 4471-4486
47. Ryckaert, J. P., and Ciccotti, G. (1986) Andersen Canonical-Ensemble Molecular-Dynamics for Molecules with Constraints. *Molecular Physics* **58**, 1125-1136
48. Evans, D. J., and Holian, B. L. (1985) The Nose-Hoover Thermostat. *Journal of Chemical Physics* **83**, 4069-4074
49. Forester, T. R., and Smith, W. (1998) SHAKE, Rattle, And Roll: Efficient Constraint Algorithms for Linked Rigid Bodies. *Journal of Computational Chemistry* **19**, 102-111
50. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781-1802
51. Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J.,

- Wolf, R. M., Wu, X., York, D. M., and Kollman, P. A. (2015) AMBER 2015. University of California, San Francisco
52. Liu, W., Schmidt, B., Voss, G., and Mueller-Wittig, W. (2008) Accelerating Molecular Dynamics Simulations Using Graphics Processing Units with CUDA. *Computer Physics Communications* **179**, 634-641
53. Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., Young, C., Batson, B., Bowers, K. J., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., Ierardi, D. J., Kolossvary, I., Klepeis, J. L., Layman, T., McLeavey, C., Moraes, M. A., Mueller, R., Priest, E. C., Shan, Y., Spengler, J., Theobald, M., Towles, B., and Wang, S. C. (2008) Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Communications of the Acm* **51**, 91-97
54. Shaw, D. E., Grossman, J. P., Bank, J. A., Batson, B., Butts, J. A., Chao, J. C., Deneroff, M. M., Dror, R. O., Even, A., Fenton, C. H., Forte, A., Gagliardo, J., Gill, G., Greskamp, B., Ho, C. R., Ierardi, D. J., Iserovich, L., Kuskin, J. S., Larson, R. H., Layman, T., Li-Siang, L., Lerer, A. K., Li, C., Killebrew, D., Mackenzie, K. M., Mok, S. Y. H., Moraes, M. A., Mueller, R., Nociolo, L. J., Peticolas, J. L., Quan, T., Ramot, D., Salmon, J. K., Scarpazza, D. P., Schafer, U. B., Siddique, N., Snyder, C. W., Spengler, J., Tang, P. T. P., Theobald, M., Toma, H., Towles, B., Vitale, B., Wang, S. C., and Young, C. (2014) Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*, 41-53
55. Nagarajan, A., Andersen, J. P., and Woolf, T. B. (2012) Coarse-Grained Simulations of Transitions in the E2-to-E1 Conformations for Ca ATPase (SERCA) Show Entropy-Enthalpy Compensation. *Journal of Molecular Biology* **422**, 575-593

56. Tozzini, V. (2005) Coarse-Grained Models for Proteins. *Current Opinion in Structural Biology* **15**, 144-150
57. Elber, R., and Karplus, M. (1990) Enhanced Sampling in Molecular-Dynamics - Use of the Time-Dependent Hartree Approximation for a Simulation of Carbon-Monoxide Diffusion Through Myoglobin. *Journal of the American Chemical Society* **112**, 9161-9175
58. Martinez, L., Sonoda, M. T., Webb, P., Baxter, J. D., Skaf, M. S., and Polikarpov, I. (2005) Molecular Dynamics Simulations Reveal Multiple Pathways of Ligand Dissociation from Thyroid Hormone Receptors. *Biophysical Journal* **89**, 2011-2023
59. Voter, A. F. (1997) Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Physical Review Letters* **78**, 3908-3911
60. Hamelberg, D., Mongan, J., and McCammon, J. A. (2004) Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *Journal of Chemical Physics* **120**, 11919-11929
61. Doshi, U., and Hamelberg, D. (2014) Achieving Rigorous Accelerated Conformational Sampling in Explicit Solvent. *Journal of Physical Chemistry Letters* **5**, 1217-1224
62. Hornak, V., and Simmerling, C. (2004) Development of Softcore Potential Functions for Overcoming Steric Barriers in Molecular Dynamics Simulations. *Journal of Molecular Graphics & Modelling* **22**, 405-413
63. Mollica, L., Decherchi, S., Zia, S. R., Gaspari, R., Cavalli, A., and Rocchia, W. (2015) Kinetics of Protein-Ligand Unbinding via Smoothed Potential Molecular Dynamics Simulations. *Scientific Reports* **5**

64. Miao, Y., Feher, V. A., and McCammon, J. A. (2015) Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **11**, 3584-3595
65. Leech, J., Prins, J. F., and Hermans, J. (1996) SMD: Visual Steering of Molecular Dynamics for Protein Design. *Ieee Computational Science & Engineering* **3**, 38-45
66. Nicolini, P., Frezzato, D., Gellini, C., Bizzarri, M., and Chelli, R. (2013) Toward Quantitative Estimates of Binding Affinities for Protein-Ligand Systems Involving Large Inhibitor Compounds: A Steered Molecular Dynamics Simulation Route. *Journal of Computational Chemistry* **34**, 1561-1576
67. Schlitter, J., Engels, M., and Kruger, P. (1994) Targeted Molecular-Dynamics - A New Approach for Searching Pathways of Conformational Transitions. *Journal of Molecular Graphics* **12**, 84-89
68. Barducci, A., Bonomi, M., and Parrinello, M. (2011) Metadynamics. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **1**, 826-843
69. Cavalli, A., Spitaleri, A., Saladino, G., and Gervasio, F. L. (2015) Investigating Drug-Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Accounts of Chemical Research* **48**, 277-285
70. Wu, X. W., and Brooks, B. R. (2003) Self-Guided Langevin Dynamics Simulation Method. *Chemical Physics Letters* **381**, 512-518
71. Wu, X., Brooks, B. R., and Vanden-Eijnden, E. (2016) Self-Guided Langevin Dynamics via Generalized Langevin Equation. *Journal of Computational Chemistry* **37**, 595-601
72. Voter, A. F. (1998) Parallel Replica Method for Dynamics of Infrequent Events. *Physical Review B* **57**, 13985-13988

73. Sugita, Y., and Okamoto, Y. (1999) Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **314**, 141-151
74. Li, H., Li, G., Berg, B. A., and Yang, W. (2006) Finite Reservoir Replica Exchange to Enhance Canonical Sampling in Rugged Energy Surfaces. *Journal of Chemical Physics* **125**
75. Gil-Ley, A., and Bussi, G. (2015) Enhanced Conformational Sampling Using Replica Exchange with Collective-Variable Tempering (vol 11, pg 1077, 2015). *Journal of Chemical Theory and Computation* **11**, 5554-5554
76. Kamberaj, H. (2015) Conformational Sampling Enhancement of Replica Exchange Molecular Dynamics Simulations Using Swarm Particle Intelligence. *Journal of Chemical Physics* **143**
77. Cheng, X. L., Cui, G. L., Hornak, V., and Sinnnerling, C. (2005) Modified Replica Exchange Simulation Methods for Local Structure Refinement. *Journal of Physical Chemistry B* **109**, 8220-8230
78. Mitsutake, A., and Okamoto, Y. (2004) Replica-Exchange Extensions of Simulated Tempering Method. *Journal of Chemical Physics* **121**, 2491-2504
79. Mitsutake, A., and Okamoto, Y. (2000) Replica-Exchange Simulated Tempering Method for Simulations of Frustrated Systems. *Chemical Physics Letters* **332**, 131-138
80. Gil-Ley, A., and Bussi, G. (2015) Enhanced Conformational Sampling Using Replica Exchange with Collective-Variable Tempering. *Journal of Chemical Theory and Computation* **11**, 1077-1085
81. Garcia, A. E., and Sanbonmatsu, K. Y. (2002) Alpha-Helical Stabilization by Side Chain Shielding of Backbone Hydrogen Bonds. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 2782-2787

82. Ihalainen, J. A., Paoli, B., Muff, S., Backus, E. H. G., Bredenbeck, J., Woolley, G. A., Caflisch, A., and Hamm, P. (2008) Alpha-Helix Folding in the Presence of Structural Constraints. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9588-9593
83. Hegefeld, W. A., Chen, S.-E., DeLeon, K. Y., Kuczera, K., and Jas, G. S. (2010) Helix Formation in a Pentapeptide Experiment and Force-Field Dependent Dynamics. *Journal of Physical Chemistry A* **114**, 12391-12402
84. Stirnemann, G., and Sterpone, F. (2015) Recovering Protein Thermal Stability Using All-Atom Hamiltonian Replica-Exchange Simulations in Explicit Solvent. *Journal of Chemical Theory and Computation* **11**, 5573-5577
85. Anselmi, M., and Pisabarro, M. T. (2015) Exploring Multiple Binding Modes Using Confined Replica Exchange Molecular Dynamics. *Journal of Chemical Theory and Computation* **11**, 3906-3918
86. Wang, K., Chodera, J. D., Yang, Y., and Shirts, M. R. (2013) Identifying Ligand Binding Sites and Poses Using GPU-Accelerated Hamiltonian Replica Exchange Molecular Dynamics. *Journal of Computer-Aided Molecular Design* **27**, 989-1007
87. Kokubo, H., Tanaka, T., and Okamoto, Y. (2013) Two-Dimensional Replica-Exchange Method for Predicting Protein-Ligand Binding Structures. *Journal of Computational Chemistry* **34**, 2601-2614
88. Kokubo, H., Tanaka, T., and Okamoto, Y. (2013) Prediction of Protein-Ligand Binding Structures by Replica-Exchange Umbrella Sampling Simulations: Application to Kinase Systems. *Journal of Chemical Theory and Computation* **9**, 4660-4671

89. Rungnim, C., Rungrotmongkol, T., Hannongbua, S., and Okumura, H. (2013) Replica Exchange Molecular Dynamics Simulation of Chitosan for Drug Delivery System Based on Carbon Nanotube. *Journal of Molecular Graphics & Modelling* **39**, 183-192
90. Armacost, K. A., Goh, G. B., and Brooks, C. L., III. (2015) Biasing Potential Replica Exchange Multisite lambda-Dynamics for Efficient Free Energy Calculations. *Journal of Chemical Theory and Computation* **11**, 1267-1277
91. Galvelis, R., and Sugita, Y. (2015) Replica State Exchange Metadynamics for Improving the Convergence of Free Energy Estimates. *Journal of Computational Chemistry* **36**, 1446-1455
92. Khavrutskii, I. V., and Wallqvist, A. (2011) Improved Binding Free Energy Predictions from Single-Reference Thermodynamic Integration Augmented with Hamiltonian Replica Exchange. *Journal of Chemical Theory and Computation* **7**, 3001-3011
93. Hess, B. (2002) Convergence of Sampling in Protein Simulations. *Physical Review E* **65**
94. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics & Modelling* **14**, 33-38
95. Zhang, J., Liang, Y., and Zhang, Y. (2011) Atomic-Level Protein Structure Refinement Using Fragment-Guided Molecular Dynamics Conformation Sampling. *Structure* **19**, 1784-1795
96. Miller, B. R., III, McGee, T. D., Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **8**, 3314-3321

97. Bro, R., and Smilde, A. K. (2014) Principal component analysis. *Analytical Methods* **6**, 2812-2831
98. Zhou, R. H., Berne, B. J., and Germain, R. (2001) The Free Energy Landscape for Beta Hairpin Folding in Explicit Water. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 14931-14936
99. Altis, A., Nguyen, P. H., Hegger, R., and Stock, G. (2007) Dihedral Angle Principal Component Analysis of Molecular Dynamics Simulations. *Journal of Chemical Physics* **126**
100. Mu, Y. G., Nguyen, P. H., and Stock, G. (2005) Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins-Structure Function and Bioinformatics* **58**, 45-52
101. Maisuradze, G. G., and Leitner, D. M. (2007) Free Energy Landscape of a Biomolecule in Dihedral Principal Component Space: Sampling Convergence and Correspondence Between Structures And Minima. *Proteins-Structure Function and Bioinformatics* **67**, 569-578
102. Altis, A., Otten, M., Nguyen, P. H., Hegger, R., and Stock, G. (2008) Construction of the Free Energy Landscape of Biomolecules via Dihedral Angle Principal Component Analysis. *Journal of Chemical Physics* **128**
103. Riccardi, L., Nguyen, P. H., and Stock, G. (2009) Free-Energy Landscape of RNA Hairpins Constructed via Dihedral Angle Principal Component Analysis. *Journal of Physical Chemistry B* **113**, 16660-16668
104. Ai, R., Fatmi, M. Q., and Chang, C.-e. A. (2010) T-Analyst: A Program for Efficient Analysis of Protein Conformational Changes by Torsion Angles. *Journal of Computer-Aided Molecular Design* **24**, 819-827

105. Chang, C. E., and Gilson, M. K. (2004) Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *Journal of the American Chemical Society* **126**, 13156-13164
106. Lazaridis, T., and Karplus, M. (1996) Orientational Correlations and Entropy in Liquid Water. *Journal of Chemical Physics* **105**, 4294-4316
107. Lin, S. T., Blanco, M., and Goddard, W. A. (2003) The Two-Phase Model for Calculating Thermodynamic Properties of Liquids from Molecular Dynamics: Validation for the Phase Diagram of Lennard-Jones Fluids. *Journal of Chemical Physics* **119**, 11792-11805
108. Shukla, D., Meng, Y., Roux, B., and Pande, V. S. (2014) Activation Pathway of Src Kinase Reveals Intermediate States as Targets for Drug Design. *Nature Communications* **5**
109. Arkhipov, A., Shan, Y., Das, R., Endres, N. F., Eastwood, M. P., Wemmer, D. E., Kuriyan, J., and Shaw, D. E. (2013) Architecture and Membrane Interactions of the EGF Receptor. *Cell* **152**, 557-569
110. Decherchi, S., Berteotti, A., Bottegoni, G., Rocchia, W., and Cavalli, A. (2015) The Ligand Binding Mechanism to Purine Nucleoside Phosphorylase Elucidated via Molecular Dynamics and Machine Learning. *Nature Communications* **6**
111. Pierce, L. C. T., Salomon-Ferrer, R., de Oliveira, C. A. F., McCammon, J. A., and Walker, R. C. (2012) Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation* **8**, 2997-3002
112. Bowman, G. R., Pande, V. S., and Noe, F. (2014) An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. Springer Science+Business Media, USA

113. Shirts, M., and Pande, V. S. (2000) Computing - Screen Savers of The World Unite! *Science* **290**, 1903-1904
114. Dellago, C., Bolhuis, P. G., Csajka, F. S., and Chandler, D. (1998) Transition Path Sampling and the Calculation of Rate Constants. *Journal of Chemical Physics* **108**, 1964-1977

## **Chapter 2 The Facet Selectivity of Ligands on Silver Nanoplates: A Molecular Mechanics Study**

### 2.1 Introduction

Inorganic nanostructures often display shape-dependent physiochemical properties<sup>1</sup>, which promise them for many elegant and practical applications in fundamental sciences and industry<sup>2</sup>. Silver nanoplates, a classic example of shape-controlled nanostructures, show tunable surface plasmon resonance<sup>2,3</sup> and can find uses in biomedicine<sup>4</sup>, catalysis<sup>5</sup>, microelectronics and data storage<sup>6</sup>, single-molecule labeling-based biological assays<sup>7,8</sup>, LED materials<sup>9</sup> and lasers<sup>7</sup>. Because of the numerous applications, substantial efforts have been spent on synthesis of silver nanoplates with well controlled dimensions. Reported methods include ligand-assisted chemical reductions<sup>10-12</sup>, electrochemical synthesis<sup>13,14</sup>, photo-induced method<sup>3,15-20</sup>, sonochemical routes<sup>21</sup>, solvothermal method<sup>22,23</sup>, and templating method<sup>24-26</sup>. The ligand-assisted chemical reduction method has been the most popular one because of its high yield and relatively simple setup.

Citrate, often considered as a “magic” reagent in the solution phase synthesis of silver nanoplates, has been studied for years, but its role in directing plate growth remains unclear. Selective surface protection is one of the hypotheses for explaining the mechanism<sup>27</sup>. In this theorem, a core with both (111) and (100) facets is used, and citrate preferentially protects the (111) facet in a non-bonded manner; thus Ag atoms deposit on other facets such as (100) and (110) selectively<sup>27,28</sup>. During the process, (111) facet actually grows thicker to some extent, so the interaction between citrate and the Ag

surface is non-covalent rather than covalent or ionic<sup>28</sup>. Several other di-carboxylates or tri-carboxylates are eligible ligands and presumably play the same roles as citrate in protecting a facet selectively<sup>29</sup>. However, one needs to show that good ligands can tightly bind to the Ag nanoplate non-covalently and discriminate one facet for anisotropic plate growth.

Previous studies modeled interactions between ligand molecules and metal nanoplates with *ab initio* calculations or classical molecular mechanics (MM) methods. Density functional theory (DFT) was used to study interactions between ligands and Ag nanoplates. For example, the DFT electronic structure calculations were applied to study the binding of citric acid on Ag nanoplates<sup>30</sup>. The calculation showed preferential binding of citric acid to the (111) rather than (100) facet, with stronger binding to the (111) facet promoting crystal growth along the (100) facet. The work also provided a structure of citric acid on the Ag nanoplate surface. However, the DFT calculations need to assume that citric acid is in its neutral form, which differs from actual synthesis conditions where the ligand molecules are in the deprotonated form<sup>29</sup>. In addition, the DFT calculation focused on only one conformation of citric acid, not considering possible contributions from other molecular conformations. The DFT method was also used to study the role of polyvinylpyrrolidone (PVP) in the shape-selective synthesis of Ag nanoplates<sup>31</sup>. In contrast to citrate, PVP binds more strongly to the (100) than (111) facet<sup>32</sup>. The DFT calculations suggested that electrostatic or van der Waals (vdW) forces govern the ligand–Ag nanoplate binding. Although experimentally, the synthesis of Ag nanoplates was carried out in solvent, DFT studies treated the ligands and silver plate

systems in vacuum and neglected the influence of solvents. DFT has also been used to explain the surface-enhanced Raman scattering (SERS) results of adenine<sup>33,34</sup> and 5-amino tetrazole<sup>35</sup> absorptions on a Ag or gold (Au) surface. In these works, the interaction between adenine and the noble metal surface was treated as the interaction between adenine and the active site metal atoms, and the binding geometry was predicted. Similarly, DFT was also combined with Fourier transform-infrared (FTIR) to explore the two distinct binding modes of p-nitroaniline on aqueous and dry Au nanoparticles<sup>36</sup>. Besides being used for investigating the interaction with small organic molecules, DFT has been used to investigate the interaction of Ag clusters with chlorine atoms<sup>37</sup>, alkali ions<sup>38</sup> and nitric oxide<sup>39</sup>.

For sampling more conformations and model dynamic natures of molecules on metal nanoplates, molecular dynamics (MD) provides a powerful tool for investigating motions and interactions of molecule and nanoplate complexes. Force-field parameters used in MM methods for metal nanoplate systems are less well established<sup>40-44</sup>. Therefore, various tests have been used to examine sets of charges and vdW parameters for metal nanoplate surfaces, e.g. by simulating nanoparticle movement on the Au surface<sup>45</sup>. MD simulations may involve explicit solvent models. To reduce computational costs, water molecules may be modeled implicitly. For example, an implicit solvent model for modeling protein–metal surface interactions has been developed<sup>42,46</sup>, and a sequential multilevel computational protocol was developed to study the interaction between peptides and Au surfaces<sup>47</sup>. Studies have provided insights into peptide and protein conformations on Au surfaces<sup>41,48,49</sup>. MD trajectories also suggested possible

supermolecular configurations of tetraphenylporphyrins on a silver (111) facet<sup>50</sup>. An MD study of the mechanism of the growth of Ag nanoparticles suggested some conformational and energy reasons for the final stage of ligand-aided formation of nanoparticles<sup>51</sup> by sampling the binding geometry from randomly placed initial conformations. Surface diffusion and growth on Ag and Au clusters were studied by using many-body potentials, and the mobility of atoms on the facets was found to be the key factor of growth of nanoparticles<sup>52</sup>.

In the present paper, we used the mining minima (M2) method to compute the binding free energy of citrate and other ligands to Ag nanoplates. The method involves an MM force field and an implicit solvent model to compute free energy that efficiently searches local energy minima and computes the configuration integral of a molecular system as a sum of the contribution from its local energy minima<sup>53,54</sup>. The same as other classical MM methods, the M2 method cannot capture quantum effects such as electron transfers between a ligand and the metal surface. However, the methods can be used to study systems with thousands of atoms and provides insight into changes in energy and configuration entropy on binding<sup>55-58</sup>. From experimental results and computational studies of molecules on Au surfaces, force-field parameters for the Ag surface were established<sup>45</sup>. In this paper, we studied nine existing and two designed ligands on both (111) and (100) Ag facets. We computed relative binding free energies for the compounds and focused on ligand conformations and the interactions between ligands and Ag nanoplate surfaces. Charge states of the carboxylate group in the ligands may

have a crucial effect on the anisotropic synthesis of Ag nanoplates. We also examined the ionic states of citrate in the plate synthesis.

## 2.2 Methods

### 2.2.1 Description of Systems

We selected nine ligands with experimental data available for the synthesis of Ag nanoplates<sup>29</sup> and designed two ligands. The molecules were prepared via Vega ZZ<sup>59</sup>. Since experimental conditions indicated that the ligands are completely deprotonated, the molecules have a basic form (COO<sup>-</sup>). In normal experimental conditions yielding Ag nanoplates, the pH of the solution was typically 9.0 initially and increased to ~9.7 near the end of the experiments. No Ag nanoplate was synthesized if the pH of the solution was < 7. The Ag surface is described with three atomic layers for the accurate calculation of ligand–Ag interactions. In total, 192 Ag atoms were used to represent both the Ag (111) (8×8×3) and (100) (6×10×3+12) facets, for a fair comparison between molecular interactions with the two facets. From our calculations, we designed two potential ligands, cyclohexane-1,3,5-tri-carboxylate and 2,2',2''-(cyclohexane-1,3,5-triyl)triacetate, and the coordinates were prepared by use of Vega ZZ.

### 2.2.2 Force-Field Parameterization

The initial force-field parameters were assigned by using the CHARMM22 force field<sup>60</sup>. The equilibrium position of the bond angles of the ligands were modified according to the crystal structures in the Cambridge Crystallographic Data Center (CCDC) database

except for the two designed ligands, whose experimental structures are not available. The CCDC IDs are listed in Table 2-1.

Ligand	CCDC ID
Acetate	AMACET
Oxalate	AMOXAL
Malonate	BIFVAN
Succinate	BURWEQ
Glutarate	ROGQEK
Pimelate	KYRIL
Citrate	FATTID
Isocitrate	N/A
Trimesic acid	N/A

Table 2.1. Cambridge Crystallographic Data Centre (CCDC) IDs for small molecules. The CCDC IDs of the crystal structures used to modify the bond angles of the ligands.

Partial charges of the ligands were computed with use of the program Vcharge with the VC/2004 parameter set <sup>61</sup>. Because the parameters of the Ag clusters are not defined in the standard CHARMM force field, the bond length between the Ag atoms was set to 2.88 Å on the basis of the experimental lattice constant (4.09 Å) <sup>62, 63</sup>. The bond force constant was set to 300.0 kcal/mol and the angle and dihedral parameters to 0. The vdW interactions are described by the 12-6 LJ-Potential in equation,

$$E_{vDW} = \varepsilon \left[ \left( \frac{r_m}{r} \right)^{12} - 2 \left( \frac{r_m}{r} \right)^6 \right] \quad \text{eq. 2.1}$$

where  $\varepsilon$  was assigned to 0.094 kcal/mol reported by Alexei, et. al <sup>45</sup>, and  $r_{min}$  to 1.444 by halving the Ag-Ag bond length. Because all the ligands have negative charges, we considered a model accounting for charge polarization of the Ag nanoplate by setting a

0.5 unit charge for the Ag atoms<sup>64, 65</sup>. Details regarding the charge models are in result and discussion.

### 2.2.3 Computational Details

M2 method is based on thorough conformational search and modified harmonic approximation using the bond-angle-torsion coordinates. The standard chemical potential of species X can be written as Equation 1. Q stands for canonical partition function, and V is the volume. The subscribe N,X means the system containing N solvent molecules and one solute molecule X, and similarly N,0 means the system with N solvent molecules but no solute.  $C^\circ$  is the standard concentration. The explicit solvent molecules are replaced by an implicit solvent model, GBSA or PBSA, to avoid numerous configurations generated by the solvent molecules.

$$\mu_X^\circ = -RT \left( \frac{1}{V_{N,X} C^\circ} \frac{Q_{N,X}(V_{N,X})}{Q_{N,0}(V_{N,0})} \right) \quad \text{eq. 2.2}$$

Because there are multiple local minima in the canonical ensemble, the total standard chemical potential is described by Equation 2.3 where  $\mu_{X,i}^\circ$  is the chemical potential of each local minimum.

$$\mu_X^\circ = -RT \ln \sum_i e^{-\beta \mu_{X,i}^\circ} \quad \text{eq. 2.3}$$

By introducing Rigid Rotor Approximation, the standard chemical potential  $\mu_{X,i}^\circ$  of species X is given by Equation 2.4. In this equation,  $M_X$  is the total mass of X,  $I_{X,a}, I_{X,b}, I_{X,c}$  are its principle moments of rotational inertia.  $n_X$  is the number of atom in

the molecule.  $h$  and  $k$  are the Planck's and Boltzmann's constants. The first two brackets are translational and rotational contributions, and are straight forward in calculation. The most computationally intensive part is the third bracket, which is the vibrational part, and is further calculated by Equation 2.5. In Equation 2.5,  $U$  and  $W$  are the potential and solvation energies, and  $q$  is the vibrational coordinates. A more specific detailed explanation can be found in the original work <sup>54</sup>.

$$e^{-\beta\mu_{X,i}^{\circ}} = \left[ \frac{1}{C^{\circ}} \left( \frac{2\pi M_X kT}{h^2} \right)^{\frac{3}{2}} \right] \times \left[ 8\pi^2 \left( \frac{2\pi kT}{h^2} \right)^{3/2} (I_{X,a} I_{X,b} I_{X,c})^{1/2} \right] \times \left[ \left( \frac{2\pi kT}{h^2} \right)^{(3n_X-6)/2} Z_X^{vib} \right] \quad \text{eq. 2.4}$$

$$Z_X^{vib} = \int e^{-\beta(U(q)+W(q))} dq \quad \text{eq. 2.5}$$

The M2 method is different from MMPBSA or other MD based analysis method. It uses analytical integration from the diagonalized Hessian matrix (modified harmonic approximation). According to Equation 2.3, the total standard chemical potential of species X is the sum of those of local minima weighed by the Boltzmann factor. Therefore, there is no fluctuation or standard deviation in the results from M2 method.

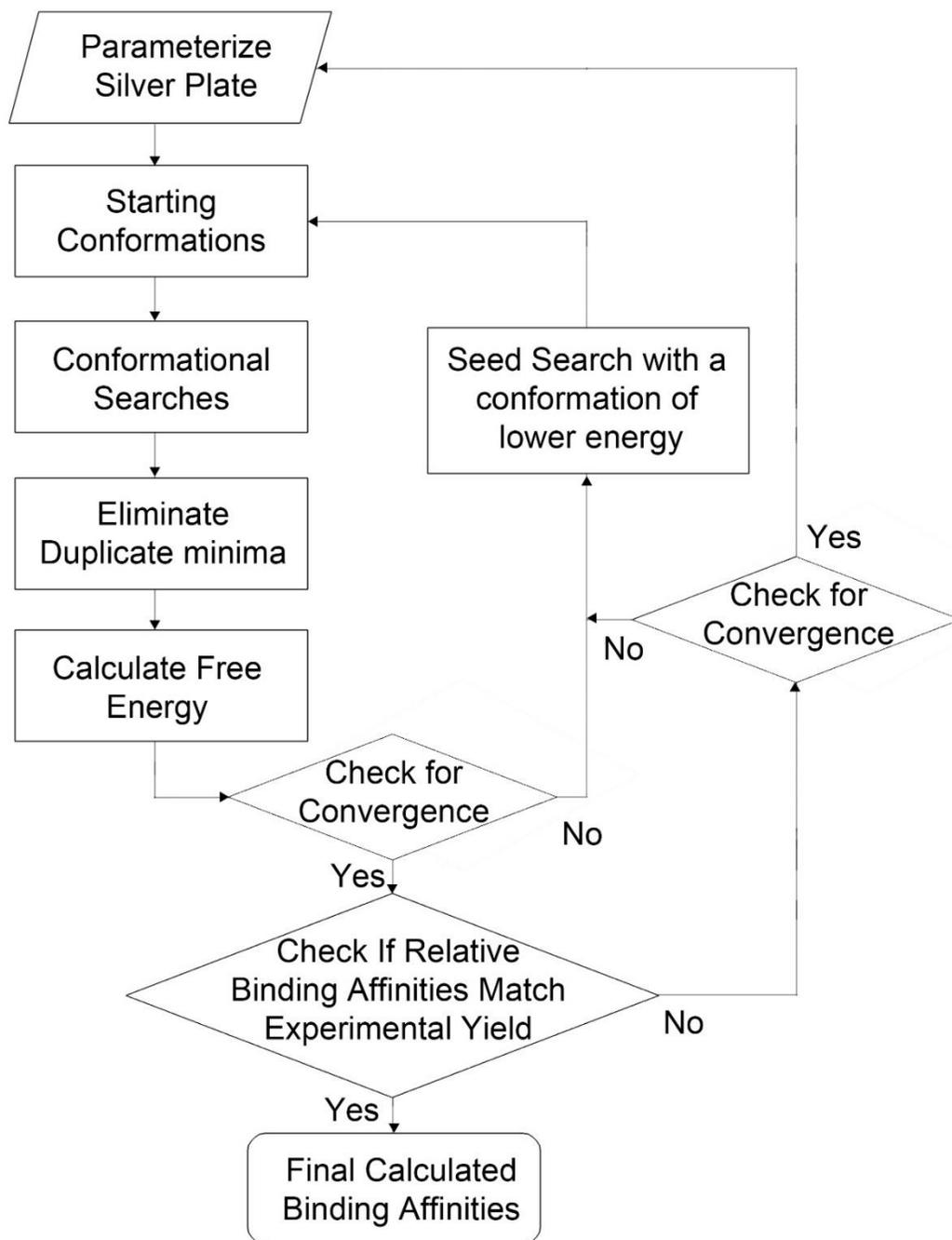


Figure 2.1. Flow chart of applying M2 method to silver plate-ligand systems.

Binding affinity of each ligand was calculated by using the second-generation mining minima method (M2 method)<sup>53-55</sup>. A flow chart of M2 method is given in Figure 2.1. The method starts by using the Tork search algorithm for thorough conformational searches for each ligand. Various conformations of a ligand were used to construct initial conformations of the complex to perform conformational searches for ligand–nanoplate complexes, where the Ag nanoplate was held fixed. Because the program may find a conformation multiple times, repeat conformations were filtered out. The details of filtering are in Section 2.4. The method then involved computing the configuration integral  $z_i$  for each unique conformation  $i$  found by the Tork search, and the free energy of this local energy minima was  $G_i = -RT \ln z_i$ , where  $R$  is the gas constant and  $T$  was set to 300 K. The standard free energy of each molecular species  $G_x$ , where  $x$  may be a ligand, a complex, or the nanoplate, is then the Boltzmann weighted sum of each computed  $G_i$ . Several iterations were performed to ensure that no new conformations with lower free energy were found, and the computed free energy,  $G_x$ , was converged. The binding affinity of each ligand was computed by  $\Delta G = G_{\text{complex}} - G_{\text{ligand}} - G_{\text{nanoplate}}$ . The free energy obtained in M2 was computed for systems in the standard condition, i.e. 1 atm pressure and 1 mol/L concentration. In experiments, though the overall concentrations of reagents and ligands are the same in all systems, the local concentrations of different ligands around the Ag surface may not be exactly identical. However, because experiments cannot detect minor changes of the local concentration, it is reasonable to approximate that the local concentrations of ligands near the Ag surface are similar, resulting in a constant in the M2 calculations. The values are canceled when

computing the related binding free energy. During the calculations, the standard pressure and concentration are separated as external degrees of freedom based on the Rigid Rotor Approximation<sup>54,66</sup>, and the analytical integration results in  $8\pi^2/C^\circ$ <sup>54</sup>, where  $C^\circ$  is the standard concentration that can be presented as  $1660 \text{ \AA}^3$  under standard pressure 1atm.

We used a distance dependent dielectric constant (DDD) model, which is widely applied in docking applications<sup>67-69</sup> to avoid unrealistic *in vacuo* Columbic interactions. This is because Poisson Boltzmann (PB) or generalized Born (GB) implicit solvent models are not yet fully developed for Ag plates. Because the distance dependent dielectric constant model is unable to provide an accurate description of desolvation energy<sup>70,71</sup>, we performed MD simulations of citrate and succinate using NAMD 2.9<sup>72</sup> to estimate the full picture of desolvation energies of di- and tri-carboxylates from the average interaction potential energies between succinate or citrate and the water molecules<sup>73-75</sup>. In this simulation, the ligands were parameterized by the CHARMM22 force field extracted from the M2 method. The solute was first solvated by a standard TIP3P water box, which is about  $20 \text{ \AA}$  from the solute itself, and then neutralized by  $\text{Na}^+$  ions. For both succinate and citrate, the water box contains 2721 water molecules. The entire system was equilibrated at 50, 100, 150, 200, 250 and 298 K for 10 ps at each step in the constant temperature, constant pressure canonical (NPT) ensemble. A product run was performed at 298 K in NPT for 5 ns with time step 1 fs. In these simulations, temperature and pressure were controlled by using Langevin dynamics. We used VMD<sup>76</sup> to calculate the potential energy of the entire system, the solute, and the water box of the simulation and then used the formula,

$$E_{interaction} = E_{total} - E_{solute} - E_{water\ box} \quad \text{eq. 2.6}$$

to calculate the average interaction energy.

#### 2.2.4 Conformation Filtering

Because the search algorithm finds the same conformation multiple times, to avoid double counting a conformation for the free energy calculation, we used a symmetry detection routine<sup>77</sup> to filter the repeats for the free ligands. Conformations are considered distinct if any dihedral angles differ larger than 15°. For the complex, in addition to applying the detection routines used for the free ligands, more criteria were used to identify repeats for the complexes. The criteria included distance between the center of mass of the ligand and the Ag nanoplate, the position of the projection point of the center of mass of the ligand on the plate surface in a periodic cell, and the orientation of the ligand relative to the nanoplate. Because the nanoplate has a periodic nature, the periodicity was considered in the filtering processes. The cutoff parameters were the same in the previous routine.

#### 2.2.5 Explicit Solvent Validation of the Distance Dependent Dielectric Constant

MD simulation was performed to optimize the arrangements of water molecules on the silver plate-ligand complex binding region, in order to investigate how many water molecules are replaced by the ligand upon binding. Releasing bound waters from the Ag surface may gain solvent entropy but lose solvent enthalpy upon ligand binding. However, the implicit solvent model may not be able to capture the effects for different ligand binding.

The global energy minima of each silver plate-ligand complex are used as an initial conformation for MD simulations using NAMD 2.9<sup>72</sup>. A complex conformation is solvated by a TIP3P water box, which is about 21 Å from the complex. In order to focus on water replacement, no counterions were placed on the Ag surface. The plate and ligand were held fixed during the simulations. The system is gradually heated at 50K, 100K, 150K, 200K, 250K and 298K in NPT for 10 ps. Finally a product run in NPT for 20 ps is performed and a trajectory is saved every 20 fs. This 1000 frame trajectory was processed to count how many water molecules on average were present at the first layer (defined by a distance 3.2 Å from the plate) on the silver plate within a region with radius 10 Å from the projection of the center of geometry of the ligand on the plate.

#### 2.2.6 Synthesis Method

The synthesis of silver nanoplates was conducted based on a previous reported method<sup>29</sup>. In a standard synthetic approach, the total volume of the reaction solution is fixed at 25.00 mL. Typically, a 24.75 mL aqueous solution combining silver nitrate (0.05 M, 50 µL), carboxylates (75 mM, 0.5 mL), and H<sub>2</sub>O<sub>2</sub> (30 wt %, 60 µL) was vigorously stirred at room temperature in air. Sodium borohydride (NaBH<sub>4</sub>, 100 mM, 250 µL) was rapidly injected into this mixture to initiate the reduction, immediately leading to a light yellow solution. After ~ 3 min, the colloidal solution turned to a deep yellow due to the formation of small silver nanoparticles. Within the next several seconds, the morphology started to change from particles to nanoplates accompanied by the solution color changing from deep yellow to red, green, and blue. The entire transition from nanoparticle to nanoplates typically took 2-3 minutes.

## 2.3 Results and Discussion

### 2.3.1 Overview of Ligands on the Ag Surface

The conformations of global energy minima for each ligand and the Ag surface complex are in Figure 2.2. The electrostatic attraction dominates the binding free energy and is the major driving force. The negatively charged oxygen atoms of the carboxylate ions all point to the (111) or (100) facet. Notably, the carboxylate group was modeled by the deprotonated form, because experiments showed that the acid form is necessary for the ligands to non-covalently bind to the surfaces. With pH changed from 9 to  $< 7$ , the carboxylate ions ( $\text{COO}^-$ ) were protonated ( $\text{COOH}$ ) and the plate formation vanished. Although the classical force field cannot model the potential electron transfers between the Ag and oxygen atoms, the strong ionic interaction may capture some characteristics of partial covalent bonds for the adsorption of a ligand on the Ag surface. The distance between an oxygen atom of the carboxylate group and its nearest Ag is  $< 3 \text{ \AA}$ , whereas the oxygen atoms of the hydroxyl group typically position toward the solvent (Figure 2.2). We modeled both (111) and (100) facets as 192-atom three-layer ( $8 \times 8 \times 3$ ) and ( $6 \times 10 \times 3 + 12$ ) Ag clusters. Because the facets are symmetric, exactly the same ligand orientation/position and conformation on the first and third layers are repeats, and only the complex conformations of ligands on the first layer were retained for further analysis and free energy calculation.

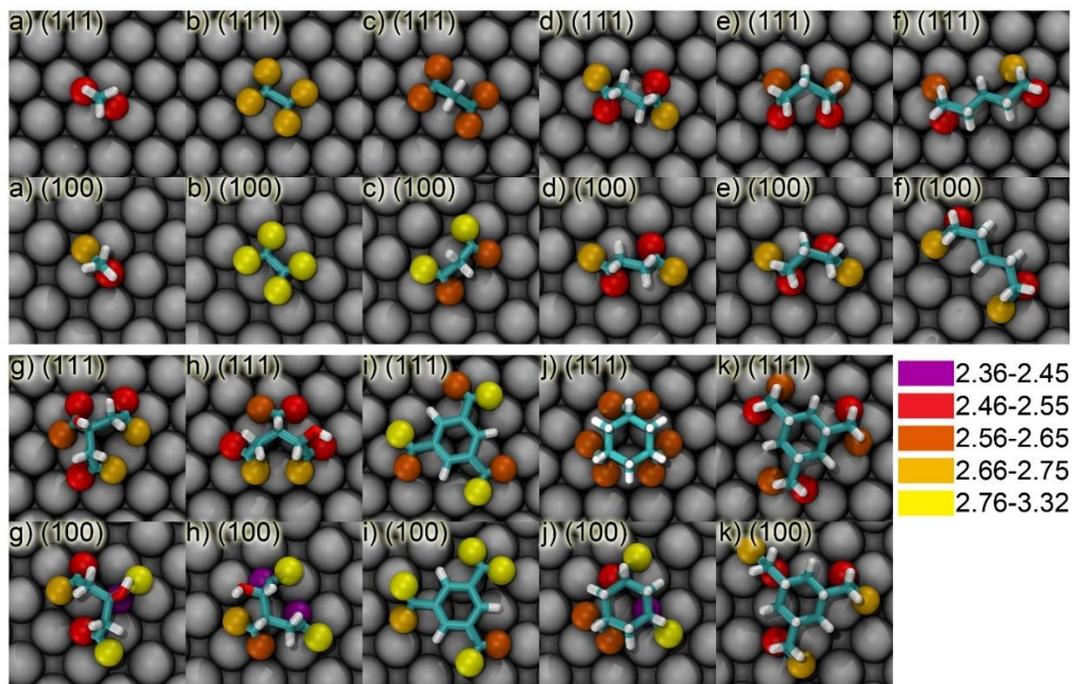


Figure 2.2. Top view of the bound geometry of the ligands on two silver (Ag) facets (top for [111] facet, bottom for [100] facet). a) acetate, b) oxalate, c) malonate, d) succinate, e) glutarate, f) pimelate, g) citrate, h) isocitrate, i) trimesic acid, j) designed ligand #1, k) designed ligand #2. The oxygen atoms are color mapped to represent the average distance between the oxygen atom and the closest three Ag atoms according to the color bar (in Å).

Table 2.2. Binding free-energy calculations and experimental data for nine ligands. Data were grouped by number of carboxylate groups. Experimental yields<sup>29</sup> were obtained from the experimentalists.  $\Delta G_{\text{comp}}$ : binding affinity of ligands on (111) facet;  $\Delta\Delta G_{\text{comp}}$ : relative binding affinity of ligands on (111) facet within each group. (For  $\Delta\Delta G$ , the reference for Group 2 is succinate and for Group 3 citrate.)

COO <sup>-</sup> group number	Ligand	Structure	Experiment al yield (%) <sup>[2]</sup>	$\Delta G_{\text{com}}$ P	$\Delta\Delta G_{\text{com}}$ P
1	Acetate		0	-129.0	
	Oxalate		0	-239.6	10.0
	Malonate		80	-251.6	-2.0
2	Succinate		100	-249.6	0.0
	Glutarate		50	-248.0	1.6
	Pimelate		0	-244.0	5.6
	Citrate		100	-361.2	0.2
3	Isocitrate		90	-361.0	0.2
	Trimesic acid		0	-345.5	15.7
New ligands	#1			-353.6	7.6
	#2			-361.3	-0.1

# of COO <sup>-</sup> group	Ligand	Facet	$\Delta G$	$\Delta E_{\text{Tot}}$	$\Delta E_{\text{Bonded}}$	$\Delta E_{\text{vdW}}$	$\Delta E_{\text{Elec}}$	$-T\Delta S$	$\Delta G_{111} - \Delta G_{100}$	
1	Acetate	111	-129.0	-130.9	-0.5	3.3	-133.7	2.0	-2.8	
		100	-126.2	-129.3	-0.6	3.3	-132.0	3.1		
2	Oxalate	111	-239.6	-246.5	-0.7	5.9	-252.4	6.9	-11.1	
		100	-228.4	-235.0	-0.9	5.4	-240.2	6.6		
	Malonate	111	-251.6	-256.2	-0.8	6.0	-261.7	4.6	-10.1	
		100	-241.5	-247.9	-0.8	5.9	-253.4	6.4		
	Succinate	111	-249.6	-257.6	-0.7	6.5	-263.5	8.0	-6.9	
		100	-242.6	-251.5	-0.4	7.2	-258.5	8.9		
	Glutarate	111	-248.0	-255.8	-0.5	5.6	-260.9	7.9	-5.9	
		100	-242.1	-251.5	-0.2	6.1	-257.4	9.4		
	Pimelate	111	-244.0	-253.1	-0.5	5.1	-257.7	9.2	-5.8	
		100	-238.1	-249.7	-0.8	5.3	-254.3	11.6		
	3	Citrate	111	-361.2	-374.9	3.0	10.3	-388.2	13.7	-10.2
			100	-351.1	-363.3	2.7	12.4	-378.5	12.3	
Isocitrate		111	-361.1	-372.7	4.0	10.9	-387.7	11.7	-7.9	
		100	-353.1	-363.4	3.5	14.3	-381.2	10.2		
Trimesic acid		111	-345.5	-357.8	4.3	6.1	-369.2	12.3	-14.9	
		100	-330.6	-340.5	3.9	4.5	-350.0	9.9		
New	1	111	-353.6	-365.4	9.0	7.8	-382.3	11.8	-5.1	
		100	-348.5	-361.1	8.9	8.7	-378.7	12.5		
	2	111	-361.3	-375.0	-1.0	7.5	-381.4	13.7	-8.0	
		100	-353.3	-367.8	1.3	8.7	-377.8	14.5		

Table 2.3. Binding energy components of ligands. The unit of this table is kcal/mol. The last column,  $\Delta G_{111} - \Delta G_{100}$ , indicates the ability of the ligands to distinguish the two Ag facets, where  $\Delta G_{111}$  and  $\Delta G_{100}$  represent the binding free energy of the same ligand to the (111) and (100) facet, respectively.

In this system, a good ligand must show reasonably strong binding affinities to the Ag surface and prefer the (111) facet. The stronger binding free energy suggests better protection on the (111) facet, thus resulting in higher yield experimentally. The absolute values of Columbic attractions are considerably different because of different numbers of the carboxylate ions. As a result, this study classified the ligands as three groups based on the number of carboxylate acids for better comparison. As shown in Tables 2.2 and 2.3, the calculated trend of binding free energies agrees with that in experimental data for each group of ligands and shows that good ligands, such as succinate and citrate, bind more tightly to the (111) than (100) facet. The following subsections detail the geometries in the bound states and changes of conformations, energy components, and configuration entropy.

### 2.3.2 Binding Preference of Ligand to the (111) and (100) Facets

All ligands bind tighter to the (111) than (100) facet (Table 2.3). Ligands that lead to high yields of anisotropic silver plate, i.e. succinate, citrate and isocitrate, the preference on (111) facet is stronger than ligands that show low yields, and the differences,  $\Delta G_{111} - \Delta G_{100}$ , are generally larger than 10 RT. The calculations support that ligands have better coverage and protection on the (111) facet from exposure to the solvent than the (100) facet, thus promoting the plate to grow on the (100) facet. Instead of covalently binding to the surface, the ligands bind to the Ag plate non-covalently and are in chemical equilibrium ( $K_{eq} = C k_{on}/k_{off}$ , where  $K_{eq}$  is the equilibrium constant, C is the concentration and  $k_{on}$  and  $k_{off}$  are the association and dissociation rate constants, respectively). As a result, the plate still can grow slowly on the (111) facet during a long experimental

process because a ligand does not permanently protect the (111) facet. Electrostatic attractions are the major contribution to stronger (111) facet–ligand binding because of the greater Ag density on the (111) facet. The Ag cluster is cubic closest-packed, and the atom density on (111) facet is  $\sim 1.155 \left(\frac{2}{3} \times \sqrt{3}\right)$  times that on the (100) facet (Figure 2.3). The closely packed (111) facet results in more non-bonded pairs between the Ag atoms and a ligand in a (111) facet–ligand complex, which yield stronger electrostatic attractions to stabilize the complex. Except for acetate, all ligands bind considerably stronger to the (111) than (100) facet, with binding free energies more negative than -6 kcal/mol and the difference comes mainly from the electrostatic interactions. Less distinguishable binding between acetate and both facets may be one reason why acetate cannot serve as a good ligand (0% plate yield).

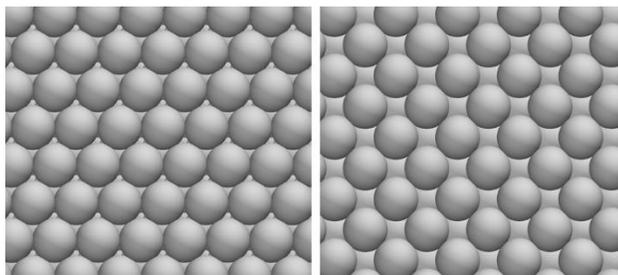


Figure 2.3. Top-down view of the (111) (left) and (100) facet (right). The atom density is greater for the (111) than (100) facet.

The higher atom density in the (111) facet also provides a geometry for all the ligands to fit better on the surface. The mean oxygen–Ag distances are shorter on the (111) than (100) facet (Table 2.4 and Figure 2.2), which contribute to stronger ionic attractions.

Ligand	(111) facet			(100) facet		
	Global Minima	Boltzmann Mean	SD	Global Minima	Boltzmann Mean	SD
Acetate	2.52	2.54	0.01494	2.6	2.60	0.00688
Oxalate	2.67	2.67	0.00073	2.85	2.85	0.00075
Malonate	2.59	2.60	0.00247	2.69	2.69	0.01128
Succinate	2.59	2.58	0.00875	2.63	2.64	0.00989
Glutarate	2.56	2.55	0.00537	2.61	2.60	0.00889
Pimelate	2.57	2.55	0.01359	2.61	2.61	0.00696
Citrate	2.57	2.58	0.00130	2.68	2.69	0.00823
Isocitrate	2.57	2.57	0.00229	2.69	2.70	0.02963
Trimesic Acid	2.69	2.69	0.00167	2.87	2.89	0.02589

Table 2.4. Computed distances between oxygen and Ag atoms. The average oxygen–Ag distance was calculated by averaging the distances between the oxygen atoms of the ligands and the three Ag atoms that are closest to each of the oxygen atoms. The unit is Å. The distances were consistently shorter on the (111) than (100) facet. On the (111) facet, the distances decrease from oxalate to pimelate. This finding can be explained by more flexible molecules fitting the periodic surface better than less flexible ones. As well, because of this factor, oxalate and malonate are unfavorable to the (111) facet.

Because the ligands have negative formal charges, and most of the negative charges are distributed on the oxygen atoms of the carboxylate group, ionic attractions can be optimized if the oxygen atoms are placed as close as possible to the positively charged Ag atoms. The ligands in this study are flexible molecules, and our study illustrated that the molecules could adjust their configurations to maximize the intermolecular attractions. The denser and shorter periodicity length of (111) facet provides a better geometry for the ligands to have closer contacts to the surface than the (100) facet.

### 2.3.3 Relative Binding Affinities of Various Ligands

Table 2.3 illustrates the computed relative binding free energies ( $\Delta\Delta G_{\text{comp}}$ ) of three groups of ligands: mono-carboxylate, di-carboxylate and tri-carboxylate. Here we focus

on ligand binding affinities to the (111) facet because this determines the coverage offered by a ligand and the corresponding experimental yield. Comparing  $\Delta\Delta G_{\text{comp}}$  for ligands within each group provides a correct ranking of the affinities, with a stronger binder able to form a more stable (111) facet–ligand complex to protect the (111) facet against growing in this direction (high % plate yield). The absolute values of the computed Columbic energies are on a different scale because of different formal charges of the ions in different groups. In typical energy calculations, the Columbic term is usually compensated by the desolvation penalty modeled by the implicit solvent models, either GB or PB. However, classical GB and PB models are not applicable to the highly charged small ionic molecules and the metal plates, and the calculations overestimate the desolvation penalty. The remarkably large desolvation penalty prevents the ligands from binding to the Ag nanoplate. A known problem is that the linear PB model overestimates the ionic strength near the charged surface<sup>78-82</sup>, and further improvement has involved the nonlinear PB equation<sup>83,84</sup>. Notably, the M2 method for computing the binding free energies relies on implicit solvent models because the method needs a well-defined distinct energy well. The explicit water model is not applicable to the method because counting numerous water configurations and explicitly integrating them over the coordinates of solvent molecules to compute a configuration integral  $z_i$  is impractical. Therefore, we used the distance-dependent dielectric (DDD) model that provides crude solvent effects to compute  $\Delta E_{\text{elec}}$ .

Because use of implicit solvent models, i.e. GB or PB, is not practical to compute the solvation free energy, we used MD simulations and linear interaction energy (LIE)

methods to approximate the desolvation penalty. We used MD simulations with an explicit solvent model for citrate and succinate to compute the average Columbic interactions between waters and each ligand to approximate the solvation enthalpy without considering the solvation entropy. The interaction energy is approximated by  $\Delta\langle E_{\text{Colu}}(\text{water-succinate})\rangle$  and  $\Delta\langle E_{\text{Colu}}(\text{water-citrate})\rangle$ , where  $\langle \rangle$  denotes the average interaction energy during the MD simulations. Having one more carboxylate ion largely increases the Columbic attractions between a ligand and water molecules, which yields values for  $\Delta\langle E_{\text{Colu}}(\text{water-succinate})\rangle$  and  $\Delta\langle E_{\text{Colu}}(\text{water-citrate})\rangle$  of -437.4 (20.5) and -808.4 (24.9) kcal/mol, respectively, where the standard deviation is shown in the bracket. The LIE method is then used to approximate the electrostatic contribution to the solvation free energy,  $\Delta G_{\text{sol}}^{\text{el}}$ , and the value equals half of the corresponding ion–water interaction energy:  $\Delta W_{\text{sol}}^{\text{el}} = \frac{1}{2} \Delta\langle E_{\text{Colu}}(\text{water-ligand})\rangle$ <sup>73-75</sup>. The difference in electrostatic contribution to the solvation free energy,  $\Delta\Delta W_{\text{succinate-citrate}}^{\text{el}}$ , between succinate and citrate is ~185.5kcal/mol. This has similar magnitudes of  $\Delta\Delta E_{\text{elec}}$  between the two groups computed with the DDD model. Use of more sophisticated solvent models developed for ions and metal plates is necessary to accurately compute absolute  $\Delta G$ . For example, an implicit solvent model for protein and Au (111) associations developed by Wade and colleagues may be modified and further developed for chemicals and Ag plate associations<sup>41, 42, 85</sup>.

In addition, we investigated gaining solvent entropy upon ligand binding by computing the number of waters on the Ag surface that had been replaced by a ligand. Our calculations showed that similar numbers of water molecules were replaced by different

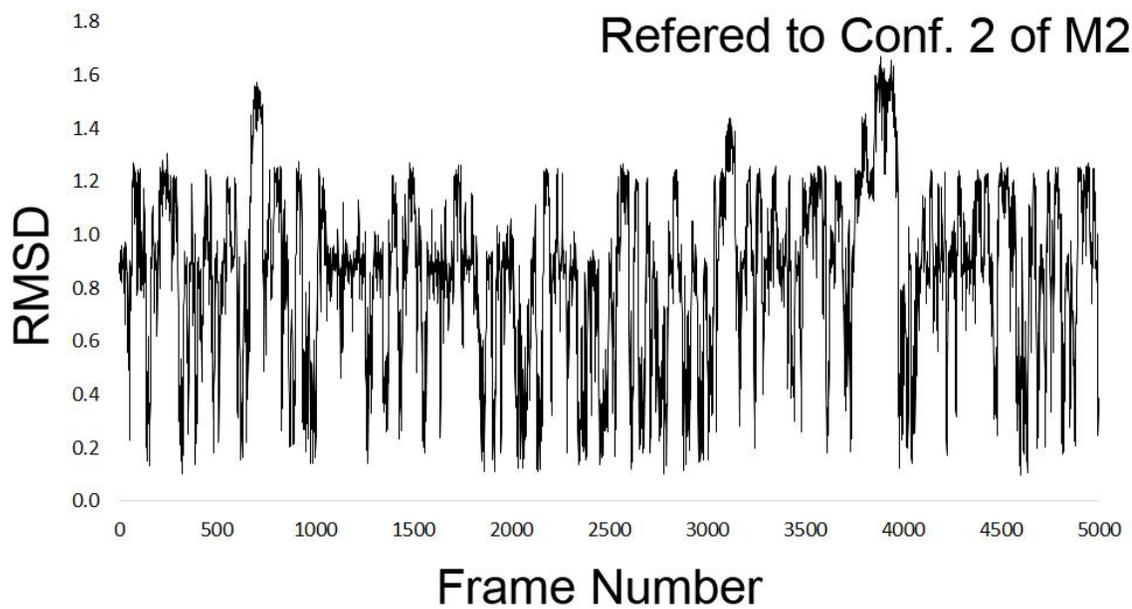
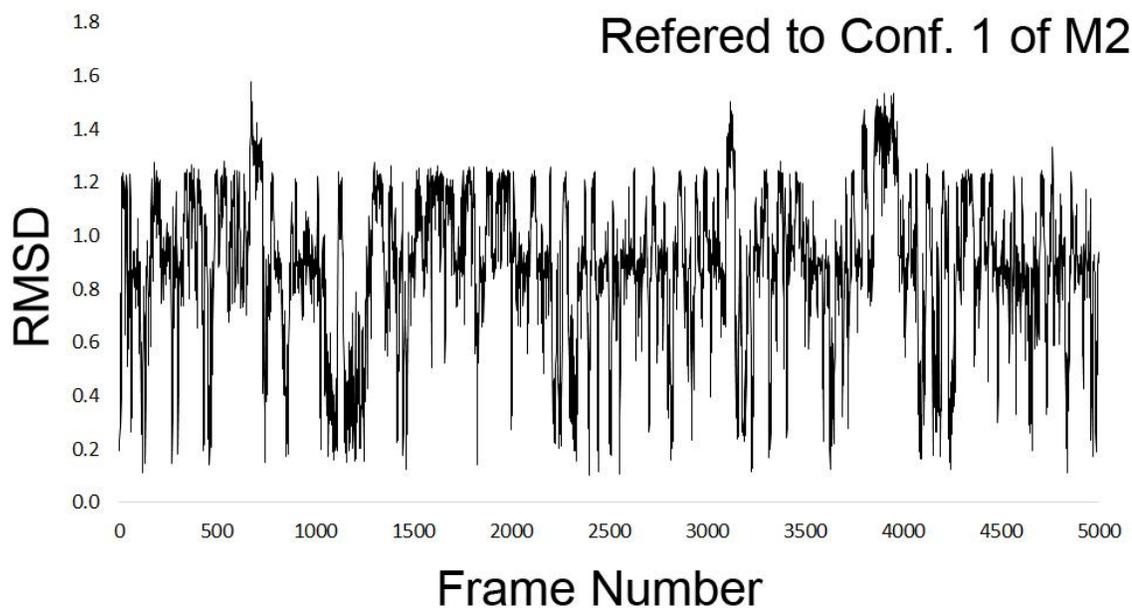
ligands, which suggests that gaining solvent entropy upon ligand binding is not a critical term that determines the relative binding affinities. The result of number of waters in defined region is given in Table 2.5. The differences in numbers of water in defined region reveal the differences of number of waters replaced by two different ligands. For example, comparing with ligands within the di-carboxylate and tri-carboxylate groups, the differences are lesser than 5.5 and 2.3 waters, respectively. Considering the standard deviation, the difference is insignificant. We therefore concluded that the solvent effects from releasing waters on the surface are similar among ligands in each group; therefore, this term does not affect results of related binding affinities.

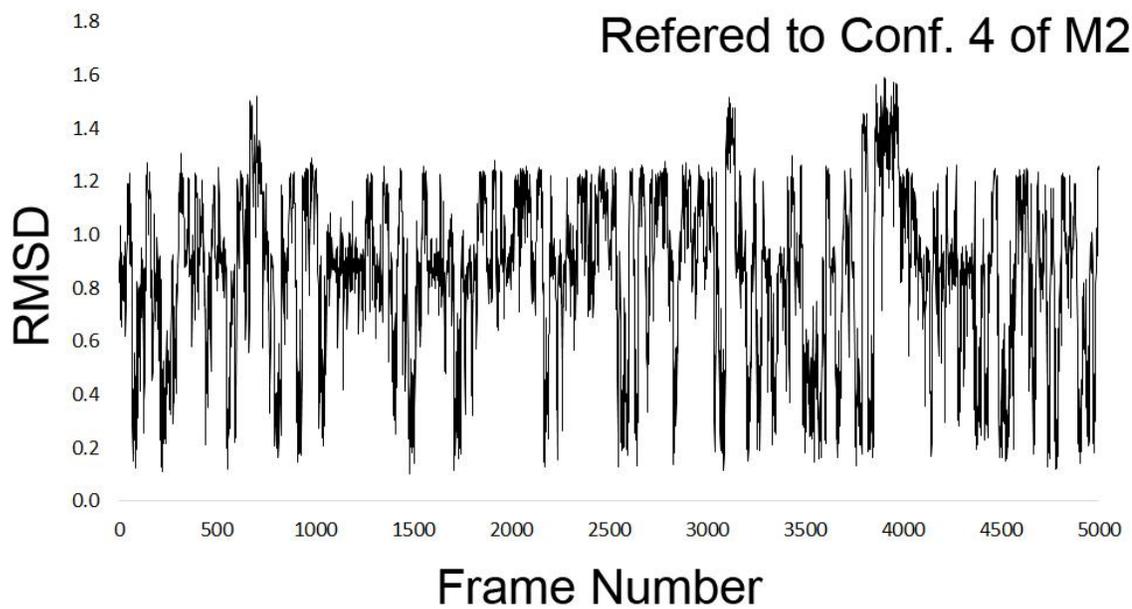
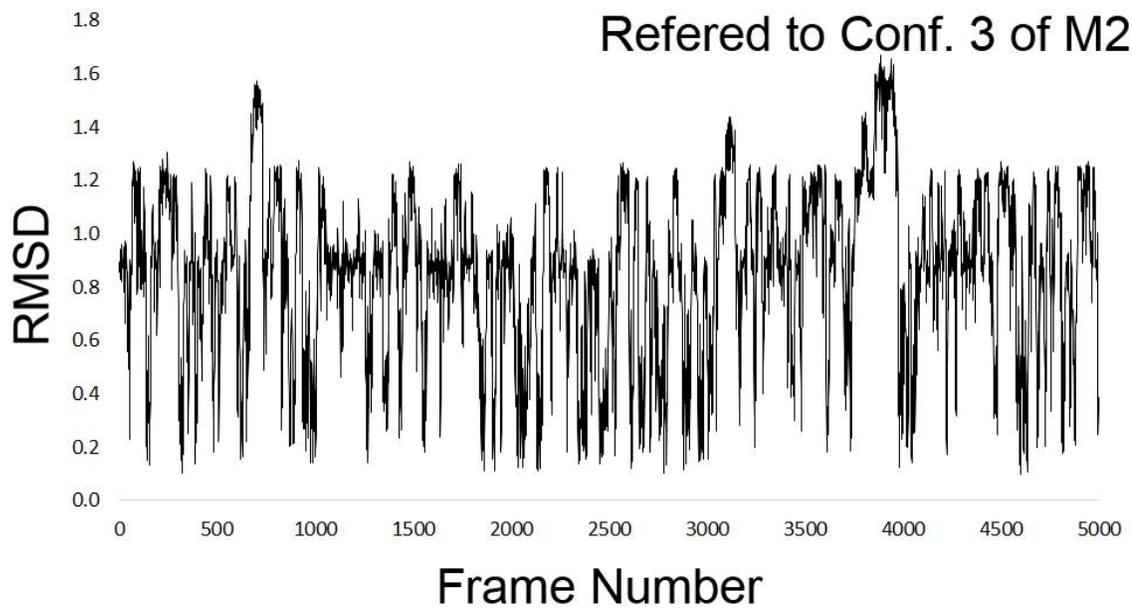
No. of Carboxylate Group	Ligand	No. of Water	Standard Deviation
1	Acetate	59.33	1.79
2	Oxalate	62.81	2.20
	Malonate	61.72	2.28
	Succinate	57.27	1.82
	Glutarate	61.39	2.01
	Pimelate	58.31	2.15
3	Citrate	55.71	1.84
	Isocitrate	53.45	1.77
	Trimesic Acid	53.40	2.06

Table 2.5. Numbers of water in the region that is the first layer (here cut off at 3.2 Å from the surface) water that is within 10 Å from the projection of center of geometry of the ligand on silver surface of silver plate-ligand complexes.

To ensure that the simple DDD solvent model brings similar conformations to those from MD simulations using the explicit solvent model, we compared the conformations of succinate obtained by MD simulations and from M2 conformation search. We computed the root-mean-square-deviation (RMSD) of MD trajectories using the first four lowest energy conformations from M2 search as the reference frames (see Figure 2.4). The comparison illustrated highly similar conformations from both methods, which suggests that the conformations are largely determined by the force field parameters instead of solvent models used in these systems<sup>86</sup>.

Figure 2.4. Plots of RMSD (in Å) computed by all atoms, including hydrogen atoms, for succinate from MD simulations. Each plot used one local energy minima found by M2 as a reference structure. . The frame was saved every 1 ps, and the MD simulation length is 5 ns.





To optimize the non-bonded interactions between a ligand and the (111) facet, ligand flexibility is necessary for a ligand to arrange fitting conformations on the Ag plate (Figure 2.2). For small or more rigid molecules such as oxalate and trimesic acid, the compounds cannot arrange themselves to fit nicely on the Ag plate; thus the  $\Delta E_{\text{Elec}}$  is largely reduced (Table 2.3 and Figure 2.2). Because the ionic attraction is the main factor that determines the relative binding affinities of a series of ligands on the (111) facet, the distribution of partial charges contributes to the strength of ionic attraction. We examined the charge distribution in the di-carboxylate group with the Vcharge model, which is based on atom electronegativity. All ligands in this group carry -2 formal charge; however, compounds with fewer carbon chains show greater negative charge density in the oxygen atoms (Table 2.6). For example, glutarate has a chain of 5 carbon atoms, and the partial charge on the oxygen atoms is -0.72, whereas pimelate has a chain of 7 carbon atoms, and the partial charge on the oxygen atoms is -0.70. To ensure that correlation between the carbon chain lengths and the charge distribution is not an artifact of the Vcharge program, we examined other charge models. The same trend held (Table 2.7). The different charge distribution is a factor of different ionic attraction. For example, the average oxygen–Ag distance for glutarate and pimelate were similar, but the computed binding  $\Delta E_{\text{Elec}}$  for two ligands to the (111) facet differed by ~3 kcal/mol.

Ligand	No. of Carbon	Partial Charge on Oxygen
Oxalate	2	-0.77
Malonate	3	-0.74
Succinate	4	-0.73
Glutarate	5	-0.72
Pimelate	7	-0.70

Table 2.6. The partial charges (Vcharge model) on oxygen atoms in the ligands. The values of partial charges decrease because as a molecule becomes larger, the -2.0 formal charges must be distributed on more atoms, and thus each of the oxygen atoms carries less negative charges.

Ligand	ccpvDZ- b3lyp	AMMP	AM3	PM3	MNDO
Oxalate	-0.492	-0.55	-0.64	-0.69	-0.66
Malonate	-0.489/ -0.475	-0.54	-0.64/-0.63	-0.69/-0.69/ -0.68/-0.67	-0.68/-0.67/ -0.63/-0.61
Succinate	-0.487/ -0.453	-0.54	-0.64/-0.61	-0.69/-0.66	-0.65/-0.63
Glutarate	-0.470/ -0.442	-0.54	-0.62/-0.61	-0.67/-0.66	-0.63
Pimelate	-0.462/ -0.431	-0.54	-0.62/-0.59	-0.67/-0.64	-0.63/-0.61

Table 2.7. Partial charges from other calculation methods. The ccpvDZ-b3lyp is *ab initio* calculated, whereas the remainder are semi-empirical methods. Multiple values indicate that the partial charges for oxygen atoms slightly differed from one another.

Ligand	No. of Atoms	No. of Rotatable Bond	Free Conf. within 3 kcal/mol	Binding Conf. within 3 kcal/mol	-TΔS (kcal/mol)
Oxalate	7	0	1	1	7
Malonate	5	2	1	3	5
Succinate	8	3	6	6	8
Glutarate	8	4	16	11	9
Pimelate	9	6	296	115	9

Table 2.8. Computed configuration entropy loss. The entropy loss increases as the ligands become more flexible. The table lists numbers of free ligand conformations and the bound complexes within 3 kcal/mol of the global energy minimum.

Binding is usually accompanied by a loss of configuration entropy,  $-T\Delta S$ , which opposes the complex formation. Because the Ag plate is held rigid, the configuration entropy loss is solely from reducing the motions in translation, rotation, and vibration degrees of freedom and the number of low-energy conformations of the ligand. Entropic penalty follows a trend whereby molecules with longer chains show larger configuration entropy loss (Tables 2.3 and 2.8). In addition, enthalpy-configuration entropy compensation is observed in this system; stronger binding energy,  $\Delta E_{\text{potential}}$ , also shows a greater loss of configuration entropy, as seen previously<sup>56, 87-89</sup>. Conceptually, more flexible molecules lose configuration entropy more because they reduce larger numbers of conformations when moving from the free to the bound form. In contrast, rigid molecules show insignificant changes in number of states in both the free and bound forms, thus resulting in smaller changes of conformation entropy. Although reducing more numbers of energy wells contributes to larger entropic penalty, the loss of translation, rotation and

vibrational entropy governs the major component of configuration entropy loss. Although the configuration entropic penalty is not the main factor that determines the strength of binding affinities in the systems studied here, the insights are useful for future molecular design.

#### 2.3.4 Partial Charge Investigation

Experiments suggested that the surface of the silver nano-plate should carry partial positive charges. In this study, a partial charge model in which all silver atoms were assigned a +0.5 partial charge was used to parameterize the silver plate. In addition to that, we also performed the same calculations with several other partial charge models, i.e. in which the silver atoms were assigned +0.2, +0.3 or +0.4 partial charge, respectively.

The calculated binding affinities are summarized in Table 2.9 (+0.2 charge), Table 2.10 (+0.3 charge), and Table 2.11 (+0.4 charge).

# of COO-group	Ligand	Facet	$\Delta G$	$\Delta E_{\text{Tot}}$	$\Delta E_{\text{Bonded}}$	$\Delta E_{\text{vdW}}$	$\Delta E_{\text{Elec}}$	$-T\Delta S$	$\Delta G_{111} - \Delta G_{100}$
1	acetic	111	-50.25	-52.07	0.00	-0.51	-51.56	1.82	-0.35
		100	-49.90	-51.19	0.02	-0.56	-50.65	1.30	
2	oxalate	111	-93.34	-97.64	0.09	-1.09	-96.78	4.30	-3.97
		100	-89.37	-93.20	0.11	-0.97	-92.45	3.83	
	malonate	111	-97.18	-101.77	0.02	-1.27	-100.52	4.59	-2.84
		100	-94.34	-98.32	0.01	-1.20	-97.16	3.98	
	succinate	111	-96.64	-101.30	0.12	-1.30	-100.13	4.66	-3.43
		100	-93.21	-98.42	0.14	-1.16	-97.43	5.21	
	glutarate	111	-95.34	-100.78	0.31	-2.56	-98.57	5.44	-2.91
		100	-92.43	-97.89	0.30	-2.54	-95.72	5.46	
	pimelate	111	-93.60	-100.10	0.07	-2.26	-97.92	6.51	-2.82
		100	-90.78	-99.19	0.08	-2.20	-97.07	8.41	
3	citrate	111	-134.51	-143.45	2.74	-2.08	-144.12	8.93	-2.78
		100	-131.73	-139.62	1.42	-1.81	-139.25	7.89	
	isocitrate	111	-134.23	-140.98	4.01	-1.37	-143.64	6.75	-1.53
		100	-132.70	-138.90	1.33	-1.07	-139.18	6.20	
	trimesic	111	-133.68	-141.73	1.04	-4.18	-138.77	8.05	-5.67
		100	-128.01	-134.65	1.38	-4.10	-132.09	6.64	

Table 2.9. Binding energy components of ligands with +0.2 partial charge set. The unit of binding energy terms and entropy loss are kcal/mol.

# of COO-group	Ligand	Facet	$\Delta G$	$\Delta E_{\text{Tot}}$	$\Delta E_{\text{Bonded}}$	$\Delta E_{\text{vdW}}$	$\Delta E_{\text{Elec}}$	$-T\Delta S$	$\Delta G_{111} - \Delta G_{100}$	
1	acetic	111	-77.28	-78.12	-0.04	0.65	-78.73	0.84	-2.56	
		100	-74.72	-76.88	-0.05	0.64	-77.45	2.16		
2	oxalate	111	-141.38	-146.85	0.03	1.05	-148.25	5.47	-6.50	
		100	-134.88	-139.75	0.02	1.03	-141.03	4.87		
	malonate	111	-147.00	-152.79	-0.04	0.94	-153.75	5.78	-3.80	
		100	-143.20	-147.80	-0.08	0.95	-148.79	4.60		
	succinate	111	-146.83	-152.83	0.04	1.08	-153.98	6.00	-4.76	
		100	-142.07	-148.68	0.14	1.37	-150.24	6.60		
	glutarate	111	-145.36	-151.35	0.35	0.16	-151.89	5.99	-4.55	
		100	-140.81	-148.29	0.74	0.56	-149.60	7.47		
	pimelate	111	-142.86	-150.57	0.10	0.11	-150.78	7.71	-4.27	
		100	-138.59	-148.37	0.11	0.13	-148.62	9.78		
	3	citrate	111	-208.22	-219.86	3.97	1.93	-225.76	11.64	-5.51
			100	-202.71	-212.59	2.89	3.37	-218.87	9.89	
isocitrate		111	-207.82	-217.42	5.45	1.93	-224.79	9.60	-4.61	
		100	-203.21	-211.18	3.37	3.92	-218.52	7.97		
trimesic		111	-202.19	-212.45	2.62	-1.06	-214.42	10.25	-8.12	
		100	-194.07	-202.06	2.70	-1.64	-203.47	7.99		

Table 2.10. Binding energy components of ligands with +0.3 partial charge set. The unit of binding energy terms and entropy loss are kcal/mol.

# of COO-group	Ligand	Facet	$\Delta G$	$\Delta E_{\text{Tot}}$	$\Delta E_{\text{Bonded}}$	$\Delta E_{\text{vdW}}$	$\Delta E_{\text{Elec}}$	$-T\Delta S$	$\Delta G_{111} - \Delta G_{100}$
1	acetic	111	-102.96	-104.50	-0.20	1.93	-106.23	1.54	-2.86
		100	-100.10	-102.77	-0.24	1.93	-104.45	2.66	
2	oxalate	111	-190.35	-196.68	-0.21	3.39	-200.41	6.32	-8.78
		100	-181.58	-187.37	-0.28	3.15	-190.63	5.80	
	malonate	111	-198.37	-204.58	-0.34	3.39	-207.79	6.21	-5.91
		100	-192.46	-197.90	-0.32	3.32	-201.16	5.44	
	succinate	111	-197.91	-205.03	-0.21	3.68	-208.57	7.12	-5.40
		100	-192.51	-200.28	-0.03	4.15	-204.51	7.77	
	glutarate	111	-196.33	-203.36	0.03	2.83	-206.23	7.03	-5.23
		100	-191.10	-199.78	0.37	3.24	-203.39	8.68	
	pimelate	111	-193.15	-201.68	-0.21	2.61	-204.08	8.53	-4.18
		100	-188.98	-199.24	-0.22	2.64	-201.66	10.26	
3	citrate	111	-284.36	-296.93	3.53	5.83	-306.29	12.57	-8.05
		100	-276.31	-287.55	2.80	7.77	-298.17	11.24	
	isocitrate	111	-283.95	-294.71	4.75	6.44	-305.90	10.76	-6.72
		100	-277.23	-286.61	3.83	9.05	-299.56	9.38	
	trimesic	111	-273.10	-284.82	3.78	2.45	-291.77	11.73	-11.07
		100	-262.03	-271.18	3.53	1.27	-276.59	9.15	

Table 2.11. Binding energy components of ligands with +0.4 partial charge set. The unit of binding energy terms and entropy loss are kcal/mol.

In +0.2 charge model, the relative binding affinities in di-carboxylate group holds the basic trend of that interpreted from the experimental yields except that malonate has a stronger binding affinity than succinate. This is similar to the result from +0.5 partial charge, but the difference of the values are smaller than the report +0.5 partial charge set. In addition, in the tri-carboxylate group, the binding affinities of citrate and isocitrate match the experimental yields, but trimesic acid is the singularity.

Similar behaviors can be observed in the +0.3 and +0.4 partial charge model, but the difference in relative binding affinities in di-carboxylate group becomes larger and larger. Furthermore, the higher partial charge on silver atoms start to differentiate the trimesic acid from the other two ligand in tri-carboxylate group.

This results from the ratio of contributions from the Columbic interactions and vdW interactions between the ligand and the surface. The Columbic attraction between silver surface and the ligand is mainly from the Columbic attraction between the oxygen atoms on the ligands and the silver atoms, while the vdW attraction is mainly from the interaction between carbon and hydrogen atoms on the ligands and the silver atoms. By comparing the +0.2, +0.3 and +0.4 charge, one can see that the columbic interaction becomes stronger and stronger. This higher positive partial charge makes oxygen-silver attraction more favorable, but in turn, makes the columbic repulsion between positively charged silver atoms and carbon, hydrogen atoms stronger. So higher positive charge will lead to the behavior of glutarate in Figure 2.5. In +0.2 charge model, the middle carbon in the chain stays on the surface, but in +0.3 and +0.4 charge models, it is push away from

the surface, due to the stronger repulsion. This behavior weakens the vdW interaction from +0.2 to +0.4 partial charge. Also in the case of trimesic acid (Figure 2.6), the rigid molecule is unable to adjust itself to maximize the columbic attraction between oxygen and silver atoms and at the same time minimize the columbic repulsion between the silver atoms and the carbon or hydrogen. This is the intrinsic reason for trimesic acid to be a weak surfactant, and only with high positive charge model the behavior appears. Based on experimental data and existing studies, we selected +0.5 partial charge model for this study that yield both more accurate experimental trends and conformations.

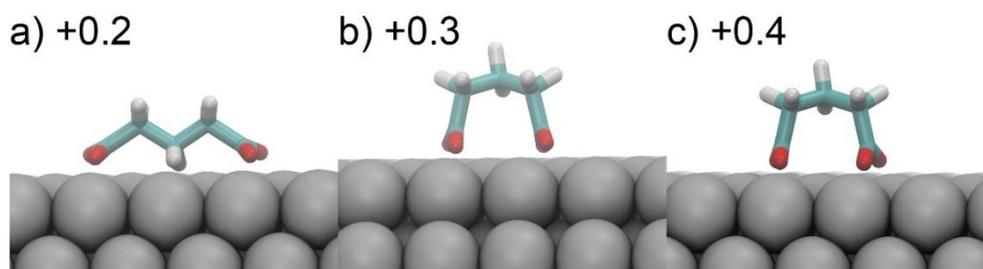


Figure 2.5. Binding pose of glutarate on silver plate in +0.2, +0.3 and +0.4 partial charge model.

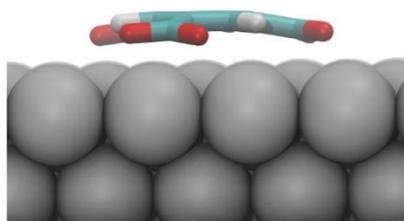


Figure 2.6. Binding pose of trimesic acid on silver plate in +0.2 partial charge model.

### 2.3.5 Design of Novel Surfactants

From findings for the series of ligands, we designed two new molecules that may be potentially effective ligands for future application (see Figure 2.2 and Table 2.2). To begin, we used trimesic acid as a template for design because it has a ring in the center as a rigid core and the length and functional groups of the side-chains may be used to optimize binding affinity to the (111) facet. Our calculations showed that the central flat aromatic ring hampered a possible side-chain rearrangement to form a (111) facet–ligand complex with good geometric complementation. Therefore, we replaced the aromatic ring by a hexane ring and retained the same side-chains as our first designed ligand, cyclohexane-1,3,5-tri-carboxylate. The calculations then revealed that the lengths of the side-chains were too short for the three carboxylate ions to fill the hollow site between three Ag atoms. Therefore, we lengthened the side chain with one more carbon to build our second designed ligand, 2,2',2''-(cyclohexane-1,3,5-triyl)triacetate. According to the calculation result in Table 2.3, ligand 2 should be able to bind as well as citrate to the (111) facet. The non-bonded terms suggested that the electrostatic interaction for designed ligand 2 was less favorable than that for citrate. However, the bonded terms favor binding, because the higher flexibility of designed ligand 2 than citrate lowers the penalty caused by distorting the ligand while adjusting itself on the Ag nanoplate surface. In addition, high flexibility in the side-chains inevitably increased the entropic penalty of binding ligand 2 to the plate, and the entropy loss could not be maintained as small as that in trimesic acid. By comparing the designed ligand 2 and citrate, we conclude that it is difficult to design new ligand that may bind tighter than citrate for this specific system

because finding small molecules with both favorable energy and small entropic penalty upon binding is challenging. Designing large ligands may satisfy better geometry fitness; however, the entropic penalty would oppose binding. In contrast, a rigid molecule efficiently reduces entropic penalty, but the requirement of geometry complementary is not likely achieved. In summary, citrate may be the best tri-carboxylate ligand.

## 2.4 Conclusions

This study involved use of the M2 method to calculate the relative binding free energies of a series of ligands on a silver cluster. The calculations support the experimental trends of anisotropic growth of the Ag plates from the perspective of ligand-nanoplate interactions, and provide insights into the mechanisms of directional anisotropic growth and factors that govern ligand-nanoplate binding. In summary, a good ligand needs to have a strong binding affinity with the Ag cluster to protect the surface against growth and also show strong preference in binding to the (111) or (100) facet to allow the plate to grow in one direction. The ligands investigated in this study use electrostatic attractions to drive (111) facet–ligand binding and to distinguish the (111) and (100) facet. The ligands need to retain certain flexibility to rearrange themselves, especially the oxygen atoms of the carboxylate ions, to fit on the Ag surface. Therefore, achieving this requirement with very small or rigid molecules may be difficult. As well, flexible molecules that may cause a large configuration entropy loss in binding are not ideal. In addition, we attempted to design derivatives of trimesic acid as potential effective ligands

and predicted the relative binding affinities of the molecules. We discussed strategies for more accurate modeling in binding chemicals on the Ag plates.

## 2.5 References

1. Burda, C., Chen, X. B., Narayanan, R., and El-Sayed, M. A. (2005) Chemistry and Properties of Nanocrystals of Different Shapes. *Chemical Reviews* **105**, 1025-1102
2. Hu, J. T., Li, L. S., Yang, W. D., Manna, L., Wang, L. W., and Alivisatos, A. P. (2001) Linearly Polarized Emission from Colloidal Semiconductor Quantum Rods. *Science* **292**, 2060-2063
3. Jin, R. C., Cao, Y. C., Hao, E. C., Metraux, G. S., Schatz, G. C., and Mirkin, C. A. (2003) Controlling Anisotropic Nanoparticle Growth Through Plasmon Excitation. *Nature* **425**, 487-490
4. Zhang, L., Gu, F. X., Chan, J. M., Wang, A. Z., Langer, R. S., and Farokhzad, O. C. (2008) Nanoparticles in Medicine: Therapeutic Applications and Developments. *Clinical Pharmacology & Therapeutics* **83**, 761-769
5. Shiraishi, Y., and Toshima, N. (2000) Oxidation of Ethylene Catalyzed by Colloidal Dispersions of Poly(Sodium Acrylate)-Protected Silver Nanoclusters. *Colloids and Surfaces a-Physicochemical and Engineering Aspects* **169**, 59-66
6. Reiss, G., and Hutten, A. (2005) Magnetic Nanoparticles - Applications Beyond Data Storage. *Nature Materials* **4**, 725-726
7. Jin, R. C., Cao, Y. W., Mirkin, C. A., Kelly, K. L., Schatz, G. C., and Zheng, J. G. (2001) Photoinduced Conversion Of Silver Nanospheres To Nanoprisms. *Science* **294**, 1901-1903
8. Schultz, S., Smith, D. R., Mock, J. J., and Schultz, D. A. (2000) Single-Target Molecule Detection with Nonbleaching Multicolor Optical Immunolabels.

*Proceedings of the National Academy of Sciences of the United States of America*  
**97**, 996-1001

9. Cho, C. Y., Kwon, M. K., Lee, S. J., Han, S. H., Kang, J. W., Kang, S. E., Lee, D. Y., and Park, S. J. (2010) Surface Plasmon-Enhanced Light-Emitting Diodes Using Silver Nanoparticles Embedded in p-GaN. *Nanotechnology* **21**, 5
10. Chen, S. H., and Carroll, D. L. (2002) Synthesis and Characterization of Truncated Triangular Silver Nanoplates. *Nano Letters* **2**, 1003-1007
11. Sun, Y. G., Mayers, B., and Xia, Y. N. (2003) Transformation of Silver Nanospheres into Nanobelts and Triangular Nanoplates Through a Thermal Process. *Nano Letters* **3**, 675-679
12. Xiong, Y. J., Siekkinen, A. R., Wang, J. G., Yin, Y. D., Kim, M. J., and Xia, Y. N. (2007) Synthesis of Silver Nanoplates at High Yields by Slowing Down the Polyol Reduction of Silver Nitrate with Polyacrylamide. *Journal of Materials Chemistry* **17**, 2600-2602
13. Sun, Y., and Wiederrecht, G. P. (2007) Surfactantless Synthesis of Silver Nanoplates and Their Application in SERS. *Small* **3**, 1964-1975
14. Liu, G., Cai, W., and Liang, C. (2008) Trapeziform Ag Nanosheet Arrays Induced by Electrochemical Deposition on Au-Coated Substrate. *Crystal Growth & Design* **8**, 2748-2752
15. Rocha, T. C. R., and Zanchet, D. (2007) Structural Defects and Their Role in the Growth of Ag Triangular Nanoplates. *Journal of Physical Chemistry C* **111**, 6989-6993
16. Jia, H. Y., Zeng, J. B., An, J., Song, W., Xu, W. Q., and Zhao, B. (2008) Preparation of Triangular and Hexagonal Silver Nanoplates on the Surface of Quartz Substrate. *Thin Solid Films* **516**, 5004-5009

17. Maillard, M., Huang, P. R., and Brus, L. (2003) Silver Nanodisk Growth by Surface Plasmon Enhanced Photoreduction of Adsorbed Ag<sup>+</sup>. *Nano Letters* **3**, 1611-1615
18. Sun, Y. G., and Xia, Y. N. (2003) Triangular Nanoplates of Silver: Synthesis, Characterization, and Use As Sacrificial Templates for Generating Triangular Nanorings of Gold. *Advanced Materials* **15**, 695-699
19. Xue, C., Metraux, G. S., Millstone, J. E., and Mirkin, C. A. (2008) Mechanistic Study of Photomediated Triangular Silver Nanoprism Growth. *Journal of the American Chemical Society* **130**, 8337-8344
20. Zhang, J. A., Langille, M. R., and Mirkin, C. A. (2010) Photomediated Synthesis of Silver Triangular Bipyramids and Prisms: The Effect of pH and BSPP. *Journal of the American Chemical Society* **132**, 12502-12510
21. Jiang, L. P., Xu, S., Zhu, J. M., Zhang, J. R., Zhu, J. J., and Chen, H. Y. (2004) Ultrasonic-Assisted Synthesis of Monodisperse Single-Crystalline Silver Nanoplates and Gold Nanorings. *Inorganic Chemistry* **43**, 5877-5883
22. Washio, I., Xiong, Y., Yin, Y., and Xia, Y. (2006) Reduction by the End Groups of Poly(Vinyl Pyrrolidone): A New and Versatile Route to the Kinetically Controlled Synthesis of Ag Triangular Nanoplates. *Advanced Materials* **18**, 1745-  
+
23. Lu, Q., Lee, K.-J., Lee, K.-B., Kim, H.-T., Lee, J., Myung, N. V., and Choa, Y.-H. (2010) Investigation of Shape Controlled Silver Nanoplates by a Solvothermal Process. *Journal of Colloid and Interface Science* **342**, 8-17
24. Chen, D., Li, L., Liu, J., Qi, S., Tang, F., Ren, X., Wu, W., Ren, J., and Zhang, L. (2007) Synthesis and Self-Assembly of Monodisperse Silver-Nanocrystal-Doped Silica Particles. *Journal of Colloid and Interface Science* **308**, 351-355

25. Liang, H. P., Wan, L. J., Bai, C. L., and Jiang, L. (2005) Gold Hollow Nanospheres: Tunable Surface Plasmon Resonance Controlled by Interior-Cavity Sizes. *Journal of Physical Chemistry B* **109**, 7795-7800
26. Hao, E. C., Kelly, K. L., Hupp, J. T., and Schatz, G. C. (2002) Synthesis of Silver Nanodisks Using Polystyrene Mesospheres as Templates. *Journal of the American Chemical Society* **124**, 15182-15183
27. Goebel, J., Zhang, Q., He, L., and Yin, Y. D. (2012) Monitoring the Shape Evolution of Silver Nanoplates: A Marker Study. *Angewandte Chemie-International Edition* **51**, 552-555
28. Zhang, Q., Hu, Y., Guo, S., Goebel, J., and Yin, Y. (2010) Seeded Growth of Uniform Ag Nanoplates with High Aspect Ratio and Widely Tunable Surface Plasmon Bands. *Nano Letters* **10**, 5037-5042
29. Zhang, Q., Li, N., Goebel, J., Lu, Z., and Yin, Y. (2011) A Systematic Study of the Synthesis of Silver Nanoplates: Is Citrate a "Magic" Reagent? *Journal of the American Chemical Society* **133**, 18931-18939
30. Kilin, D. S., Prezhdo, O. V., and Xia, Y. (2008) Shape-Controlled Synthesis of Silver Nanoparticles: Ab Initio Study of Preferential Surface Coordination with Citric Acid. *Chemical Physics Letters* **458**, 113-116
31. Al-Saidi, W. A., Feng, H., and Fichthorn, K. A. (2012) Adsorption of Polyvinylpyrrolidone on Ag Surfaces: Insight into a Structure-Directing Agent. *Nano Letters* **12**, 997-1001
32. Sun, Y. G., Mayers, B., Herricks, T., and Xia, Y. N. (2003) Polyol Synthesis of Uniform Silver Nanowires: A Plausible Growth Mechanism and the Supporting Evidence. *Nano Letters* **3**, 955-960

33. Pagliai, M., Caporali, S., Muniz-Miranda, M., Pratesi, G., and Schettino, V. (2012) SERS, XPS, and DFT Study of Adenine Adsorption on Silver and Gold Surfaces. *Journal of Physical Chemistry Letters* **3**, 242-245
34. Muniz-Miranda, M., Gellini, C., Pagliai, M., Innocenti, M., Salvi, P. R., and Schettino, V. (2010) SERS and Computational Studies on MicroRNA Chains Adsorbed on Silver Surfaces. *Journal of Physical Chemistry C* **114**, 13730-13735
35. Thomas, S., Biswas, N., Venkateswaran, S., Kapoor, S., Naumov, S., and Mukherjee, T. (2005) Studies on Adsorption of 5-Amino Tetrazole on Silver Nanoparticles by SERS and DFT Calculations. *Journal of Physical Chemistry A* **109**, 9928-9934
36. Ma, W., and Fang, Y. (2006) Experimental (FT-IR) and Theoretical (DFT) Studies on the Adsorption Behavior of p-Nitroaniline (PNA) on Gold Nanoparticles. *Journal of Nanoparticle Research* **8**, 761-767
37. Shuang, Z., Zhen-Hua, L., Wen-Ning, W., and Kang-Nian, F. (2005) Density Functional Study of the Interaction of Chlorine Atom with Small Neutral and Charged Silver Clusters. *Journal of Chemical Physics* **122**, 144701-144701-144708
38. Ignaczak, A. (2001) DFT Calculations of the Interaction of Alkali Ions with Copper and Silver. *Journal of Electroanalytical Chemistry* **495**, 160-168
39. Matulis, V. E., and Ivaskevich, O. A. (2006) Comparative DFT Study of Electronic Structure and Geometry of Copper and Silver Clusters: Interaction with NO Molecule. *Computational Materials Science* **35**, 268-271
40. Schravendijk, P., Ghiringhelli, L. M., Delle Site, L., and van der Vegt, N. F. A. (2007) Interaction of Hydrated Amino Acids with Metal Surfaces: A Multiscale Modeling Description. *Journal of Physical Chemistry C* **111**, 2631-2642

41. Iori, F., Di Felice, R., Molinari, E., and Corni, S. (2009) GoIP: An Atomistic Force-Field to Describe the Interaction of Proteins With Au(111) Surfaces in Water. *Journal of Computational Chemistry* **30**, 1465-1476
42. Kokh, D. B., Corni, S., Winn, P. J., Hoefling, M., Gottschalk, K. E., and Wade, R. C. (2010) ProMetCS: An Atomistic Force Field for Modeling Protein-Metal Surface Interactions in a Continuum Aqueous Solvent. *Journal of Chemical Theory and Computation* **6**, 1753-1768
43. Heinz, H., Vaia, R. A., Farmer, B. L., and Naik, R. R. (2008) Accurate Simulation of Surfaces and Interfaces of Face-Centered Cubic Metals Using 12-6 and 9-6 Lennard-Jones Potentials. *Journal of Physical Chemistry C* **112**, 17281-17290
44. Heinz, H., Jha, K. C., Luettmmer-Strathmann, J., Farmer, B. L., and Naik, R. R. (2011) Polarization at Metal-Biomolecular Interfaces in Solution. *Journal of the Royal Society Interface* **8**, 220-232
45. Akimov, A. V., Nemukhin, A. V., Moskovsky, A. A., Kolomeisky, A. B., and Tour, J. M. (2008) Molecular Dynamics of Surface-Moving Thermally Driven Nanocars. *Journal of Chemical Theory and Computation* **4**, 652-656
46. Cohavi, O., Corni, S., De Rienzo, F., Di Felice, R., Gottschalk, K. E., Hoefling, M., Kokh, D., Molinari, E., Schreiber, G., Vaskevich, A., and Wade, R. C. (2010) Protein-Surface Interactions: Challenging Experiments and Computations. *Journal of Molecular Recognition* **23**, 259-262
47. Di Felice, R., and Corni, S. (2011) Simulation of Peptide-Surface Recognition. *Journal of Physical Chemistry Letters* **2**, 1510-1519
48. Wright, L. B., Rodger, P. M., Corni, S., and Walsh, T. R. (2013) GoIP-CHARMM: First-Principles Based Force Fields for the Interaction of Proteins

with Au(111) and Au(100). *Journal of Chemical Theory and Computation* **9**, 1616-1630

49. Siwko, M. E., and Corni, S. (2013) Cytochrome C on a Gold Surface: Investigating Structural Relaxations and Their Role in Protein-Surface Electron Transfer by Molecular Dynamics Simulations. *Physical Chemistry Chemical Physics* **15**, 5945-5956
50. Barone, V., Casarin, M., Forrer, D., Monti, S., and Prampolini, G. (2011) Molecular Dynamics Simulations of the Self-Assembly of Tetraphenylporphyrin-Based Mono Layers and Bilayers at A Silver Interface. *Journal of Physical Chemistry C* **115**, 18434-18444
51. Zeng, Q., Jiang, X., Yu, A., and Lu, G. (2007) Growth Mechanisms of Silver Nanoparticles: a Molecular Dynamics Study. *Nanotechnology* **18**
52. Baletto, F., Mottet, C., and Ferrando, R. (2000) Molecular Dynamics Simulations of Surface Diffusion and Growth on Silver and Gold Clusters. *Surface Science* **446**, 31-45
53. Chang, C. E., and Gilson, M. K. (2003) Tork: Conformational Analysis Method for Molecules and Complexes. *Journal of Computational Chemistry* **24**, 1987-1998
54. Chang, C. E., Potter, M. J., and Gilson, M. K. (2003) Calculation of Molecular Configuration Integrals. *Journal of Physical Chemistry B* **107**, 1048-1055
55. Chang, C. E., and Gilson, M. K. (2004) Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *Journal of the American Chemical Society* **126**, 13156-13164

56. Chen, W., Chang, C. E., and Gilson, M. K. (2004) Calculation of Cyclodextrin Binding Affinities: Energy, Entropy, and Implications for Drug Design. *Biophysical Journal* **87**, 3035-3049
57. Chen, W., Gilson, M. K., Webb, S. P., and Potter, M. J. (2010) Modeling Protein-Ligand Binding by Mining Minima. *Journal of Chemical Theory and Computation* **6**, 3540-3557
58. Huang, Y.-m. M., Chen, W., Potter, M. J., and Chang, C.-e. A. (2012) Insights from Free-Energy Calculations: Protein Conformational Equilibrium, Driving Forces, and Ligand-Binding Modes. *Biophysical Journal* **103**, 342-351
59. Pedretti, A., Villa, L., and Vistoli, G. (2004) VEGA - An Open Platform to Develop Chemo-Bio-Informatics Applications, Using Plug-In Architecture and Script Programming. *Journal of Computer-Aided Molecular Design* **18**, 167-173
60. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **4**, 187-217
61. Gilson, M. K., Gilson, H. S. R., and Potter, M. J. (2003) Fast Assignment of Accurate Partial Atomic Charges: An Electronegativity Equalization Method That Accounts for Alternate Resonance Forms. *Journal of Chemical Information and Computer Sciences* **43**, 1982-1997
62. Liu, L., and Bassett, W. A. (1973) Compression of Ag and Phase-Transformation of NaCl. *Journal of Applied Physics* **44**, 1475-1479
63. Swanson, H. E., Morris, M. C., and Evans, E. H. (1966) Standard X-Ray Diffraction Power Patterns. Section 4-Data for 103 Substances. *National Bureau of Standards Monograph*, 83 pp.-83 pp.

64. Schweighofer, K. J., Xia, X. F., and Berkowitz, M. L. (1996) Molecular Dynamics Study of Water Next to Electrified Ag(111) Surfaces. *Langmuir* **12**, 3747-3752
65. Sanchez, C. G. (2003) Molecular Reorientation of Water Adsorbed on Charged Ag(111) Surfaces. *Surface Science* **527**, 1-11
66. Potter, M. J., and Gilson, M. K. (2002) Coordinate Systems and The Calculation of Molecular Properties. *Journal of Physical Chemistry A* **106**, 563-566
67. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **47**, 1739-1749
68. Meiler, J., and Baker, D. (2006) ROSETTALIGAND: Protein-Small Molecule Docking with Full Side-Chain Flexibility. *Proteins-Structure Function and Bioinformatics* **65**, 538-548
69. Wang, J., Kollman, P. A., and Kuntz, I. D. (1999) Flexible Ligand Docking: A Multistep Strategy Approach. *Proteins-Structure Function and Genetics* **36**, 1-19
70. Gilson, M. K., and Zhou, H. X. (2007) Calculation of Protein-Ligand Binding Affinities. in *Annual Review of Biophysics and Biomolecular Structure*, Annual Reviews, Palo Alto. pp 21-42
71. Huang, S.-Y., Grinter, S. Z., and Zou, X. (2010) Scoring Functions and Their Evaluation Methods for Protein-Ligand Docking: Recent Advances and Future Directions. *Physical Chemistry Chemical Physics* **12**, 12899-12908

72. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781-1802
73. Warshel, A., and Russell, S. T. (1984) Calculations of Electrostatic Interactions in Biological-Systems and in Solutions. *Quarterly Reviews of Biophysics* **17**, 283-422
74. Aqvist, J., Medina, C., and Samuelsson, J. E. (1994) New Method for Predicting Binding-Affinity in Computer-Aided Drug Design. *Protein Engineering* **7**, 385-391
75. Roux, B., Yu, H. A., and Karplus, M. (1990) Molecular-Basis for the Born Model of Ion Solvation. *Journal of Physical Chemistry* **94**, 4683-4688
76. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics & Modelling* **14**, 33-38
77. Chen, W., Huang, J., and Gilson, M. K. (2004) Identification of Symmetries in Molecules and Complexes. *Journal of Chemical Information and Computer Sciences* **44**, 1301-1313
78. Borukhov, I., Andelman, D., and Orland, H. (1997) Steric Effects in Electrolytes: A Modified Poisson-Boltzmann Equation. *Physical Review Letters* **79**, 435-438
79. Ziebert, F., and Lacoste, D. (2010) A Poisson-Boltzmann Approach for a Lipid Membrane in an Electric Field. *New Journal of Physics* **12**
80. Fogolari, F., Zuccato, P., Esposito, G., and Viglino, P. (1999) Biomolecular Electrostatics with the Linearized Poisson-Boltzmann Equation. *Biophysical Journal* **76**, 1-16

81. Larsson, D. S. D., and van der Spoel, D. (2012) Screening for the Location of RNA Using the Chloride Ion Distribution in Simulations of Virus Capsids. *Journal of Chemical Theory and Computation* **8**, 2474-2483
82. Fenley, M. O., Harris, R. C., Jayaram, B., and Boschitsch, A. H. (2010) Revisiting the Association of Cationic Groove-Binding Drugs to DNA Using a Poisson-Boltzmann Approach. *Biophysical Journal* **99**, 879-886
83. Rocchia, W., Alexov, E., and Honig, B. (2001) Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *Journal of Physical Chemistry B* **105**, 6507-6514
84. Boschitsch, A. H., and Fenley, M. O. (2004) Hybrid Boundary Element and Finite Difference Method for Solving the Nonlinear Poisson-Boltzmann Equation. *Journal of Computational Chemistry* **25**, 935-955
85. Cohavi, O., Reichmann, D., Abramovich, R., Tesler, A. B., Bellapadrona, G., Kokh, D. B., Wade, R. C., Vaskevich, A., Rubinstein, I., and Schreiber, G. (2011) A Quantitative, Real-Time Assessment of Binding of Peptides and Proteins to Gold Surfaces. *Chemistry-a European Journal* **17**, 1327-1336
86. Chang, C. E., Chen, W., and Gilson, M. K. (2005) Evaluating the Accuracy of the Quasiharmonic Approximation. *Journal of Chemical Theory and Computation* **1**, 1017-1028
87. Rekharsky, M. V., and Inoue, Y. (1998) Complexation Thermodynamics of Cyclodextrins. *Chemical Reviews* **98**, 1875-1917
88. Rekharsky, M., and Inoue, Y. (2000) Chiral Recognition Thermodynamics of Beta-Cyclodextrin: the Thermodynamic Origin of Enantioselectivity and the Enthalpy-Entropy Compensation Effect. *Journal of the American Chemical Society* **122**, 4418-4435

89. Williams, D. H., Stephens, E., O'Brien, D. P., and Zhou, M. (2004) Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angewandte Chemie-International Edition* **43**, 6596-6616

## **Chapter 3 A Modeling Investigation on Inhibitor MitoBloCK-10 Binding in Wildtype and Mutants of Tim44**

### 3.1 Introduction

Protein import is highly regulated by translocons on the outer and inner membranes of mitochondria <sup>1</sup>. The Translocase of the Outer Membrane (TOM) is the major gateway for mitochondria protein import. The Translocase of the Inner Membrane (TIM23), which resides in the matrix, mediates the import of proteins which typically have an N-terminal targeting sequence <sup>2</sup>. The two complex systems coordinate the assembly of the mitochondrion.

Complete translocation of the protein across the inner membrane requires the membrane potential and the ATP-driven protein-associated motor (PAM) <sup>3</sup>. The PAM is a dynamic and highly regulated protein complex that is composed of TIM44, mitochondrial hsp70, the nucleotide exchange factor mGrpE (Mge1) and J-proteins Pam16 and Pam18 <sup>4-8</sup>. Among them, TIM44 is upregulated in diabetes and can be used as drug target <sup>9</sup>.

Our collaborator conducted drug screening for TIM44 and identified a potential drug, MitoBloCK-10 (MB-10), and investigated a variety of biological activities of MB-10 by *in vitro* and mutation studies. MB-10 inhibited the import of substrates that use the TIM23 translocon but not those imported via other pathways. MB-10 targeted TIM44 and reduced the import of substrates that require the PAM complex. In mammalian cells,

MB-10 treatment supports an important role for TIM44 in the import of SOD2. Finally, TIM44 was required for heart development in zebrafish.

A series of mutants of the wild type TIM44 was obtained in the mutation study. Among them, 3H1 (T290S), 4E5 (I297), 3B7 (F284L) and 3G5 (V298A) show resistance with 10  $\mu$ M treatment of MB-10 while the wildtype (WT) is completely inhibited. Mutants 3H1 and 4E5 demonstrated partial resistance upon 15  $\mu$ M MB-10 treatment while the other two mutants were completely inhibited.

Computational methods are widely used in the study of conformational search for ligand-protein binding and binding affinity evaluation. Among them, docking is a fast and efficient method which systematically or stochastically searches for appropriate binding site in the protein and the low energy bound conformations<sup>10,11</sup>. It utilizes a scoring function<sup>12</sup> rather than advanced free energy calculation to rank the result conformations. To further evaluate the binding affinity for the conformations found through docking, it is common to perform molecular dynamics (MD) simulation and calculate the binding free energy via MMPB/SA<sup>13</sup> method. Mining Minima 2 (M2)<sup>14,15</sup>, which is based on normal mode, is another method that evaluates the binding affinity with high efficiency. It systematically searches for the global conformation minima and calculates the binding affinity based on harmonic approximation. Replica exchange molecular dynamics<sup>16,17</sup> introduces an alchemical parameter to accelerate the exploration of conformation space near the global minimum of bound state, thus yielding high accuracy of free energy evaluation.

In this study, we used docking and MD to investigate the selective binding of MB-10 to the wildtype TIM44 and its four mutants. Accelerated MD starting from the crystal structure was employed to generate the initial conformations for docking. Conventional MD simulation was performed followed by MMPB/SA calculation to evaluate the MB-10 binding affinities in the wildtype proteins and its mutants. A shallow binding site and a deep binding sites were identified, and the deep binding site was confirmed as the MB-10 binding site. The selective binding of MB-10 in the wildtype and resistance of the mutants were explained by the MD conformation analysis and MMPB/SA calculation results.

## 3.2 Methods

### 3.2.1 Overview

To begin, molecular dynamics and docking simulations were performed from the crystal structure of yeast TIM44. In the first round of docking and MD conformation search, an accelerated MD simulation was performed to generate conformations for rigid protein docking. After the MB-10 was docked to the protein, the wildtype enzyme was artificially mutated to obtain the structures of its mutants, i.e. 3H1, 4E5, 3B7 and 3G5. With the docked wildtype and mutants structures, conventional MD simulations were performed on each of the structures followed by MMPB/SA calculations.

In the second round of docking and MD conformation search, conformations from the wildtype conventional MD from the first round were selected for flexible sidechain docking. Similarly, the conformations of mutant complexes were created from the docked

wildtype conformation. From the results of the second round of docking, conventional MD simulations were performed, and the energy was calculated using MMPB/SA. A third round of flexible sidechain docking was conducted in the same approach continuing from the result of the second round.

### 3.2.2 System Preparation

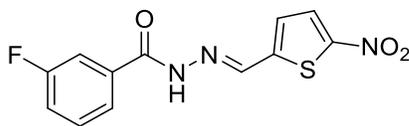


Figure 3.1. Structure of MitoBloCK -10 (MB-10).

The crystal structure of yeast TIM44 was obtained from the Protein Data Bank <sup>18</sup> (PDB ID: 2FXT), and parameterized with AMBER 99SB force field <sup>19</sup>. The structure of MB-10 (Figure 3.1) was manually created by using Vega ZZ <sup>20</sup> and parameterized with GAFF force field <sup>21</sup>. The partial charge of the ligand was calculated by Vcharge <sup>22</sup>. All 3D structures are generated with VMD <sup>23</sup>.

### 3.2.3 Molecular Dynamics

In the initial accelerated MD simulation <sup>24</sup>, the wildtype TIM44 crystal structure was solvated using a box of TIP3P water <sup>25</sup> of at least 12 Å around the protein. This was accomplished by the tleap module in AMBER12 <sup>19</sup>, creating a system of about 43400 atoms. Next, the water molecules was equilibrated at 298K before the entire solvated system was equilibrated at 200K, 250K and 298K. The production runs were performed for 20 ns at 298K, and the temperature was maintained at 298K by Langevin dynamics <sup>26</sup>.

<sup>27</sup>. The particle mesh Ewald (PME) <sup>28</sup> was used for the long-range electrostatic interactions, while the van der Waals interactions were cutoff using the Amber default setting of 8.0 Å. A time step of 2.0 fs was used and MD frames were save every 1 ps. For the two rounds of MD simulations, the protocol from accelerated MD was used except that conventional MD was performed. The low energy docking conformations from accelerated MD, Frame 82, 112, 136 and 138, were used in the first round of conventional MD and the production run was 20 ns. Similarly, the low energy docking conformation from Frame 3059 in the first round of conventional MD was used in the second round of conventional MD and the production run was 70 ns. The last 10 ns of the 70 ns trajectories were used for MMPB/SA calculation.

#### 3.2.4 Docking

In the first round, Frame 82, 86, 93, 104, 110, 112, 116, 125, 136, 138 and 194 were selected from the accelerated MD of wildtype TIM44 for rigid protein docking using Autodock <sup>29</sup>. The Lamarckian genetic algorithm <sup>30</sup> was used to search for minimum energy ligand conformations and orientations. A total of 20 unique docking results were generated and analyzed. Those that had the lowest energy conformation which placed MB-10 in the binding site were used to create mutant structures.

In the second round, a total of 17 conformations from the conventional MD of wildtype TIM44 in the first round were selected including Frame 3059, and same protocol as the first round docking was used except that residue 315, 316, 404, 415 and 417 were kept

flexible during docking. The selected low energy conformation from this round of docking was used to create the mutant structures.

### 3.2.5 Mutations

The mutant structures were created from the selected wildtype docking conformations from the two rounds of searching. The mutated residues atoms were removed and new heavy atoms were added manually in the pdb file to obtain the mutations of 3H1 (T290S), 4E5 (I297), 3B7 (F284L) and 3G5 (V298A). The hydrogen atoms of the mutated residues were added by tleap during MD setting up.

### 3.2.6 MMPB/SA

The two round of conventional MD trajectories (wildtype and four mutants) were used in the energy calculation using the MMPB/SA module <sup>13</sup> of AMBER 12. The module produced three separate trajectories of a protein-ligand, a free protein and a free ligand, from each original trajectory with the explicit water molecules removed. Then the energies were calculated using an implicit Poisson Boltzmann solvent model <sup>31</sup>. The energies of the trajectories were analyzed to evaluate the MB-10 binding to the wildtype TIM44 and the four mutants.

## 3.3 Results and Discussion

To determine where MB-10 bind to the C-terminus of TIM44, we took advantage of computational approaches using docking and molecular dynamics (MD) simulations. In the binding pose found in the first round conformation search, a shallow binding pose

was found on the surface of the space between  $\alpha 4$  helix,  $\alpha 5$  helix, and a loop that connects the  $\beta 3$  and  $\beta 4$  strands (Figure 3.2 and Figure 3.6). This binding site is discussed in Section 3.3.3.

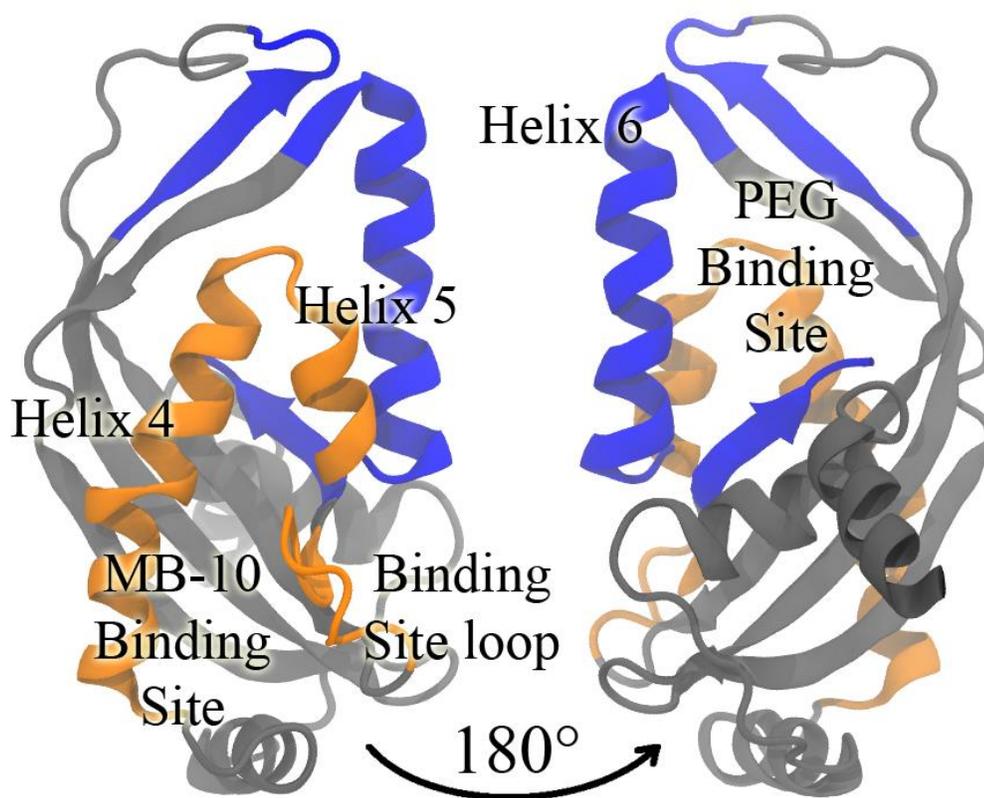


Figure 3.2. The MB-10 binding site in TIM44 wildtype structure, which is located differently from the PEG binding site.

To further explore the buried space below the shallow binding site, the second round docking and MD conformation search was performed. One deep binding pose was discovered from the docking on Frame 3059 in the second round docking, while the other 16 initial protein conformations result in the same shallow binding pose as the first round

of docking. The novel MB-10 binding site discovered by the docking and MD conformation search is shown in Figure 3.2.

### 3.3.1 MB-10 Binding Site

Based on the mutational analysis from experiments, the MD simulations revealed a cavity between  $\alpha$ 4 helix,  $\alpha$ 5 helix, the  $\beta$  strands and a loop that connects the  $\beta$ 3 and  $\beta$ 4 strands, where MB-10 can bind into the pocket (Figure 3.2). Notably, this is a novel binding site that is different from the large hydrophobic binding site for ligand pentaethylene glycol (PEG) on the opposite face of TIM44 (Figure 3.2), which has been suggested to interact with the membrane<sup>32</sup>. During the course of MD simulations, the  $\alpha$ -helices and  $\beta$ -strands were relatively rigid, compared with the loop structures, but the loops were flexible and the size of the MB-10 binding site fluctuated. Molecular docking illustrated that the thiophene ring could successfully fit into the binding cavity formed by I297, Y296 and H292 of the  $\alpha$ 4 helix and other residues in the loop (Figure 3.3). Binding of MB-10 may inhibit important interactions between TIM44 and Hsp70; thus resulting in inhibition of protein import.

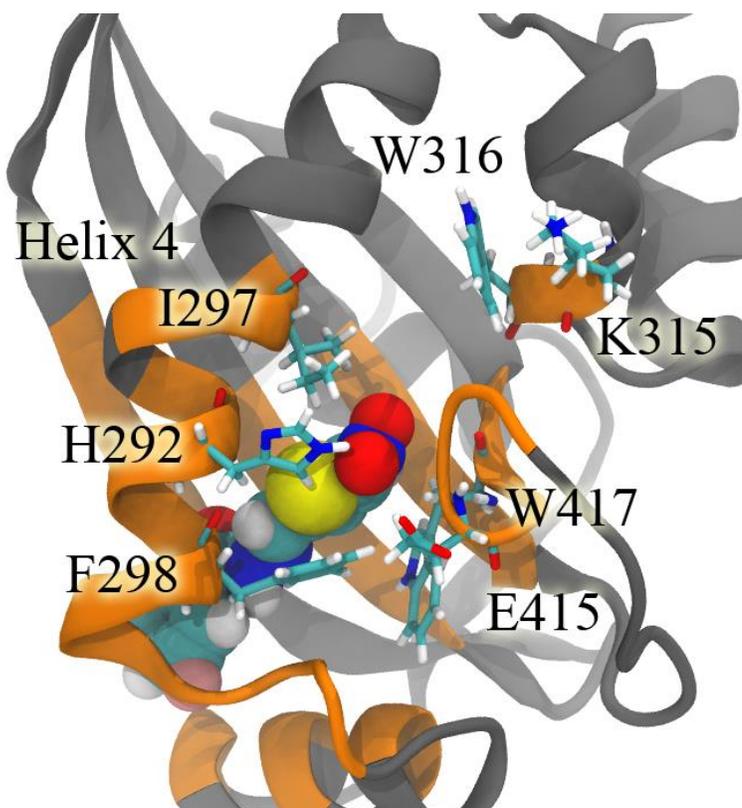


Figure 3.3. The deep binding pose of MB-10 in the wildtype TIM44. The residues within 9 Å of the ligand is rendered in orange. The ligands which blocks the binding site are shown with licorice representation. The ligand is shown in vdW representation.

### 3.3.2 Wildtype vs Mutants

	$\Delta G$	SD	$E_{vDW}$	$E_{Coul}$	$E_{PB}$	$E_{Cavity}$	$E_{Disp}$	$E_{Total}$
WT	-14.89	2.89	-46.44	-2.01	14.92	-28.81	47.45	-14.89
3H1	-9.66	2.94	-42.64	0.47	13.92	-25.24	43.82	-9.66
4E5	-7.39	3.19	-41.84	-8.83	23.74	-26.74	46.28	-7.39
3B7	-10.30	3.02	-43.78	-1.98	16.60	-27.80	46.66	-10.30
3G5	-7.76	3.50	-39.83	-2.78	14.94	-25.72	45.63	-7.76

Table 3.1. The binding energy decomposition from MMPB/SA of MB-10 to wildtype TIM44 and its four mutants. The units are in kcal/mol.

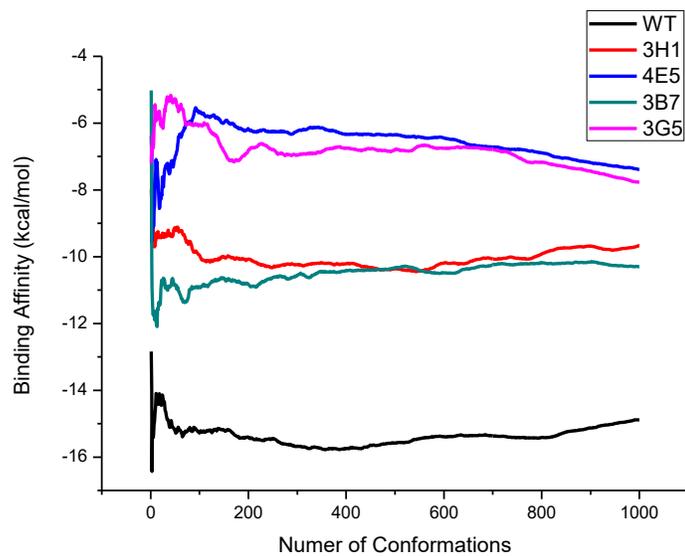


Figure 3.4. The convergence of MMPB/SA calculation. The average accumulative binding affinity over number of conformations was plotted against the number of conformations.

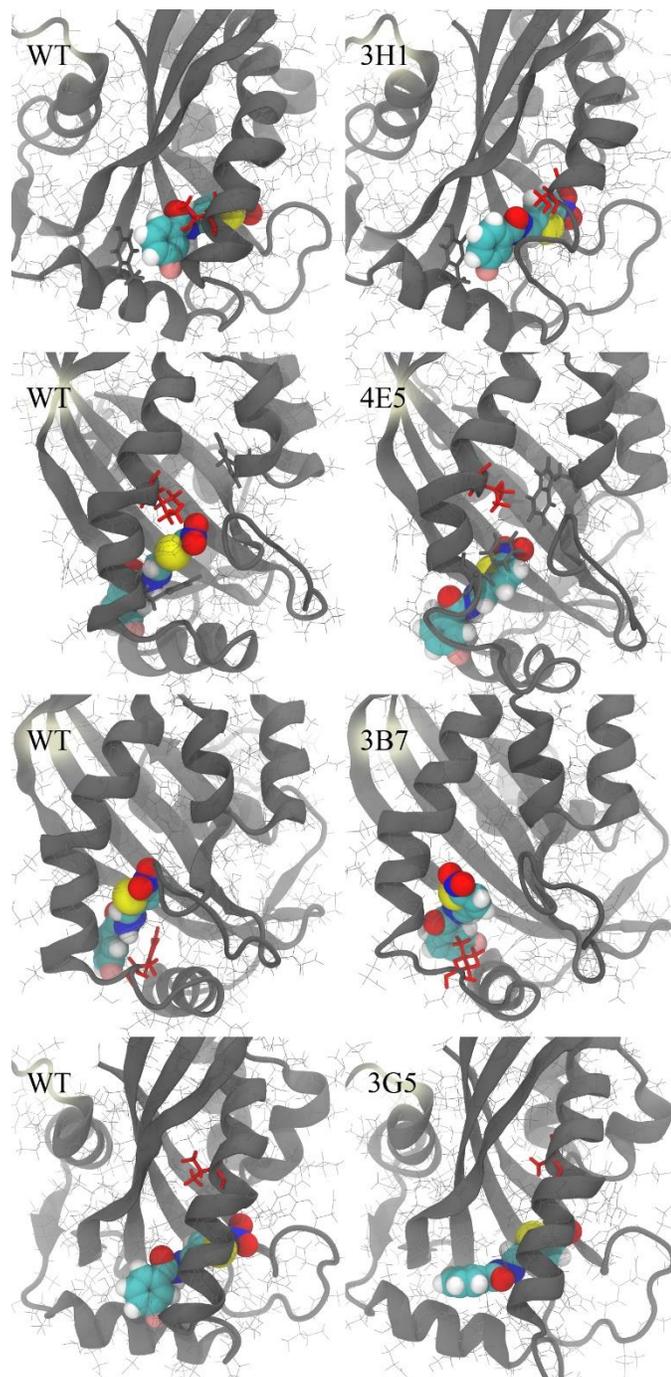


Figure 3.5. The binding pose of the wildtype of TIM44 and the mutants. For comparison, the wildtype is shown for each mutant. The mutation site residues are rendered in red, and other relevant residues are in gray.

According to the mutation study, the wildtype has a stronger MB-10 binding affinity than its four mutants. The binding affinities to 3G5 and 3B7 were weaker compared to the wildtype but stronger than 3H1 and 4E5. The MMPB/SA binding energy decomposition of wildtype TIM44 and its four mutants are listed in Table 3.1 and the convergence of MMPB/SA calculation is shown in Figure 3.4. The MB-10 binding affinity to the wildtype is the strongest of the five listed. Overall, it is roughly 5 kcal/mol stronger than that to the second strongest mutant, 3B7. This agrees with the experimental observation in which only 10  $\mu$ M of MB-10 completely inhibits the activity of the wildtype TIM44 while the four mutants were partially resistant. However, the relative binding affinities to the four mutants reveal little from the experimental result, where 15  $\mu$ M fully inhibits the 3B7 and 3G5 but not 3H1 or 4E5. Nonetheless, considering that the affinities of the four mutants are within the range of 3 kcal/mol and the standard deviation of the affinities are roughly 3 kcal/mol, the mismatch between calculated values and the experimental observations may be partially attributed to the fluctuation in the calculation. Another reason may be due to the total lengths of the MD simulation (70 ns), which is not adequate for complete sampling of the correct binding pose.

The structural differences explain the stronger binding affinity of the wildtype.

Comparing 3H1 to the wildtype, the mutation site, from Thr290 to Ser, which is one less methyl group on the sidechain of the residue, results in the loss of the hydrophobic interaction between the aromatic ring on the ligand and the Ser290 (Figure 3.5).

Therefore the  $\alpha$ 4 helix shifts upwards and the ligand moves towards unbinding, resulting in the roughly 5 kcal/mol binding affinity loss. This is also indicated by the 4 kcal/mol

vdW interaction loss in the 3H1 mutant (Table 3.1). Similar behavior was observed in the 3B7 case (Figure 3.5). The mutation from Phe284 to Leu284 removes the  $\pi$ - $\pi$  interaction between the two aromatic rings, leading the ligand to unbind. In the 3G5 case (Figure 3.5), the mutation from Val298 to Ala, on the other hand, does not contribute to direct interaction loss. Instead, allosteric effect makes the ligand bind weaker. The Val298 has a bulkier hydrophobic sidechain which stabilizes the relative position of the  $\alpha$ 4 helix and holds the cavity open for the ligand. The mutation delocalizes the  $\alpha$ 4 helix and shrinks the binding pocket. For this reason, the ligand unbinds with greater amplitude in the 3G5 mutation. In the 4E5 mutant (Figure 3.5), the mutation from Ile297 to Val 297 also reduces one methyl group from the sidechain. The loss of the methyl group weakens the direct interaction between the sidechain and the ligand. This results in a delocalization of the  $\alpha$ 4 helix making it shift towards the  $\alpha$ 5 helix, which closes up the binding pocket. In this way, the ligand is pushed outwards by the  $\alpha$ 4 helix forcing the ligand to unbind. Overall, the two major reasons for the weaker affinity in the mutants are loss of hydrophobic interaction and delocalization of the  $\alpha$ 4 helix.

### 3.3.3 The Shallow Binding Site

The ligand binds to the shallow binding site in the conformations found in the first round of docking (Figure 3.6). The ligand is in contact with H292, I297, F298, K315, W316, E415 and W417, but no significant hydrophobic interaction or hydrogen bond was observed. The ligand is unable to bind to the deep binding site because the deep binding pocket is closed in the apo-form, which is where the accelerated MD starts.

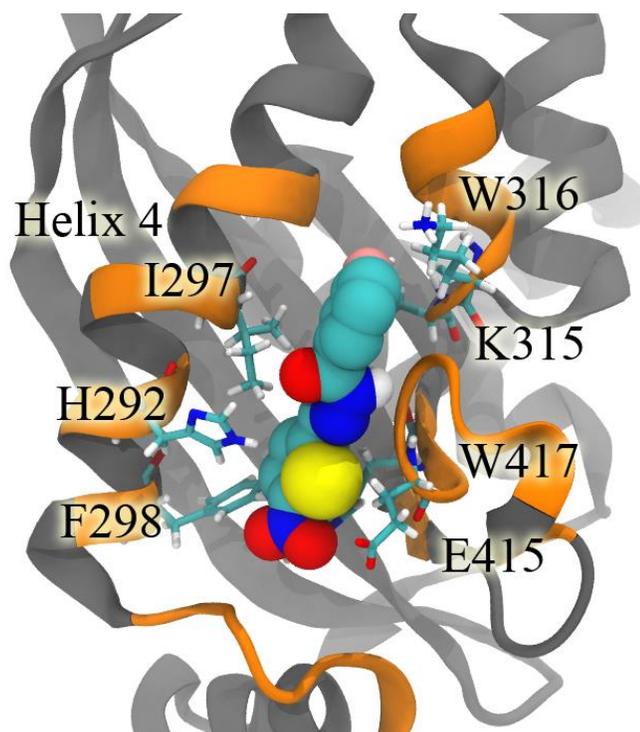


Figure 3.6. The shallow MB-10 binding site. It is on the surface of the deep MB-10 binding.

0 ns to 20 ns	F82	F112	F136	F138
WT	-6.68(3.75)	-5.44(3.49)	-0.78(3.72)	-7.96(3.49)
3H1	-3.92(4.11)	-4.81(3.71)	-2.24(3.88)	-3.86(5.18)
4E5	-7.95(3.53)	-4.75(3.55)	-0.22(4.40)	-7.74(4.51)
3B7	-6.15(3.69)	-9.18(3.36)	-8.23(4.59)	-5.09(4.24)
3G5	-4.48(3.39)	-4.10(3.75)	-7.05(4.98)	-5.76(3.93)
10 ns to 20 ns	F82	F112	F136	F138
WT	-8.07(0.04)	-5.49(3.24)	-0.20(3.56)	-9.05(3.01)
3H1	-5.98(3.71)	-4.04(3.72)	-3.10(2.89)	-6.97(4.58)
4E5	-9.32(2.94)	-4.90(3.39)	-1.99(3.83)	-9.22(3.79)
3B7	-6.99(3.80)	-9.58(3.18)	-10.91(3.85)	-5.32(3.27)
3G5	-4.96(3.15)	-2.95(3.63)	-9.29(4.34)	-7.72(3.07)

Table 3.2. The MMPB/SA results of the simulations starting from the shallow binding site. Frame 82, 112, 136 and 138 in the accelerated MD were used in the first round docking, and the lowest energy conformations were used for conventional MD simulation. MMPB/SA calculations were done on the corresponding trajectories. The units are kcal/mol, and the standard deviations are indicated in the parenthesis.

The MMPB/SA results from the first round of conventional MD simulation are listed in Table 3.2. In general, by looking at the total 20 ns and the last 10 ns of the MMPB/SA results, the ligand adjusts to a better binding pose and forms stronger interaction with the proteins in the last 10 ns simulation. However the relative binding affinities in the wildtype and four mutants remain the same in the first and last 10 ns. Looking at the relative binding affinities of the four initial conformations, no meaningful trend could be observed. 4E5 and 3B7 adopt a better binding pose and stronger affinities than the wildtype in most cases, while Frame 137 led to trivial binding of the ligand to the wildtype. These counter experimental results indicated that the shallow binding pose is not the final bound state of the MB-10.

Despite the failure in finding the correct bound state conformation, the first round docking and MD conformation search opens the way for the second round docking. In the wildtype conventional MD simulation, the ligand induces the opening of the deep binding pocket, which led to the discovery of the deep binding pose.

However this does not imply that the shallow binding pose is the entrance to the deep binding site. First the ligand adopts a binding pose in which it is up-side-down oriented as in the deep binding pose. Second, entrance from the shallow binding site requires a significant conformational shift; the  $\alpha 4$  helix must move away from  $\alpha 5$  helix. Third, from the discussion about deep binding pose in the mutants, it is implied that the ligand unbinds from the open space between  $\alpha 4$  helix,  $\alpha 3$  helix and  $\beta 1$  strand. Therefore the ligand is likely to enter the binding pocket from the opposite direction rather than the shallow binding site.

### 3.3.4 Investigation of Deeper Binding

To eliminate the possibility of other binding pockets, Frame 3919 and 4639 were selected from the second round wildtype conventional MD simulation. The ligand in the simulation was removed and flexible sidechain re-docking of MB-10 was performed on these two conformations. No deeper binding pose was obtained in both initial conformations. This is because the internal space of the TIM44 between the four  $\beta$ -strands,  $\alpha 4$  helix and  $\alpha 5$  helix is filled with the protein sidechains and it is impossible to open up that space without deforming the protein. For this reason, it is safe to conclude that near the deep binding site, no other possible deeper binding pocket exists.

### 3.4 Conclusions

In this study, we used docking and MD simulation to find the binding site of MB-10 in the wildtype TIM44 and explained the selective resistance of the mutants. We concluded that the deep binding site is the correct MB-10 binding site because 1) the mutation study suggested that the MB-10 binding site is near the  $\alpha 4$  helix, 2) the deep binding site is the deepest binding site that can be located near  $\alpha 4$  helix, 3) the MMPB/SA results explain the experimental observed selectivity of MB-10 in the wildtype TIM44, and 4) the structural analysis based on MD simulations explains the resistances in the mutants.

### 3.5 References

1. Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T., and Pfanner, N. (2009) Importing Mitochondrial Proteins: Machineries and Mechanisms. *Cell* **138**, 628-644
2. Stojanovski, D., Bohnert, M., Pfanner, N., and van der Laan, M. (2012) Mechanisms of Protein Sorting in Mitochondria. *Cold Spring Harbor Perspectives in Biology* **4**
3. Neupert, W., and Herrmann, J. M. (2007) Translocation of Proteins into Mitochondria. *Annual Review of Biochemistry* **76**, 723-749
4. Frazier, A. E., Dudek, J., Guiard, B., Voos, W., Li, Y. F., Lind, M., Meisinger, C., Geissler, A., Sickmann, A., Meyer, H. E., Bilanchone, V., Cumsky, M. G., Truscott, K. N., Pfanner, N., and Rehling, P. (2004) Pam16 Has an Essential Role in the Mitochondrial Protein Import Motor. *Nature Structural & Molecular Biology* **11**, 226-233
5. Marom, M., Azem, A., and Mokranjac, D. (2011) Understanding the Molecular Mechanism of Protein Translocation Across the Mitochondrial Inner Membrane: Still a Long Way to Go. *Biochimica Et Biophysica Acta-Biomembranes* **1808**, 990-1001
6. Marom, M., Dayan, D., Demishtein-Zohary, K., Mokranjac, D., Neupert, W., and Azem, A. (2011) Direct Interaction of Mitochondrial Targeting Presequences with Purified Components of the TIM23 Protein Complex. *Journal of Biological Chemistry* **286**, 43809-43815
7. Li, Y. F., Dudek, J., Guiard, B., Pfanner, N., Rehling, P., and Voos, W. (2004) The Presequence Translocase-Associated Protein Import Motor Of Mitochondria -

- Pam16 Functions in an Antagonistic Manner to Pam18. *Journal of Biological Chemistry* **279**, 38047-38054
8. Slutsky-Leiderman, O., Marom, M., Iosefson, O., Levy, R., Maoz, S., and Azem, A. (2007) The Interplay Between Components of the Mitochondrial Protein Translocation Motor Studied Using Purified Components. *Journal of Biological Chemistry* **282**, 33935-33942
  9. Wada, J., and Kanwar, Y. S. (1998) Characterization of Mammalian Translocase of Inner Mitochondrial Membrane (TIM44) Isolated from Diabetic Newborn Mouse Kidney. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 144-149
  10. Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2006) Protein-Ligand Docking: Current Status and Future Challenges. *Proteins-Structure Function and Bioinformatics* **65**, 15-26
  11. Wong, C. F. (2008) Flexible Ligand-Flexible Protein Docking in Protein Kinase Systems. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1784**, 244-251
  12. Huang, S.-Y., Grinter, S. Z., and Zou, X. (2010) Scoring Functions and Their Evaluation Methods for Protein-Ligand Docking: Recent Advances and Future Directions. *Physical Chemistry Chemical Physics* **12**, 12899-12908
  13. Miller, B. R., III, McGee, T. D., Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **8**, 3314-3321
  14. Chang, C. E., and Gilson, M. K. (2004) Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *Journal of the American Chemical Society* **126**, 13156-13164

15. Chang, C. E., and Gilson, M. K. (2003) Tork: Conformational Analysis Method for Molecules and Complexes. *Journal of Computational Chemistry* **24**, 1987-1998
16. Mitsutake, A., and Okamoto, Y. (2000) Replica-Exchange Simulated Tempering Method for Simulations of Frustrated Systems. *Chemical Physics Letters* **332**, 131-138
17. Anselmi, M., and Pisabarro, M. T. (2015) Exploring Multiple Binding Modes Using Confined Replica Exchange Molecular Dynamics. *Journal of Chemical Theory and Computation* **11**, 3906-3918
18. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) Protein Data Bank - Computer-Based Archival File for Macromolecular Structures. *Journal of Molecular Biology* **112**, 535-542
19. Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., York, D. M., and Kollman, P. A. (2015) AMBER 2015. University of California, San Francisco
20. Pedretti, A., Villa, L., and Vistoli, G. (2004) VEGA - An Open Platform to Develop Chemo-Bio-Informatics Applications, Using Plug-In Architecture and Script Programming. *Journal of Computer-Aided Molecular Design* **18**, 167-173

21. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **25**, 1157-1174
22. Gilson, M. K., Gilson, H. S. R., and Potter, M. J. (2003) Fast Assignment of Accurate Partial Atomic Charges: An Electronegativity Equalization Method That Accounts for Alternate Resonance Forms. *Journal of Chemical Information and Computer Sciences* **43**, 1982-1997
23. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics & Modelling* **14**, 33-38
24. Hamelberg, D., Mongan, J., and McCammon, J. A. (2004) Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *Journal of Chemical Physics* **120**, 11919-11929
25. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935
26. Doll, J. D., and Dion, D. R. (1976) Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - Numerical Techniques for Gaussian Generalized Langevin Dynamics. *Journal of Chemical Physics* **65**, 3762-3766
27. Adelman, S. A. (1979) Generalized Langevin Theory for Many-Body Problems in Chemical-Dynamics - General Formulation and the Equivalent Harmonic Chain Representation. *Journal of Chemical Physics* **71**, 4471-4486
28. Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **98**, 10089-10092

29. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry* **30**, 2785-2791
30. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *Journal of Computational Chemistry* **19**, 1639-1662
31. Fogolari, F., Zuccato, P., Esposito, G., and Viglino, P. (1999) Biomolecular Electrostatics with the Linearized Poisson-Boltzmann Equation. *Biophysical Journal* **76**, 1-16
32. Handa, N., Kishishita, S., Morita, S., Akasaka, R., Jin, Z., Chrzas, J., Chen, L., Liu, Z.-J., Wang, B.-C., Sugano, S., Tanaka, A., Terada, T., Shirouzu, M., and Yokoyama, S. (2007) Structure of the Human TIM44 C-Terminal Domain in Complex with Pentaethylene Glycol: Ligand-Bound Form. *Acta Crystallographica Section D-Biological Crystallography* **63**, 1225-1234

## Chapter 4 Accurate Solvent Entropy Estimation: Beyond the Implicit Solvent

### Model

#### 4.1 Introduction

Evaluating free energy accurately is a critical component in fundamental studies and other applications. Studies such binding affinity, crystal morphology prediction and drug discovery relies on the theoretical guidance of the free energy. Enthalpy, which is an important contribution to free energy, can be evaluated through potential energy calculations. On the other hand, entropy has many intrinsic difficulties in its evaluation. It requires more thorough sampling and proper weighting of the states for population evaluation. Also, from a fundamental point of view, it is challenging to define the distinct states.

Various methods<sup>1-3</sup> was developed to evaluate the entropy of solutes in an implicit solvent model. These implicit solvent models (Poisson Boltzmann<sup>4</sup> model and generalized Born<sup>5</sup> model) calculate the solvation energy which includes the solvent enthalpy change and solvent entropy change, with reasonable accuracy especially for small and neutral solutes. However this method is unable to separate the solvent entropy from enthalpy, capture the solvation shell properties and correctly describe the solvation of highly charged species. To overcome these shortcomings, other methods can be applied to simulations using an explicit solvent model. Free energy perturbation<sup>6</sup> and thermodynamic integration<sup>7</sup> can calculate the change in entropy. This requires finite movement of the system from one state to another and intensive sampling of the states

along the transition paths. Umbrella sampling<sup>8</sup>, which is similar to the two aforementioned methods, enforces external restraints on the system along a certain transition pathway. It aggressively samples different states including rare populations, and reproduces an unbiased free energy profile after removing the restraints from the biased populations. These methods are computationally expensive and temporally inefficient. Therefore, efficient solvent entropy calculation method based on explicit water molecules is desirable.

The pair correlation function (PCF) based method<sup>9</sup>, developed by Martin Karplus and Themis Lazaridis is among the many methods developed in the past twenty years. The original method focuses on water molecules and decouples the six dimension PCF into low dimensional functions and thus calculates the entropy of the water. Its descendants, Inhomogeneous Fluid Solvation Theory (IFST)<sup>10, 11</sup> and Grid Inhomogeneous Solvation Theory (GIST)<sup>12</sup> extend its capacity to solutions with solutes and carefully treat the correlation between solutes and solvent molecules. These methods are coupled with WaterMap<sup>13, 14</sup> in later applications<sup>15</sup>. The two phase theory (2PT) approaches the entropy of solvent in a different way, in which the velocity correlation function is used to calculate the density of states and the partition function of the solvent. Due to the generalizability to other solvents other than water and its computational efficiency, it has been widely applied in solvent entropy calculation in many fields. The cell method<sup>16, 17</sup>, a harmonic approximation approach, treats water molecules as if they are confined by forces from its neighbors and vibrates in a local crystal cell. The advantages of this

method are that it can focus on individual water molecules and the implementation is straightforward.

This chapter describes the PCF based methods (Section 4.2), the 2PT method (Section 4.3), and a benchmark evaluation of 2PT method on bulk water (Section 4.4). Evaluations and practical comments are given on these theories. Due to the advantages of the cell method, it is utilized in later studies and will be discussed in Chapter 5.

## 4.2 Entropy from Pair Correlation Function

### 4.2.1 Entropy of Fluids

This method was first discussed by Lazaridis *et al*<sup>9</sup> in 1996. It starts from a fundamental description of the structure of the fluids which makes use of the N-body correlation function  $g^{(N)}$ <sup>18</sup> given by,

$$g^{(N)}(r^N, \omega^N) \quad \text{eq. 4.1}$$

where the  $r$  denotes the distance and  $\omega$  denotes the orientation.

Because this function is impossible to be evaluated completely due to its high dimension, its lower order representations, namely,  $g^{(2)}$ ,  $g^{(3)}$  or several higher order correlation functions, are used to describe the structure of the fluids. Among them, the  $g^{(2)}$ , which is the pair correlation function (PCF), is the most important one because it captures the majority of the information in the fluids.

The entropy  $S$  of a liquid can be expressed as integrals over the multiparticle correlation functions in the equation below,

$$\begin{aligned}
S &= S^{id} - \frac{1}{2} k \frac{\rho}{\Omega^2} \int [g^{(2)} \ln g^{(2)} - g^{(2)} + 1] dr d\omega^2 \\
&\quad - \frac{1}{3!} k \frac{\rho^2}{\Omega^3} \int [g^{(3)} \ln g^{(3)} - g^{(3)} + +3g^{(2)}g^{(2)} - 3g^{(2)}1] dr d\omega^2 - \dots
\end{aligned} \tag{eq. 4.2}$$

where  $S^{id}$  is the entropy of ideal gas,  $k$  is the Boltzmann constant,  $\rho$  is the number density and  $\Omega$  is the integral over the Euler angles of one molecule. The difference between the ideal gas entropy and the fluid entropy is the excess entropy of the fluid.

Due to efficiency and convergence, the excess entropy is only evaluated at the PCF level, which truncates the expression of entropy at the second term. By limiting the number of particles to two, the coordinates of the two particles can be simplified to the distance between these two particles. Because the radial distribution of a liquid is straightforward to evaluate, the PCF is separated into two contributions,

$$g^{(2)}(r, \omega^2) = g^r(r)g(\omega^2|r) \tag{eq. 4.3}$$

where  $g^r(r)$  is the radial distribution function and  $g(\omega^2|r)$  is the orientational correlation function at each distance.

With these concerns, the entropy of the liquid can be decomposed into two contributions: translation and orientation,

$$S_{trans} = -\frac{1}{2} k \rho \int [g^r(r) \ln g^r(r) - g^r(r) + 1] dr \tag{eq. 4.4}$$

$$S_{or} = \rho \int g^r(r) S^{or}(r) dr \tag{eq. 4.5}$$

where

$$S^{or}(r) = -\frac{1}{2}k \frac{1}{\Omega^2} \int g(\omega^2|r) \ln g(\omega^2|r) d\omega^2 \quad \text{eq. 4.6}$$

The theory of this method is straightforward but the practical application of this method experiences several obstacles. First, the  $g(\omega^2|r)$ , which is a fivefold function in the case of water PCF, is still too expensive in terms of memory and requires exponentially increasing number of sampling for convergence. To address this issue, one or two dimensional marginals of the  $g(\omega^2|r)$  are calculated instead of the full fivefold function. Second, the calculation of the  $g(\omega^2|r)$  requires the calculation of relative orientations of each pair of water molecules in the water box, and the construction of a corresponding histogram at each water-water distance. The expensive computational cost hinders the application of this method.

#### 4.2.2 Inhomogeneous Fluid Solvation Theory

Inhomogeneous fluid solvation theory (IFST)<sup>10,11</sup> was proposed two years later since the original work<sup>9</sup> by including the interaction between solvent and solutes. Similar to the original work, IFST treats the excess entropy of the solvent using correlation functions.

The excess entropy of a solvent is given by,

$$S_S^{ex} = S^{VE} + S_{sw} + \Delta S_{ww}^{corr} + \Delta S_{ww}^{lib} \quad \text{eq. 4.7}$$

where the  $S^{VE}$  is the volume entropy originating from the volume excluded by the solute,  $S_{sw}$  is the entropy from the correlation between solute and solvent,  $\Delta S_{ww}^{corr}$  is the entropy of the solvent reorganization from orientational correlations, and  $\Delta S_{ww}^{lib}$  is the solvent reorganization entropy from radial correlations. The four terms are given by,

$$S^{VE} = -k(1 - \rho_w^\circ v_s^\infty) \quad \text{eq. 4.8}$$

$$S_{sw} = -k\rho_w^\circ \int [G_{sw} \ln G_{sw} - G_{sw} + 1] dr \quad \text{eq. 4.9}$$

$$\begin{aligned} \Delta S_{ww}^{corr} &= -\frac{1}{2} k \rho_w^\circ{}^2 \int G_{sw}(r) [G_{sw}(r') \{g_{ww}^{inh} \ln g_{ww}^{inh} - g_{ww}^{inh} + 1\} \\ &\quad - \{g_{ww}^\circ \ln g_{ww}^\circ - g_{ww}^\circ + 1\}] dr dr' \end{aligned} \quad \text{eq. 4.10}$$

$$\Delta S_{ww}^{lib} = -\rho_w^\circ{}^2 \kappa k T \frac{\partial}{\partial \rho} \left[ -\frac{1}{2} k \rho_w \int \{g_{ww}^\circ \ln g_{ww}^\circ - g_{ww}^\circ + 1\} dr \right] \quad \text{eq. 4.11}$$

where  $\rho_w^\circ$  is the density of pure solvent,  $G_{sw}$  is the solute-solvent PCF,  $g_{ww}^\circ$  is the pure solvent PCF,  $g_{ww}^{inh}$  is the inhomogeneous solvent-solvent PCF,  $\kappa$  is the isothermal compressibility,  $v_s^\infty$  is the molar volume of the solute and  $k$  is the Boltzmann constant.

The IFST theory was designed to capture the interaction between solvent and solute and the change in the solvent induced by the presence of the solute. Therefore it has been widely applied in the study of solvation<sup>15,19</sup>, binding site water molecules<sup>13,20,21</sup>, hydrophobicity<sup>22,23</sup>, and water networks<sup>24</sup>.

Several drawbacks were witnessed from the various applications of the IFST method.

The binding free energy of water calculated by IFST is not directly comparable to absolute binding free energy determined by other methods<sup>25</sup>. The IFST does not take into consideration the relaxation of solute molecules after removal of a water molecule<sup>26</sup>.

Lastly this theory relies on readily equilibrated MD simulation, which is not always guaranteed in practice <sup>27</sup>.

A Grid Inhomogeneous Solvation Theory (GIST) <sup>12</sup> was developed by replacing the integrals over the space in the IFST by discrete sums over the voxels in the three-dimensional grid.

#### 4.3 Two-Phase Thermodynamics Model

The two phase thermodynamic model was developed by Lin *et al* <sup>28</sup> in 2003, aiming at efficient solvent entropy calculation. This theory decomposes the liquid phase, which is abstruse to describe with pure diffusive model or harmonic approximation, into the solid phase and the gas phase. Since the solid phase is non-diffusive and has a zero density of state at zero frequency, a harmonic oscillator around the local lattice can describe the solid phase, which resembles the crystal structure of the liquid. On the other hand, the gas phase can be described by had sphere motion and it contribute to all zero frequency density of state. By combining the results from the two phases, the original liquid phase can be restored.

The calculation starts from the determination of the velocity correlation function  $C(\tau)$ ,

$$C(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2\tau} \int_{-\tau}^{\tau} x(t)x(t + \tau)dt \quad \text{eq. 4.12}$$

where the  $x(t)$  is the velocity of an atom at time  $t$ .

Then the velocity correlation is Fourier transformed to obtain the density of state,  $S(\nu)$ , based on the Wiener-Khintchine theorem <sup>29</sup>,

$$S(\nu) = \frac{2}{kT} \lim_{\tau \rightarrow \infty} \int_{-\tau}^{\tau} C(t) e^{-i2\pi\nu t} dt \quad \text{eq. 4.13}$$

where  $k$  is the Boltzmann constant,  $T$  is the temperature,  $\nu$  is the frequency. Thus the total density of state, or  $S(\nu)$ , is obtained and the next step is to decompose it into the gas phase contribution  $S_{Gas}(\nu)$  and the solid phase contribution  $S_{Solid}(\nu)$ .

Because the solid phase makes no contribution to the zero frequency density of state  $S(0)$  which means the diffusion, the  $S(0)$  gives the bound condition for the  $S_{Gas}(\nu)$  and  $S_{Gas}(0) = S(\nu)$ . Considering the velocity correlation function of hard sphere gas decays exponentially <sup>29</sup>, the gas phase density of state can be derived and given by,

$$S_{gas}(\nu) = \frac{S(0)}{1 + \left[ \frac{\pi s_0 \nu}{6fN} \right]^2} \quad \text{eq. 4.14}$$

where  $N$  is the number of atoms, and  $f$  is the fluidicity factor which can be resolved from the molecular dynamics simulation. From the  $S_{gas}(\nu)$ , the  $S_{solid}(\nu)$  can also be solved.

The weighting functions <sup>28, 30</sup> for the entropy from solid component is,

$$W_{solid}^S(\nu) = \frac{\beta h \nu}{\exp(\beta h \nu) - 1} - \ln[1 - \exp(-\beta h \nu)] \quad \text{eq. 4.15}$$

and the weighting functions for the entropy from gas component is,

$$W_{gas}^S(v) = \frac{1}{3} \frac{S_{gas}(v)}{k} \quad \text{eq. 4.16}$$

Finally, the entropy is given by,

$$S = k \ln Q + \beta^{-1} \left( \frac{\partial \ln Q}{\partial T} \right)_{N,V} = k \int_0^\infty dv S(v) W^S(v) \quad \text{eq. 4.17}$$

By summing up the entropies from the gas and solid contributions, the total solvent entropy can be obtained.

The 2PT method requires only roughly 20 ps for converged entropy evaluation<sup>28, 31</sup>, which is more efficient than other similar methods. Due to this reason, it has been widely applied in studies including solvent molar entropy<sup>32</sup>, water molecules in nanotubes<sup>33</sup>, solute solvation free energy<sup>34</sup>, solubility<sup>35, 36</sup>, solvation of DNA<sup>37</sup>, sugar interaction with lipid<sup>38</sup>, entropy of solvent mixture<sup>39</sup> and phase transition of alloys<sup>40</sup>.

However, there are several downsides to this method. The converged molar entropy result from 2PT method has a thermal fluctuation of roughly 0.2 kcal/mol. In a typical small molecule simulation with, for example, 500 water molecules, the entropy uncertainty in the entire water box is 100 kcal, which is 100 kcal/mol for one molar solute. In the presence of solute this uncertainty is overwhelming to the solvent entropy change. To successfully apply this method, the solute-solvent ratio needs to be high enough so that the solvent entropy change due to the presence of the solute is several folds higher than the thermal fluctuation. Another technical issue of this method is that it can only treat the entire water box as a whole part and is unable to calculate the water entropy at specific location, for example, at the binding site of a protein. This stems from the utilization of

the velocity correlation function, which makes it impossible to focus on a region instead of a diffusive water molecule.

#### 4.4 Application of 2PT to Bulk Water

##### 4.4.1 Method

Eight cubic empty water boxes were built with tleap in AMBER 12<sup>41</sup>, with roughly 1k, 2k, 3k, 4k, 5k, 6k, 7k and 8k TIP3P<sup>30</sup> water molecules respectively. Each water box was equilibrated at 30K, 60K, 90K, 120K, 150K, 180K, 210K, 240K, 250K and then 298K for 50 ps at constant volume. Then each water box was equilibrated at 298K at constant pressure for 6 ns and 12 ns respectively in two sets of tests. In the production run, MD simulation was performed for each water box at 298K at constant volume for 20 ps with the coordinates and velocities saved every 1 fs. The temperature of the simulations was maintained by Langevin thermostat<sup>42,43</sup>. The rigid conformations of TIP3P molecules were retained by SHAKE algorithm<sup>44</sup>. All MD simulations were performed with AMBER 12 package. The trajectories were analyzed by 2PT tool<sup>31</sup> to calculate the molar water entropy.

##### 4.4.2 Results and Discussion

The molar water entropies calculated by 2PT method are shown in Figure 4.1 for the total 16 trajectories (8 for 6 ns equilibrium and 8 for 12 ns equilibrium). Overall, the calculated molar water entropy increases with an increase in the number of water molecules. Also it can be either smaller or greater than the experimental value (69.95 J/mol/K), depending on the number of water molecules in the water box. Two

observations can be made by comparing the results from 6 ns equilibrium and 12 ns equilibrium. First, the length of equilibrium results in the fluctuations in the calculated molar water entropy. Second, the increasing trend becomes less significant in the 12 ns equilibrium results. These observations may suggest that the dependence of the calculated molar entropy on the number of water molecules can be alleviated by performing the equilibrium for longer period.

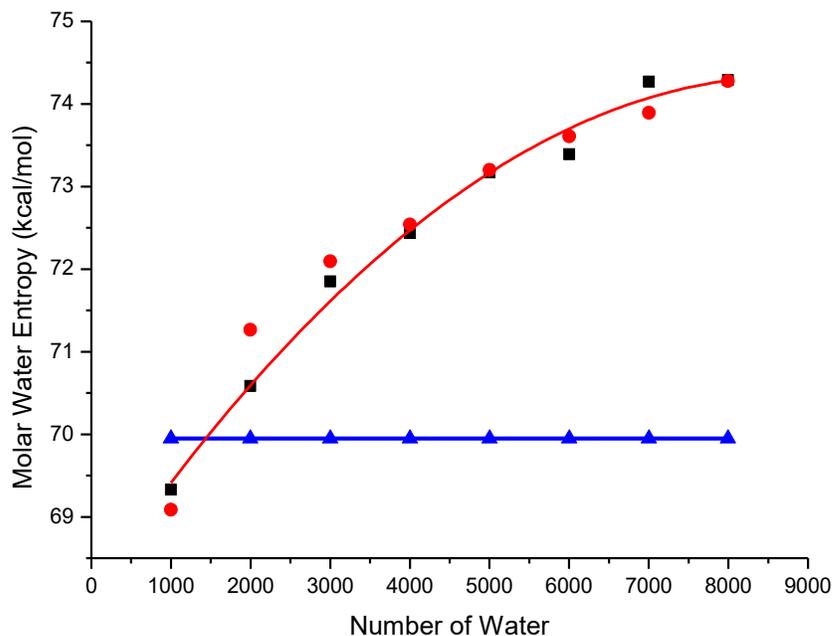


Figure 4.1. The calculated molar water entropy from 2PT plotted against number of water molecules in the water box. The black rectangle indicates the results from 6 ns equilibration, and red dots for the 12 ns equilibration. The blue triangle marks the experimental molar water entropy (69.95 J/mol/K). The second order fit curve for the 2PT results is only shown for the data from 6 ns equilibration.

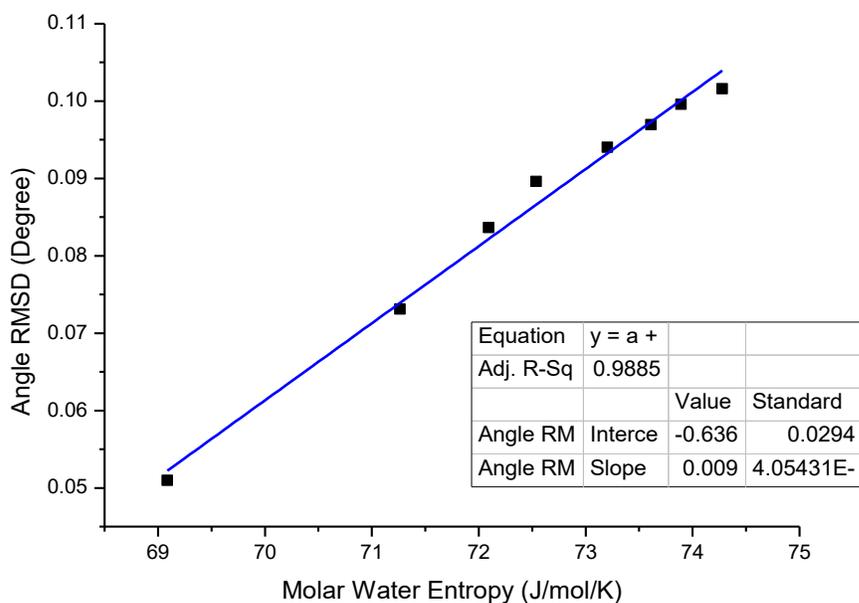


Figure 4.2. The molar water entropy (6 ns) from each trajectory plotted against the H-O-H angle RMSD in each trajectory. Noticeably, the correlation between these two quantities is 0.9885.

To further examine the origin of such behavior, the fluctuation in the H-O-H angles of the TIP3P water molecules was investigated and the results were plotted in Figure 4.2. In the 12 ns equilibrium results, the H-O-H angle RMSD in each of the 8 trajectories correlates with the calculated molar water entropy in a linear fashion. This means the internal vibration of the water molecules contributes to the molar entropy increase. TIP3P is a rigid body water model, in which the H-O-H angle is supposed to be fixed in the MD simulation by the SHAKE algorithm. However the RMSD result of H-O-H angle change indicates that the internal vibration of the water molecules is greater with increased number of water molecules in the water box. Thus the SHAKE algorithm does not constrain the vibration of every single water molecule but restrains it overall. To

conclude, the extra internal vibration is captured by the velocity correlation function, and results in the increase of molar water entropy.

Number of Water	Calc. Molar Entropy (J/mol/K)	Deviation from Expt. Value (J/mol/K)	Deviation of 1000 Water (kcal/mol)
999	69.09	-0.86	-61.50
1998	71.26	1.31	93.62
3000	72.09	2.14	152.66
4003	72.54	2.59	184.32
5001	73.20	3.25	231.66
6003	73.61	3.66	260.67
7003	73.89	3.94	280.73
7997	74.28	4.33	308.26

Table 4.1. Calculated molar entropies from 12 ns equilibrium results. The deviation is from the experimental data (69.95 J/mol/K). The deviation of 1000 water is calculated by multiplying the previous column by 1000.

In addition to the dependence of system size, the fluctuation in the calculation is overwhelming. The calculation is done on empty water boxes for benchmark. However typical calculations are done in a water box with solutes, where the entropy change of the entire water box is equivalent to the molar solvent entropy change of the solute molecule. This means the fluctuation of the change in solvent entropy is induced by the presence of the solute. Mathematically, this is equal to the fluctuation of the calculated molar entropy multiplied by the number of water molecules in that water box. With the 12 ns equilibrium results, the deviations of each trajectory from experimental molar entropy are multiplied by the number of molecules (1000 for fair comparison) (Table 4.1). From the calculation, it is clear that the deviation from experimental molar entropy is magnified by the number of water molecules and is greater than the amplitude of entropy change of

non-covalent binding. Suppose this method was applied to a water box with a small ligand, which only contributed a little portion to the entire system, the entropy change induced by the solute would be overwhelmed by the thermal fluctuation of the solvent and results in no meaningful calculated value.

#### 4.4.3 Summaries

2PT method is capable of capturing the molar entropy of the bulk water. However the SHAKE algorithm implemented in AMBER does not constrain the internal fluctuation of the rigid TIP3P water molecule completely and this results in the system size dependence of the calculated molar entropy from 2PT method. Furthermore, 2PT method is more suitable for application in systems with high portion of water molecule that is affected by the solute rather than dilute systems. Therefore, this method is not further utilized in later applications.

#### 4.5 Conclusions

The original PCF method only treats the bulk water but not solvent. Therefore it is not a suitable method for solvent entropy calculation. Its descendants, IFST and GIST, have several aforementioned drawbacks in Section 4.2, and due to the difficulty in implementation, they are not further utilized in other applications. The 2PT method gave unsatisfying results used together with the AMBER package (Section 4.4). Although a public version of the method is ready for use, it is not further applied in any studies due to its limitations.

Considering the feature of the cell method which can focus on individual water molecules, it allows for entropy calculation of water molecules of great importance. Since implementation is straightforward, this method is further studied and utilized in later works. This method is discussed in detail and applied to host-guest systems in Chapter 5.

#### 4.6 References

1. Miller, B. R., III, McGee, T. D., Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **8**, 3314-3321
2. Ai, R., Fatmi, M. Q., and Chang, C.-e. A. (2010) T-Analyst: A Program for Efficient Analysis of Protein Conformational Changes by Torsion Angles. *Journal of Computer-Aided Molecular Design* **24**, 819-827
3. Chang, C. E., and Gilson, M. K. (2004) Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *Journal of the American Chemical Society* **126**, 13156-13164
4. Fogolari, F., Zuccato, P., Esposito, G., and Viglino, P. (1999) Biomolecular Electrostatics with the Linearized Poisson-Boltzmann Equation. *Biophysical Journal* **76**, 1-16
5. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of the American Chemical Society* **112**, 6127-6129
6. Zwanzig, R. W. (1954) High-Temperature Equation of State by a Perturbation Method .1. Nonpolar Gases. *Journal of Chemical Physics* **22**, 1420-1426
7. Straatsma, T. P., and McCammon, J. A. (1991) Multiconfiguration Thermodynamic Integration. *Journal of Chemical Physics* **95**, 1175-1188
8. Torrie, G. M., and Valleau, J. P. (1977) Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics* **23**, 187-199

9. Lazaridis, T., and Karplus, M. (1996) Orientational Correlations and Entropy in Liquid Water. *Journal of Chemical Physics* **105**, 4294-4316
10. Lazaridis, T. (1998) Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *Journal of Physical Chemistry B* **102**, 3531-3541
11. Lazaridis, T. (1998) Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to Simple Fluids. *Journal of Physical Chemistry B* **102**, 3542-3550
12. Nguyen, C. N., Young, T. K., and Gilson, M. K. (2012) Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit 7 Uril. *Journal of Chemical Physics* **137**
13. Abel, R., Young, T., Farid, R., Berne, B. J., and Friesner, R. A. (2008) Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *Journal of the American Chemical Society* **130**, 2817-2831
14. Young, T., Abel, R., Kim, B., Berne, B. J., and Friesner, R. A. (2007) Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein-Ligand Binding. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 808-813
15. Wang, L., Berne, B. J., and Friesner, R. A. (2011) Ligand Binding to Protein-Binding Pockets with Wet and Dry Regions. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1326-1330
16. Henchman, R. H. (2007) Free Energy of Liquid Water From a Computer Simulation via Cell Theory. *Journal of Chemical Physics* **126**
17. Gerogiokas, G., Calabro, G., Henchman, R. H., Southey, M. W. Y., Law, R. J., and Michel, J. (2014) Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory. *Journal of Chemical Theory and Computation* **10**, 35-48

18. Joslin, C. G., Gray, C. G., and Gubbins, K. E. (1984) *Theory of Molecular Fluids* Oxford
19. Gentilcore, A. N., Michaud-Agrawal, N., Crozier, P. S., Stevens, M. J., and Woolf, T. B. (2010) Examining the Origins of the Hydration Force Between Lipid Bilayers Using All-Atom Simulations. *Journal of Membrane Biology* **235**, 1-15
20. Michel, J., Tirado-Rives, J., and Jorgensen, W. L. (2009) Energetics of Displacing Water Molecules from Protein Binding Sites: Consequences for Ligand Optimization. *Journal of the American Chemical Society* **131**, 15403-15411
21. Huggins, D. J. (2015) Quantifying the Entropy of Binding for Water Molecules in Protein Cavities by Computing Correlations. *Biophysical Journal* **108**, 928-936
22. Lazaridis, T. (2001) Solvent Size Vs Cohesive Energy as the Origin of Hydrophobicity. *Accounts of Chemical Research* **34**, 931-937
23. Snyder, P. W., Mecinovic, J., Moustakas, D. T., Thomas, S. W., III, Harder, M., Mack, E. T., Lockett, M. R., Heroux, A., Sherman, W., and Whitesides, G. M. (2011) Mechanism of the Hydrophobic Effect in the Biomolecular Recognition of Arylsulfonamides by Carbonic Anhydrase. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 17889-17894
24. Breiten, B., Lockett, M. R., Sherman, W., Fujita, S., Al-Sayah, M., Lange, H., Bowers, C. M., Heroux, A., Krilov, G., and Whitesides, G. M. (2013) Water Networks Contribute to Enthalpy/Entropy Compensation in Protein-Ligand Binding. *Journal of the American Chemical Society* **135**, 15579-15584
25. Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997) The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophysical Journal* **72**, 1047-1069

26. Li, Z., and Lazaridis, T. (2007) Water at Biomolecular Binding Interfaces. *Physical Chemistry Chemical Physics* **9**, 573-581
27. Michel, J., Tirado-Rives, J., and Jorgensen, W. L. (2009) Prediction of the Water Content in Protein Binding Sites. *Journal of Physical Chemistry B* **113**, 13337-13346
28. Lin, S. T., Blanco, M., and Goddard, W. A. (2003) The Two-Phase Model for Calculating Thermodynamic Properties of Liquids From Molecular Dynamics: Validation for the Phase Diagram of Lennard-Jones Fluids. *Journal of Chemical Physics* **119**, 11792-11805
29. McQuarrie, D. A. (2000) *Statistical Mechanics*. University Science Books
30. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935
31. Lin, S.-T., Maiti, P. K., and Goddard, W. A., III. (2010) Two-Phase Thermodynamic Model for Efficient and Accurate Absolute Entropy of Water from Molecular Dynamics Simulations. *Journal of Physical Chemistry B* **114**, 8191-8198
32. Pascal, T. A., Lin, S.-T., and Goddard, W. A., III. (2011) Thermodynamics of Liquids: Standard Molar Entropies and Heat Capacities of Common Solvents From 2PT Molecular Dynamics. *Physical Chemistry Chemical Physics* **13**, 169-181
33. Kumar, H., Mukherjee, B., Lin, S.-T., Dasgupta, C., Sood, A. K., and Maiti, P. K. (2011) Thermodynamics of Water Entry in Hydrophobic Channels of Carbon Nanotubes. *Journal of Chemical Physics* **134**

34. Yang, L., and Lin, S.-T. (2015) Rapid Prediction of Solvation Free Energy and Vapor Pressure of Liquid and Solid From Molecular Dynamics Simulation. *Aiche Journal* **61**, 2298-2306
35. Tolonen, L. K., Bergenstrahle-Wohlert, M., Sixta, H., and Wohlert, J. (2015) Solubility of Cellulose in Supercritical Water Studied by Molecular Dynamics Simulations. *Journal of Physical Chemistry B* **119**, 4739-4748
36. Wang, J., Chakraborty, B., and Eapen, J. (2014) Absolute Thermodynamic Properties of Molten Salts Using the Two-Phase Thermodynamic (2PT) Superpositioning Method. *Physical Chemistry Chemical Physics* **16**, 3062-3069
37. Pascal, T. A., Goddard, W. A., III, Maiti, P. K., and Vaidehi, N. (2012) Role of Specific Cations and Water Entropy on the Stability of Branched DNA Motif Structures. *Journal of Physical Chemistry B* **116**, 12159-12167
38. Tian, J., Sethi, A., Swanson, B. I., Goldstein, B., and Gnanakaran, S. (2013) Taste of Sugar at the Membrane: Thermodynamics and Kinetics of the Interaction of a Disaccharide with Lipid Bilayers. *Biophysical Journal* **104**, 622-632
39. Pascal, T. A., and Goddard, W. A., III. (2012) Hydrophobic Segregation, Phase Transitions and the Anomalous Thermodynamics of Water/Methanol Mixtures. *Journal of Physical Chemistry B* **116**, 13905-13912
40. An, Q., Samwer, K., Goddard, W. A., III, Johnson, W. L., Jaramillo-Botero, A., Garret, G., and Demetriou, M. D. (2012) Predicted Optimum Composition for the Glass-Forming Ability of Bulk Amorphous Alloys: Application to Cu-Zr-Al. *Journal of Physical Chemistry Letters* **3**, 3143-3148
41. Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P.,

Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., York, D. M., and Kollman, P. A. (2015) AMBER 2015. University of California, San Francisco

42. Doll, J. D., and Dion, D. R. (1976) Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - Numerical Techniques for Gaussian Generalized Langevin Dynamics. *Journal of Chemical Physics* **65**, 3762-3766
43. Adelman, S. A. (1979) Generalized Langevin Theory for Many-Body Problems in Chemical-Dynamics - General Formulation and the Equivalent Harmonic Chain Representation. *Journal of Chemical Physics* **71**, 4471-4486
44. Forester, T. R., and Smith, W. (1998) SHAKE, Rattle, and Roll: Efficient Constraint Algorithms for Linked Rigid Bodies. *Journal of Computational Chemistry* **19**, 102-111

## **Chapter 5 Calculation of Beta-Cyclodextrin and Guest Binding Enthalpy and Entropy: From Thermodynamics to Kinetics**

### 5.1 Introduction

The study of thermodynamics and kinetics is important in fundamental sciences; thermodynamics focuses on enthalpy, entropy and free energy of a specific state or their changes between two states marking the ends of a process while the kinetics studies detailed transitions of a process. These properties are under the government of several factors, the hydrophobic and hydrophilic interactions, the internal flexibility changes, and the effect from the solvent molecules which is equally important as the previous two but less well understood. In the field of drug discovery, binding affinity has been the most pivotal and dominant guidance because the drug has to win the competition with the natural substrate for the occupancy of the target protein binding site. This was the case until the residence time<sup>1,2</sup> was proposed in the last decade due to stronger correlation between dissociation rate and drug efficacy. A structure-kinetics study on CDK8<sup>3</sup> revealed that some key interactions between the compounds and the protein binding pocket determines the binding affinity and also the residence time but stronger affinity does not necessarily leads to visible residence time. An investigation of hydrogen bond along the binding pathway of xk263 and ritonavir into HIVp showed how hydrogen bonding and desolvation process may effects the association rate<sup>4</sup>. Such related studies asserted the importance of not only thermodynamics but also kinetics.

Enthalpy, which is one major contribution to free energy, is more straightforward to evaluate by potential energy calculation from an MD trajectory. On the other hand, accurate entropy evaluation remains as an open task. A variety of methods, namely perturbation theory<sup>5</sup>, thermodynamic integration<sup>6-9</sup>, umbrella sampling<sup>10</sup>, and partition function from density of states<sup>11</sup>, require intensive sampling and treat the free energy as an entity. Normal mode based configurational entropy calculation<sup>12-14</sup>, successfully tackled the configurational entropy of the solute with less expensive computational cost compared to the methods mentioned above, but solvent entropy, in most cases, the entropy of water, is obscure to approach, due to the intrinsic diffusive property and intractable number of conformations of the solvent. Generalized Born (GB) model<sup>15</sup> and solution to Poisson Boltzmann equation<sup>16-18</sup> give implicit solvation level insights into the free energy of solvent, but they lack an anisotropic and local description of the solvation shell and combine the enthalpy and entropy into the free energy term.

Because of the lack of fundamental understanding of the behavior of water molecule, many efforts have been devoted to this topic<sup>19,20</sup>. Hydrophobicity/hydrophilicity and dewetting process<sup>21,22</sup> were used to explain the role of water molecules from an overall perspective of view while individual water molecules are sometime determinant in thermodynamics and kinetics. Solute interaction with the water molecules and correlation change between the water molecules directly determine the desolvation energy. It also influences the reaction rates by imposing a relaxation time of desolvation or explicit bridge water hydrogen bonding structures which slows down the chemical process.

Theories have been proposed aiming at this topic. From distribution correlation function

method<sup>23</sup>, to inhomogeneous solvation theory (IST),<sup>24,25</sup>, grid inhomogeneous solvation theory (GIST)<sup>26</sup>, two-phase thermodynamics model (2PT)<sup>27,28</sup>, and cell theory<sup>29,30</sup>, many attempts have shed lights on this issue. The original distribution correlation function method was not computationally efficient. The followers including IST and GIST required a frozen conformation of the solutes for entropy calculation purpose. The 2PT method works with flexible solutes, but it treats the entropy of the entire water box and is unable to look at specific water molecules for local entropy changes. The cell theory, on the other hand, treats each water molecule individually so that it is possible to look at local water entropy with flexible solute conformations.

The study of kinetics, however, is much behind thermodynamics. It is an intractable task to experimentally determine the pathway of a binding process with atomistic details so therefore computational approaches became crucial in revealing the binding pathways and providing the corresponding thermal profiles. Molecular dynamics (MD), which evolves the system with time under the government of Newton's law, is a suitable method to investigate the binding events. Although the contemporary computational power is limited at millisecond level<sup>31-34</sup> and can rarely sample the full association/dissociation events, it still provides insights into the binding events<sup>35-39</sup>. Regarding complex protein-ligand systems, Markov State model (MSM)<sup>40</sup> and transition path sampling<sup>41</sup> are methods which successfully unraveled important kinetic information from the massive conformational space.

$\beta$ -Cyclodextrin, a 7-membered cyclic oligosaccharide compound, is a well-studied host molecule in the family of cyclodextrins. It has a hydrophilic outer surface due to the presents of hydroxyl groups while its inner surface is hydrophobic. This property grants many applications to  $\beta$ -Cyclodextrin and its derivatives in fields, such as drug delivery<sup>42-44</sup>, phase transfer catalyst<sup>45-48</sup>, food industry<sup>49-52</sup> and recently biomimetic catalyst<sup>53</sup>. In addition, the cavity enclosed by the glucopyranose units resembles the binding sites of proteins, which makes this small molecule a good model to study binding process of proteins. For those reasons, an immense reservoir of experimental thermodynamics and kinetics data is available for  $\beta$ -Cyclodextrin complexes<sup>7, 54-68</sup>. The relatively small size, which consists of only 147 atoms, ensures the computational efficiency of investigation of the complexation of  $\beta$ -Cyclodextrin using molecular modeling. Due to this reason, many efforts have been spent on various aspects of this specific complex family, including dynamics<sup>69-76</sup>, binding free energy<sup>77-83</sup> and accurate enthalpy evaluation<sup>84</sup>. However these works focus on the structural-thermal relationships but not on kinetics. From the kinetics data it is obvious that for most small ligand- $\beta$ -Cyclodextrin complexes the association ( $k_{on}$ ) and dissociation ( $k_{off}$ ) rates predict an average expected residence time on the scale of several hundred nano seconds, making it feasible to sample the full association and dissociation pathway using graphics processor unit (GPU) accelerated MD simulation.

In this work, we present a study of seven choice  $\beta$ -Cyclodextrin complexes on the thermodynamics profiles, association/dissociation rates, and binding pathways. GPU based MD is employed to sample the behaviors of complex, free  $\beta$ -Cyclodextrin, free

ligand, and empty water box in microsecond-level simulations. Post analysis is conducted on the total binding enthalpy, solute entropy, solvent entropy, rate constants, binding pathways. Enthalpy-entropy compensation and thermal-kinetic relationship are discussed.

## 5.2 Method

### 5.2.1 System Selection and Parameterization

In this work, the ligands for  $\beta$ -cyclodextrin complex were selected under these criteria, 1) availability of experimental data of binding affinity, binding enthalpy, binding entropy, association rate and dissociation rate constants, 2) average bound periods and unbound periods predicted from the association rate and dissociation rate constants are in reasonable time scale, 3) different order of the association rate and dissociation rate constants for kinetics comparison, 4) entropy driven and enthalpy driven binding thermodynamics. Based on these criteria, five ligands, namely 1-butanol, t-butanol, 1-propanol, methyl butyrate, aspirin, 1-naphthylethanol and 2-naphthylethanol, were selected for MD simulations.

The ligand structures were manually created with Vega ZZ<sup>85</sup>.  $\beta$ -cyclodextrin structure was obtained from Cambridge Crystal Data Center (PDB ID: WEWTOJ). GAFF force field<sup>86</sup> was used to parameterize  $\beta$ -cyclodextrin and the ligands. The partial charge of  $\beta$ -cyclodextrin was taken from the reference<sup>77</sup> and the partial charges of the ligands are modeled by 6-31+g(d,p)/B3LYP ChelpG calculation with Gaussian package<sup>87</sup> following an optimization with the same settings. Initial conformations of the complexes are obtained by manually locating the ligand in the center of the cavity of  $\beta$ -cyclodextrin.

A q4md-CD force field <sup>88</sup> was also used to parameterize  $\beta$ -cyclodextrin in a separate setting up. All corresponding calculations were performed in the same way as the set in GAFF. The corresponding results are shown in Section 5.3.3.

## 5.2.2 Molecular Dynamics Simulation

For each of the complex system four MD run series were performed, i.e. on the complex, the free  $\beta$ -cyclodextrin, the free ligand and an empty water box <sup>89</sup>. For each run, the system was first solvated with exactly 1737 water molecules (roughly 12Å away from the solute) and then optimized to eliminate possible clashes. Afterwards the water molecules were equilibrated at 298 K in NPT (Isothermal-isobaric ensemble) for 1 ns. This step was followed by an equilibration of entire system from 200K to 298K in NPT for 150 ps. Then a production run was performed at 298K in NPT for at least 2.5  $\mu$ s for complexes and 1.0  $\mu$ s for runs of single solute. All MD simulations were performed using Amber 14 GPU version <sup>86,90</sup>. The temperature was maintained by Langevin thermostat <sup>91,92</sup>. Trajectories were visualized by VMD <sup>93</sup>. The corresponding bound state conformations in the trajectory of the complexes were collected for the thermodynamics calculation.

## 5.2.3 Enthalpy Calculation

$$\Delta H = \langle E \rangle_{\text{Complex}} + \langle E \rangle_{\text{Water}} - \langle E \rangle_{\beta\text{-cyclodextrin}} - \langle E \rangle_{\text{Ligand}} \quad \text{eq. 5.1}$$

The binding enthalpy was calculated by the formula above <sup>89</sup>, where  $\langle E \rangle$  is the average potential energy of all molecules in the corresponding trajectory. Here the enthalpy is replaced by average potential energy because the change in thermal energy, kinetic

energy and PV term are negligible in the above described conditions. The average potential energy calculation was done by considering the closet periodic image of possible atom pairs to capture the periodicity of the water box. For energetic decomposition, potential energy of the solutes, water, solute-solute interactions, solute-water interactions were also calculated for the trajectories. The four trajectories of each complex were resaved every 10 ps for potential energy evaluation. The standard deviation of mean of the enthalpy was evaluated by using block analysis<sup>94</sup> with 2 ns in a block.

#### 5.2.4 Solute Entropy Calculation

The solute entropy change was decomposed into conformational/vibrational and translational/rotational contributions. The conformational/vibrational term was calculated by using  $S = - \int P(\mathbf{d}) \ln(P(\mathbf{d}))$ , where  $P(\mathbf{d})$  is the probability distribution of conformations defined by key dihedrals in the species. 14 dihedrals defined in Figure 5.1 were selected for  $\beta$ -cyclodextrin while every dihedral except methyl rotations were selected for the ligands. The selected dihedrals were analyzed by fitting Gaussian curves to the corresponding population histogram from MD simulation samplings. The averages ( $\mu$ ) and the deviations ( $3\sigma$ ) were used to define conformations of the corresponding dihedral. Different combinations of dihedral conformations were defined as distinct conformations of the species. The entropy change was then evaluated by using the equation below.

$$\Delta S_{\text{Solute}} = S_{\text{Complex}} - S_{\beta\text{-cyclodextrin}} - S_{\text{Ligand}} \quad \text{eq. 5.2}$$

The translational/rotational entropy of the ligand relative to the host was calculated by aligning the trajectories of the complexes to  $\beta$ -cyclodextrin, and evaluating  $\int \sin(\theta) dx dy dz d\phi d\psi d\theta$  numerically from the MD trajectory. The bin size of the Cartesian position of the ligand was  $2\text{\AA}$  and the bin size of the rotational position of the ligand was  $30^\circ$ .

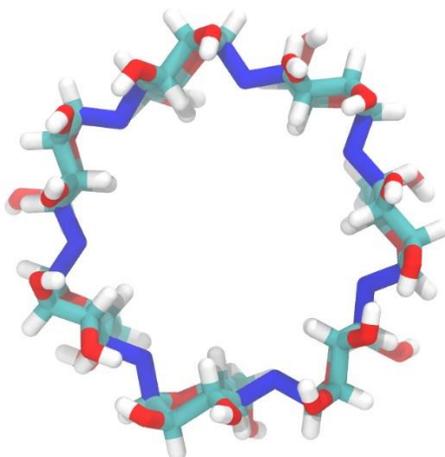


Figure 5.1. The dihedral used to define the macrocyclic ring conformation of  $\beta$ -cyclodextrin (Blue). The  $\beta$ -cyclodextrin is a macrocyclic compound consisting of 7 glucopyranose units. The glucopyranose unit is more rigid than the single bonds linking them so the 14 dihedral rotations were selected to characterize the macrocyclic ring conformation.

### 5.2.5 Solvent entropy Calculation

The solvent entropy was calculated based on the cell method<sup>29, 30, 95</sup>. It treats the water molecules in the water box as if they are vibrating in a local cell and confined by the forces from surrounding water molecules and solutes. It calculates the vibrational entropy from such forces and torques on the three principal axes of the target water molecule. On

the other hand, it treats the conformational entropy of water based on the Pauling model of ice water<sup>96</sup>. Here we proposed a modified version of the conformational entropy based on the previous work<sup>30,95</sup>.

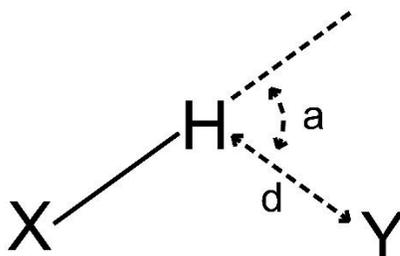


Figure 5.2. The definition of a loose hydrogen bond. It is based on normal hydrogen bond definition except that the complementary angle of X-H...Y is  $80^\circ$  and the distance between H and Y is  $2.65\text{\AA}$ .

First a loose hydrogen bond was defined as shown in Figure 5.2, where the complementary angle of X-H-Y is less than  $80^\circ$  and the H-Y distance is less than  $2.65\text{\AA}$ . The purpose of the loose definition is to capture the forming and breaking hydrogen bonds due to thermal fluctuation. Based on that, the conformational entropy was defined according to Figure 5.3. The water molecules around the target water molecule are classified into three groups, 1) acceptor waters (AW), which are the loose hydrogen bond acceptors from the target water, 2) donor waters (DW), which are the loose hydrogen bond donors to the target water, and 3) surrounding waters (SW), whose oxygen atoms are within  $3.2\text{\AA}$  of the oxygen atom of the target water molecule. Every hydroxyl group on the solute was considered as a DW and every oxygen and nitrogen with no hydrogen on them was considered as an AW. The number of conformations of the target water is

calculated by Eq 5.3. In the first term, the 2 means the two AW sites, regardless of the existence of water molecules there. This term can be interpreted as choosing two sites of the water molecules around as the acceptor sites. The second term means in the rest of the water around to choose # of SW of waters as the SW. The appearance of the last term is to compensate the double counting of water molecules around, because, for example, the target water molecule is the SW when considering the current SW. For example, when there are exactly two AWs and two DWs, which reproduces the Pauling's model, the number of conformation is  $C_4^2 \cdot C_2^0 \cdot \left(\frac{1}{2}\right)^2 = 6 \times \frac{1}{4} = 1.5$ , which is identical to the Pauling's model. When there are two AWs, two DWs and one SW, the number of conformation is  $C_5^2 \cdot C_3^1 \cdot \left(\frac{1}{2}\right)^3 = 6 \times \frac{1}{4} = 3.75$ , which means an additional surrounding water molecule increases the number of conformations of the target water. Finally the conformational entropy of the water is calculated by  $S_{\text{Conf}} = RT \ln(\Omega)$ .

$$\# \text{ of conformation } \Omega = C_{2+\# \text{ of DW}+\# \text{ of SW}}^2 \cdot C_{\# \text{ of DW}+\# \text{ of SW}}^{\# \text{ of SW}} \cdot \left(\frac{1}{2}\right)^{\# \text{ of DW}+\# \text{ of SW}} \quad \text{eq. 5.3}$$

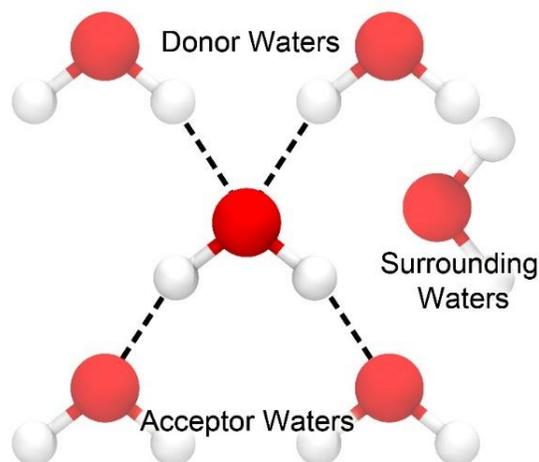


Figure 5.3. The conformational entropy definition of water. It is based on the Pauling's model, where the donor waters (DW) and acceptor waters (AW) are defined with the loose hydrogen bond definition and any water molecule that is within 3.2 Å from the oxygen atom of the target water molecule but cannot form a loose hydrogen bond is considered as a surrounding water (SW).

Technically, a cutoff 50.0 Å was used for force and torque calculation in the vibrational term, and periodically repeated atom images within this cutoff were considered. The complex and  $\beta$ -cyclodextrin trajectories were aligned against the crystal structure of free  $\beta$ -cyclodextrin, and a sub water box which is 20Å  $\times$  26Å  $\times$  26Å around  $\beta$ -cyclodextrin was used for entropy calculation. For ligands trajectories they were aligned against a free ligand conformation, and for empty water box case, a sub water box with size 22Å  $\times$  22Å  $\times$  22Å was used. The above described entropy calculation was performed on every water molecules in the selected sub water box.

#### 5.2.6 Association ( $k_{on}$ ) and Dissociation ( $k_{off}$ ) Rate Calculation

As mentioned above, the system underwent multiple association and dissociation events in the same complex trajectory, and for this reason the association and dissociation rates

were computed by the protocol described here. The system is defined as bound when the center of mass of the ligand is within 7.5 Å from the center of mass of β-cyclodextrin. The bound/unbound state were scanned through the trajectory with the complex, and any bound/unbound period that lasted longer than 1.0 ns were considered as an individual bound/unbound event. Followed by that, the average bound/unbound time length were calculated and plugged into the formula below to compute the  $k_{on}$  and  $k_{off}$ . The solute concentration was directly averaged from the MD trajectory.

$$k_{on} = \frac{1}{\text{unbound time} \times [\text{solute}]} \quad \text{eq. 5.4}$$

$$k_{off} = \frac{1}{\text{bound time}} \quad \text{eq. 5.5}$$

### 5.3 Results and Discussion

The thermodynamics results from MD simulation reproduce the experimental binding free energy ( $\Delta G$ ) (Table 5.1) reasonably well except for 1-butanol and t-butanol. The error is mainly from the enthalpic term rather than the entropic term. The kinetic results from MD simulation generally agree with the experimental data (Table 5.2). The calculated association rate constants ( $k_{on}$ ) reproduce the experimental value, while the calculated dissociation rate constants ( $k_{off}$ ) are systematically greater than the experimental values, especially for the stronger binders whose  $k_{off}$  are one order greater than the experimental value. Because of the  $k_{off}$ , the equilibrium constants ( $K$ ) are systematically smaller than the experimental value except 1-naphthylethanol. The results of enthalpy, entropy, the driven force in binding and water contributions are discussed in Section 5.3.1. The kinetics results are discussed in Section 5.3.2.

	1-Propanol	1-Butanol	Methyl Butyrate	t-Butanol	1-Naphthyl Ethanol	Aspirin	2-Naphthyl Ethanol
$\Delta G$	-1.88	1.10	-2.79	0.33	-3.24	-4.06	-3.03
Calc. $\Delta H$	1.35 (1.17)	3.60 (0.88)	1.45 (0.86)	3.14 (0.76)	-1.41 (0.55)	-0.01 (0.60)	-0.88 (0.54)
$-T\Delta S$	-3.23	-2.51	-4.24	-2.81	-1.83	-4.05	-2.14
$\Delta G$	-0.88	-1.67	-1.99	-2.22	-3.22	-3.74	-3.97
Expt. $\Delta H$	1.43	0.69	/	/	/	/	/
$-T\Delta S$	-2.34	-2.34	/	/	/	/	/

Table 5.1. Calculated and experimental free energies, enthalpies and entropies of binding. The standard deviations of mean of the enthalpy changes are shown in parentheses. All values are in kcal/mol.

	1-Propanol	1-Butanol	Methyl Butyrate	t-Butanol	1-Naphthyl Ethanol	Aspirin	2-Naphthyl Ethanol
$k_{on}$	$1.60 \times 10^8$	$1.12 \times 10^8$	$2.27 \times 10^8$	$2.12 \times 10^8$	$6.21 \times 10^8$	$1.18 \times 10^9$	$6.39 \times 10^8$
Calc. $k_{off}$	$1.96 \times 10^8$	$4.55 \times 10^7$	$9.80 \times 10^7$	$5.44 \times 10^7$	$1.55 \times 10^6$	$6.22 \times 10^6$	$3.65 \times 10^6$
K	0.81	2.46	2.78	3.89	539.53	143.12	174.98
$k_{on}$	$5.10 \times 10^8$	$2.80 \times 10^8$	$3.70 \times 10^8$	$3.60 \times 10^8$	$4.70 \times 10^8$	$7.20 \times 10^8$	$2.90 \times 10^8$
Expt. $k_{off}$	$1.20 \times 10^8$	$3.80 \times 10^7$	$1.28 \times 10^7$	$8.50 \times 10^6$	$4.80 \times 10^5$	$1.30 \times 10^6$	$1.80 \times 10^5$
K	4.20	7.20	29.00	42.60	230.00	553.00	820.00

Table 5.2. Calculated and experimental  $k_{on}$ ,  $k_{off}$  and K. The units of  $k_{on}$  is in  $s^{-1}M^{-1}$ ,  $k_{off}$  in  $s^{-1}$  and K in  $M^{-1}$ .

### 5.3.1 Thermodynamics

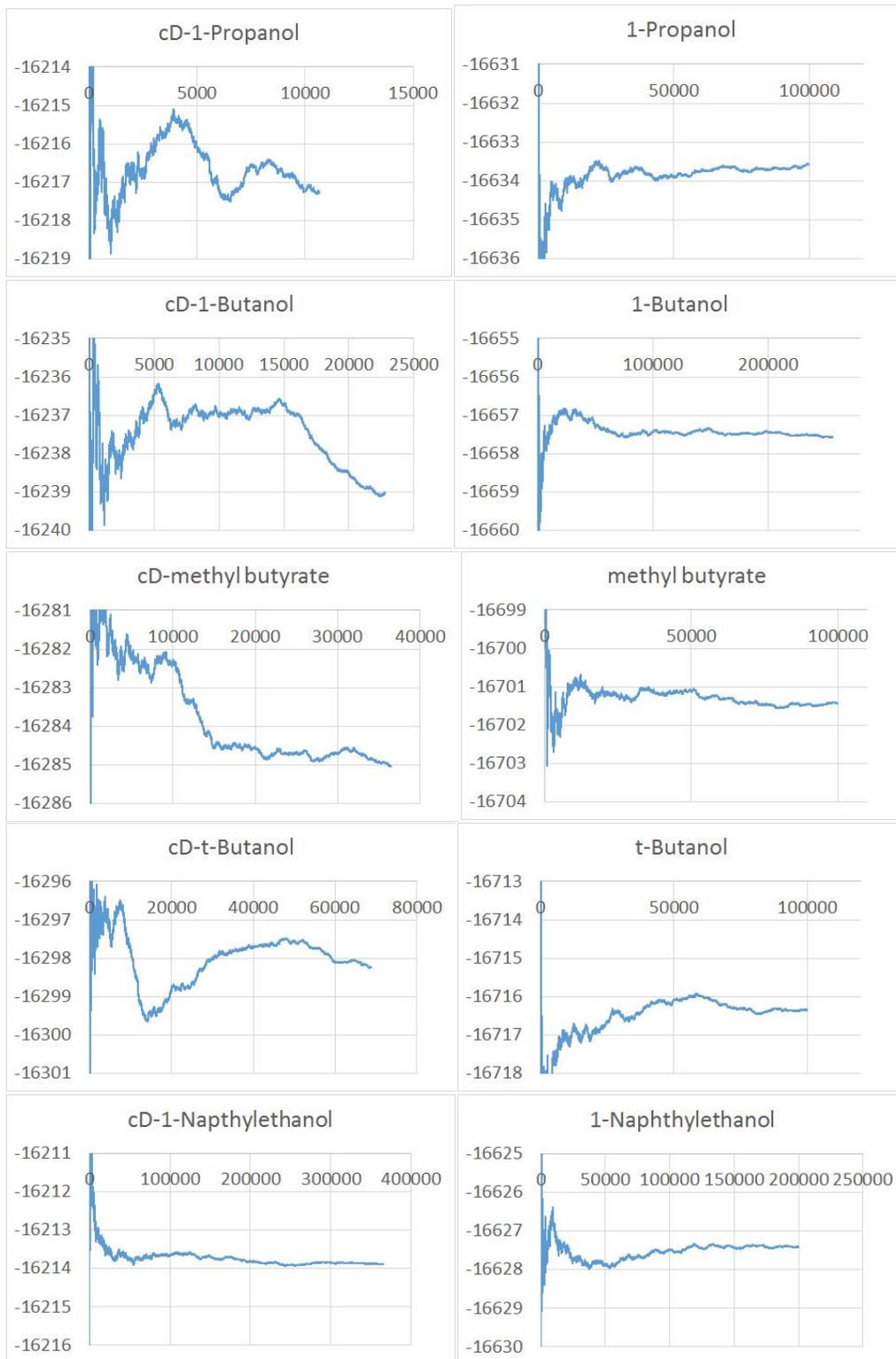
In the first part of Table 5.3, the enthalpy changes of binding process are decomposed into the changes of solute-solute interaction, water-water interaction and solute-water interaction. It can be noticed that the stronger binders including aspirin, 1-naphthylethanol and 2-naphthylethanol have more gains in solute-solute interaction energy. This term is further decomposed in later discussion. Besides the solute-solute interaction, the stronger binders also have a better water-water interaction gain. This is because the stronger binders generally have a larger volume, and in the binding process, more water molecules get freed from the capture of  $\beta$ -cyclodextrin and the ligand, and reform interaction with other water molecules. However, as a compensation, more solute-water interaction disappears in the stronger binder cases. As an entity, the desolvation energy (solute-water interaction loss and water-water interaction gain) of the stronger binders compensates most part of the solute-solute interaction gain and results in a relatively weak total enthalpy gain in the binding process. This total enthalpy change terms are positive for the weak binders and negative for relatively stronger binders in general. This is mainly because for relatively stronger binders, the ligands have direct vdW contacts with the host by occupying the entire space in the cavity of  $\beta$ -cyclodextrin, while the weak binders are small in size so that they do not form stable vdW interactions with the host. Instead the weak ligands tumble with great freedom in the cavity of  $\beta$ -cyclodextrin.

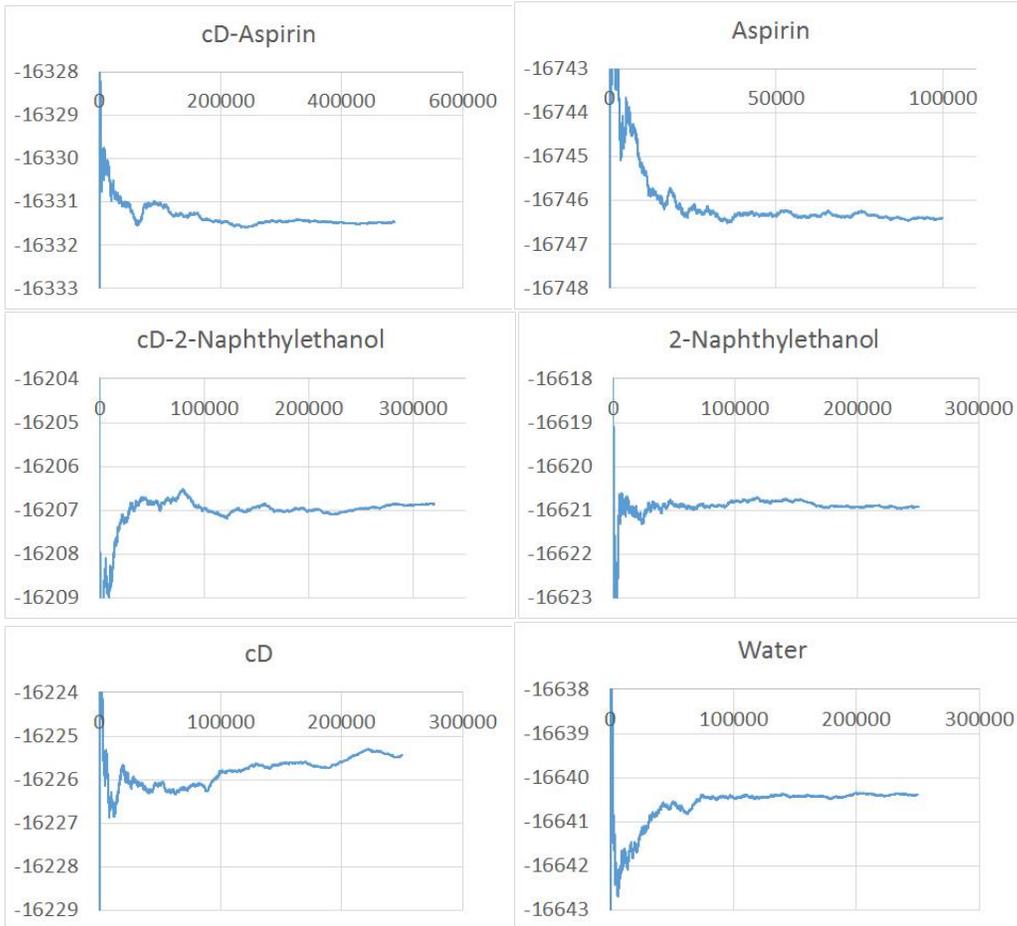
	1-Propanol	1-Butanol	Methyl Butyrate	t-Butanol	1-Naphthyl Ethanol	Aspirin	2-Naphthyl Ethanol
Total Enthalpy Change ( $\Delta H$ )	1.35	3.60	1.45	3.14	-1.41	-0.01	-0.88
Solute-Solute Interaction ( $\Delta H$ )	-11.67	-16.04	-17.55	-16.08	-30.60	-31.53	-28.45
Solute-Water Interaction ( $\Delta H$ )	21.97	30.61	32.88	29.91	50.38	53.75	46.97
Water-Water Interaction ( $\Delta H$ )	-8.95	-10.96	-13.88	-10.68	-21.19	-22.22	-19.40
Desolvation energy ( $\Delta H$ )	13.03	19.65	19.00	19.22	29.19	31.52	27.57
Host Conformational Change ( $\Delta H$ )	-1.10	-2.45	-2.12	-2.70	-4.45	-3.77	-3.97
Guest Conformational Change ( $\Delta H$ )	-0.04	-0.01	0.04	-0.04	0.03	0.23	0.13
Host-Guest Interaction ( $\Delta H$ )	-10.53	-13.59	-15.47	-13.34	-26.18	-27.99	-24.61

Table 5.3. The decompositions of binding enthalpies of the complexes. All values are in kcal/mol.

Regarding the solute-solute interaction, it is further decomposed into the changes of host enthalpy, guest enthalpy and host-guest interaction (Table 5.3). It is not surprising that the ligands have minor conformational changes and result in trivial enthalpy change because first the ligands in this study are either small or relatively rigid and second the binding cavity of  $\beta$ -cyclodextrin is big enough for the ligands to relax as if they are in free environment. On the other hand, the host,  $\beta$ -cyclodextrin, gains enthalpy in the conformational rearrangement. This can be explained by the relaxation of  $\beta$ -cyclodextrin in the binding process. The host is a flexible molecule and it undergoes certain self-adjustment in fully solvated free states but this self-adjustment is not fully necessary due to the elimination of solvation water by the ligand in the bound state. The stronger binders are larger in size and remove more water molecules from the host in the binding process, and thus induce more enthalpy gain in the host. Similarly, the host-guest interaction gain can also be explained by more non-polar interaction formed by more intact surface area between the two molecules.

Figure 5.4. The convergence of the trajectories represented by average potential energy plotted against the number of frames.





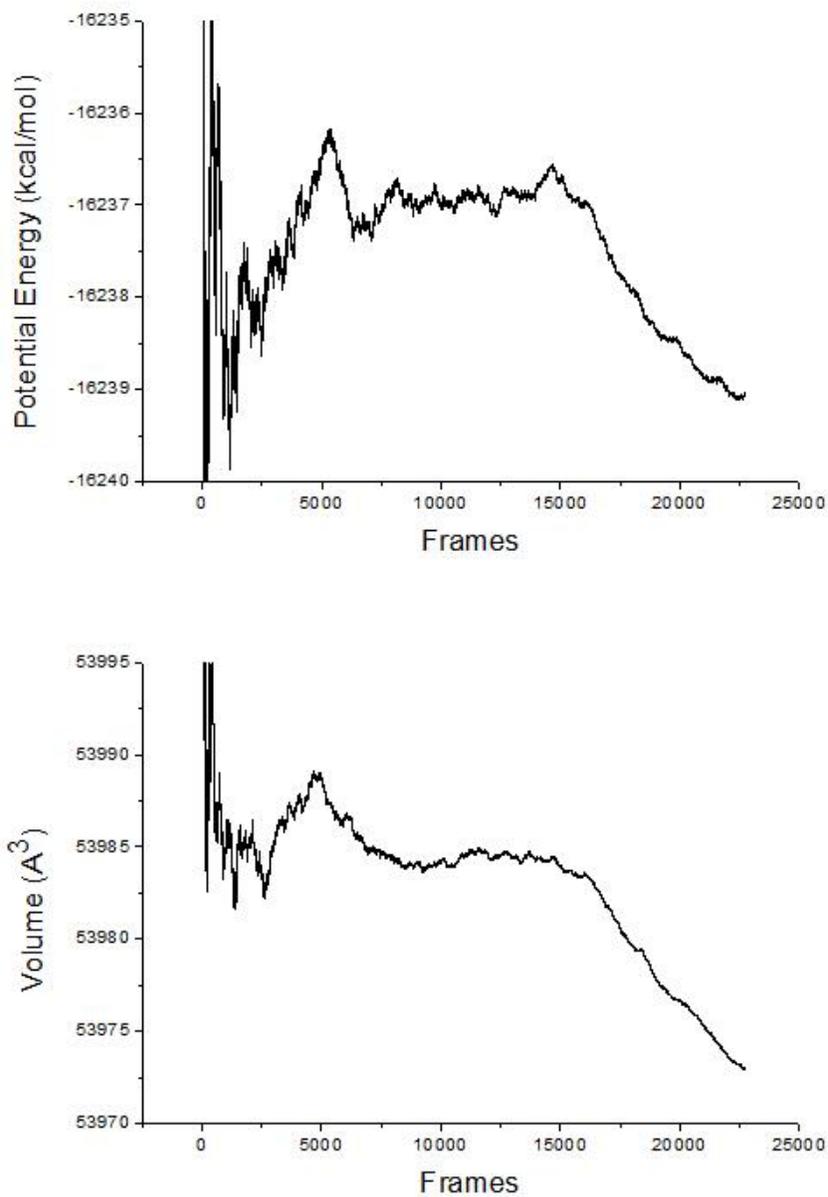


Figure 5.5. The plot of average volume and average potential energy against number of frames in TIP3P solvent model. After convergence, which takes roughly 4000-5000 frame, the average volume and average potential show same trend. Both of them converge at a higher value, and then drop to a lower value.

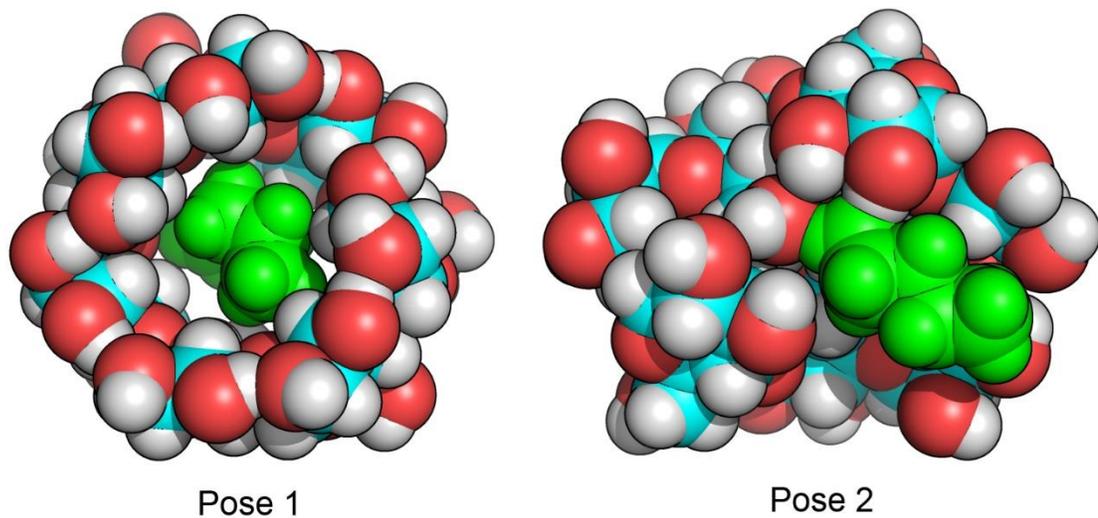


Figure 5.6. The two binding poses of  $\beta$ -cyclodextrin-1-butanol complex. In Pose 1, the ligand stays in the middle of the cavity of  $\beta$ -cyclodextrin and the space between the two molecules is too small for any water molecule to sneak into and thus becomes vacuum. On the other hand in Pose 2, the ligand sticks to the surface of the  $\beta$ -cyclodextrin and the cavity of  $\beta$ -cyclodextrin is closed up by glucopyranose unit flipping.

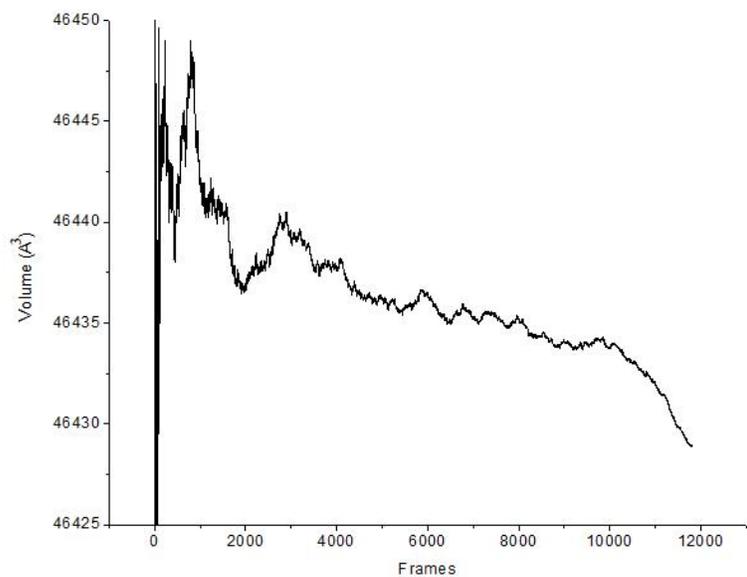
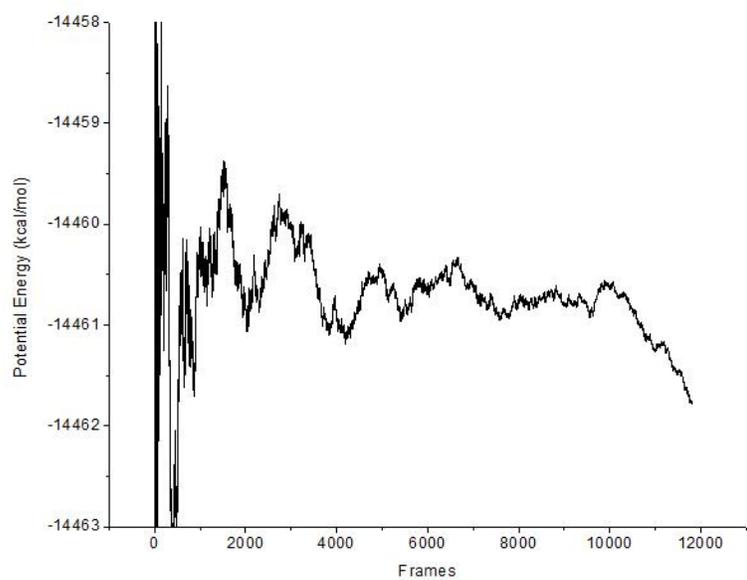


Figure 5.7. The plot of average volume and average potential energy against number of frames of  $\beta$ -cyclodextrin-1-butanol complex in TIP4P solvent model.

The convergence of the enthalpy calculations are examined by plotting the average potential energy against the number of frames (Figure 5.4). A closer look at the plot of  $\beta$ -cyclodextrin-1-butanol complex trajectory (Figure 5.5) reveals some interesting facts regarding the conformations and energy of  $\beta$ -cyclodextrin complexes. The energy converges at a higher value, and then drops to a lower value. By examining the corresponding conformations, it is clear that the higher potential energy corresponds to an “inner bound pose” or Pose 1 of the complex and the lower potential energy corresponds to an “outer bound pose” or Pose 2 (Figure 5.6). Pose 1 has a positive binding enthalpy while Pose 2 has an almost zero binding enthalpy. To further identify the reason of the abnormal enthalpy behavior, the accumulative average volume of the corresponding conformations (Figure 5.5) were examined against the cumulative average potential. It is obvious that the higher potential energy correlates with the greater volume after convergence is reached. The volume change can be attributed to the thermal fluctuation and more importantly the excluded volume of the binding process. In Pose 1, where the ligand sits in the cavity of  $\beta$ -cyclodextrin comfortably, the space between the two molecules cannot be filled with water molecules, and thus more excluded volumes are created. On the other hand, in Pose 2, the space is more efficiently utilized by a denser placing of molecules. To eliminate the possible artifact arising from water model, a similar MD simulation was run with TIP4P water model, and similar behavior of excluded volume was observed (Figure 5.7). Similar excluded volume behaviors were also observed in other complexes (Table 5.4). The excluded volume changes are also

experimentally confirmed, but a tiny fluctuation in the volume change effects the totally enthalpy change significantly due to the fact that interaction energy of each water molecules is as strong as -9.6 kcal/mol and the volume of one water molecules is roughly 30.3 Å<sup>3</sup>. Therefore the calculated excluded volume change subtly changes the totally enthalpy change by 1-2 kcal/mol. A flexible water model or a more accurate force field of the solute may improve the discrepancy between calculated enthalpy and experimental data.

	1-Propanol	1-Butanol	Methyl Butyrate	t-Butanol	1-Naphthyl Ethanol	Aspirin	2-Naphthyl Ethanol
$\Delta G$	-1.88	1.10	-2.79	0.33	-3.24	-4.06	-3.03
Calc. $\Delta H$	1.35 (1.17)	3.60 (0.88)	1.45 (0.86)	3.14 (0.76)	-1.41 (0.55)	-0.01 (0.60)	-0.88 (0.54)
$-T\Delta S$	-3.23	-2.51	-4.24	-2.81	-1.83	-4.05	-2.14
$\Delta G$	-0.88	-1.67	-1.99	-2.22	-3.22	-3.74	-3.97
Expt. $\Delta H$	1.43	0.69	/	/	/	/	/
$-T\Delta S$	-2.34	-2.34	/	/	/	/	/
Excluded Volume	15.63	25.56	12.58	16.89	25.33	23.03	24.72

Table 5.4. The calculated and experimental values of the thermodynamic properties with excluded volumes for each complex. Thermodynamic values are in kcal/mol. Excluded volumes are in  $\text{\AA}^3$ .

The free  $\beta$ -cyclodextrin undergoes glucopyranose unit flipping through out the trajectory. In the majority of the time,  $\beta$ -cyclodextrin adopts the conformations with flipping of two glucopyranose units (Figure 5.8). In relatively shorter periods it adopts conformations with flipping of one glucopyranose unit. Very seldom does  $\beta$ -cyclodextrin visit the crystal like conformations. Upon ligand binding, the dominant population is the crystal structure like conformations with short periods of glucopyranose unit flipping. This is because the cavity of  $\beta$ -cyclodextrin is occupied by the ligand and the ring flipping is hindered. The deviations of the RMSDs along the trajectories (Figure 5.8) were calculated for free  $\beta$ -cyclodextrin (0.268 Å),  $\beta$ -cyclodextrin-2-naphthyl ethanol (0.327 Å) and  $\beta$ -cyclodextrin-t-butanol (0.546 Å). This indicates that the population distribution in the free  $\beta$ -cyclodextrin is narrower than the bound state and the entropy of  $\beta$ -cyclodextrin increases upon ligand binding, and suggests an entropy gain of the host upon binding.

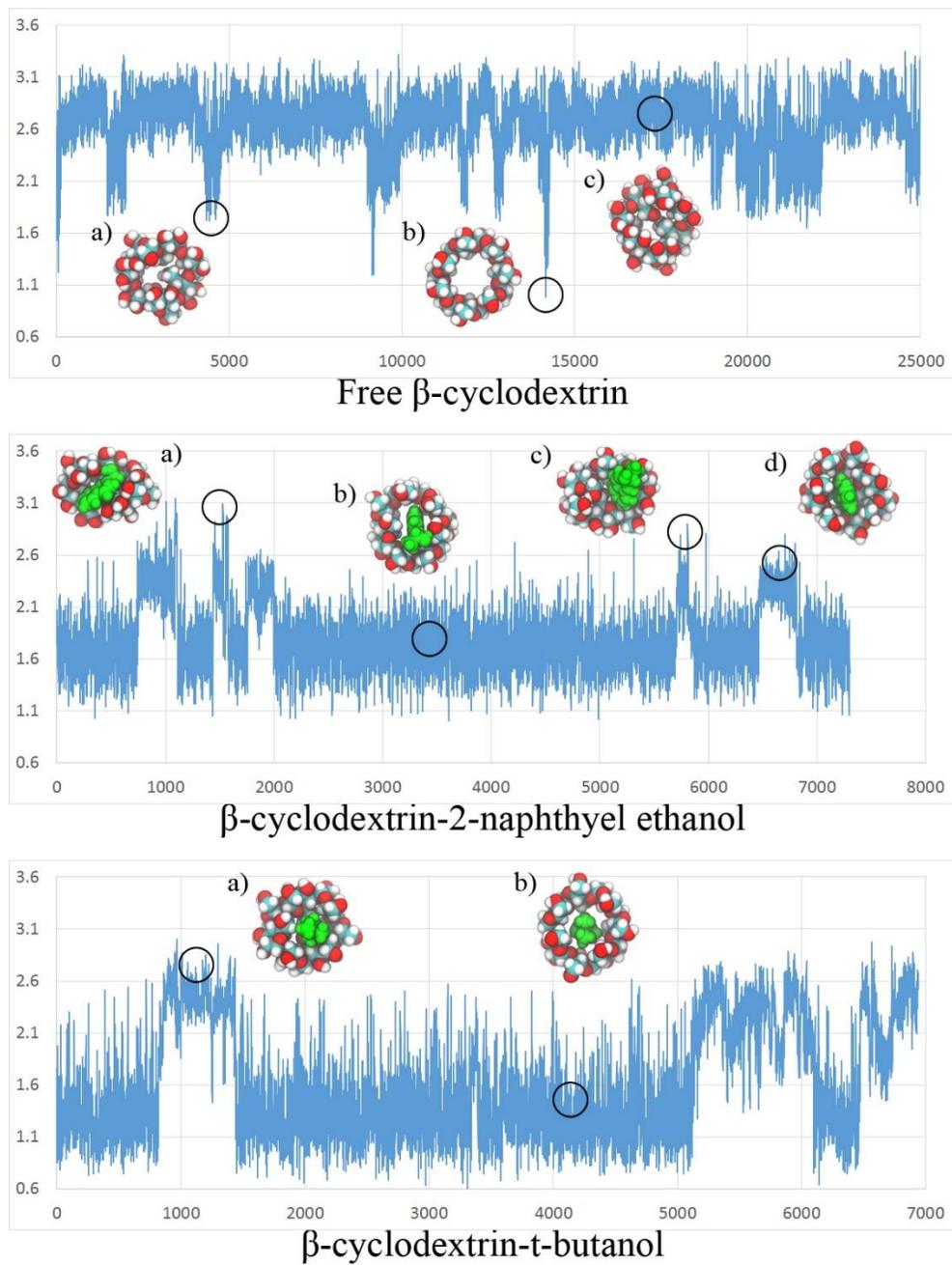


Figure 5.8. The RMSDs in reference with the crystal structure plotted against the frame number. The RMSD is in Å. The representative conformations are shown above the plots and circled on the plots.

	1-Propanol	1-Butanol	Methyl Butyrate	t-Butanol	1-Naphthyl Ethanol	Aspirin	2-Naphthyl Ethanol
Water (-TΔS)	-3.39	-2.37	-3.71	-2.14	-2.19	-3.99	-2.28
cD Internal (-TΔS)	-1.28	-1.46	-1.66	-2.04	-1.40	-1.69	-1.54
Ligand Internal (-TΔS)	-0.01	-0.10	0.01	0.00	0.00	0.08	-0.01
Ligand External (-TΔS)	1.44	1.42	1.12	1.37	1.76	1.55	1.68
Total (-TΔS)	-3.23	-2.51	-4.24	-2.81	-1.83	-4.05	-2.14

Table 5.5. The entropy decompositions of the complexes. All values are in kcal/mol.

The binding entropy decompositions are shown in Table 5.5. The solute entropy change can be decomposed into external entropy change (translational/rotational) and internal entropy change (conformational/vibrational). The external entropy change is captured in this work by numerical integration of the external degrees of freedom of the ligand inside the cavity of  $\beta$ -cyclodextrin under the assumption that the diffusion of  $\beta$ -cyclodextrin is not effected significantly by the presence of the ligand. Normally in host-guest calculations, the external entropy of the ligand is considered as completely lost due to the confinement from the host<sup>13, 52, 97</sup> and this leads to a rough 7 kcal/mol loss of external entropy. However from the numerical integration for  $\beta$ -cyclodextrin-complexes, the loss of external entropy of the ligand upon binding is only roughly 1-2 kcal/mol. This is because the weak interaction between the guest and host is not strong enough to full stop the ligand from tumbling and vibrating in the cavity of  $\beta$ -cyclodextrin. This is actually observed in the MD trajectory, that the ligands do not only stay in the center of the binding site but also tangentially stick to the entrance of binding site on the top or bottom of  $\beta$ -cyclodextrin for some short but observable period up to 100 ns with different poses (Figure 5.6).

The solute entropy change also gives counter intuitive results. From previous work<sup>70, 77</sup>  $\beta$ -cyclodextrin complexes lose entropy upon complexation because of the attraction between the two molecules. However from the conformation/vibrational entropy calculation of  $\beta$ -cyclodextrin, it is obvious that  $\beta$ -cyclodextrin actually gains entropy due to the presence of the ligands. The two major differences are first the presence of the explicit water molecules instead of a continuum of implicit water solvent model and

second the ligands in this work are smaller in size. As discussed above, the water molecules is unable to fit into the space between the guest and host, and the vacuum space allows additional relaxation of the host molecule. Thus, more conformations of  $\beta$ -cyclodextrin can be observed in the bound state than its free state (Figure 5.8) and therefore the internal entropy of  $\beta$ -cyclodextrin increases rather than decreases. On the other hand, the ligands in this study are generally rigid, and they don't undergo much entropy changes in the complexation process. In implicit water model calculations, the vacuum space is not captured due to a continuum description of the solvation and therefore the internal flexibility of  $\beta$ -cyclodextrin is not present due to the absence of hydrophobic environment.

Because of the intrinsic diffusive property of water molecules, it is intractable and meaningless to calculate the average water entropy by focusing on one another water molecules in MD simulations. Therefore a grid cell method<sup>30</sup> approach was employed to partition the water entropy into small spaces near the solute region. The water entropy decompositions of pure water (the reference), free  $\beta$ -cyclodextrin, 1-butanol, and  $\beta$ -cyclodextrin-1-butanol complex are shown in Figure 5.9. In general, the translation entropy decreases at the surface of the solute because the existence of the solute hinders the free diffusion of the water molecules. The rotational entropy either increases on hydrophobic surfaces or decreases near hydrophilic sites. The conformational entropy fluctuates with the density of water molecules.

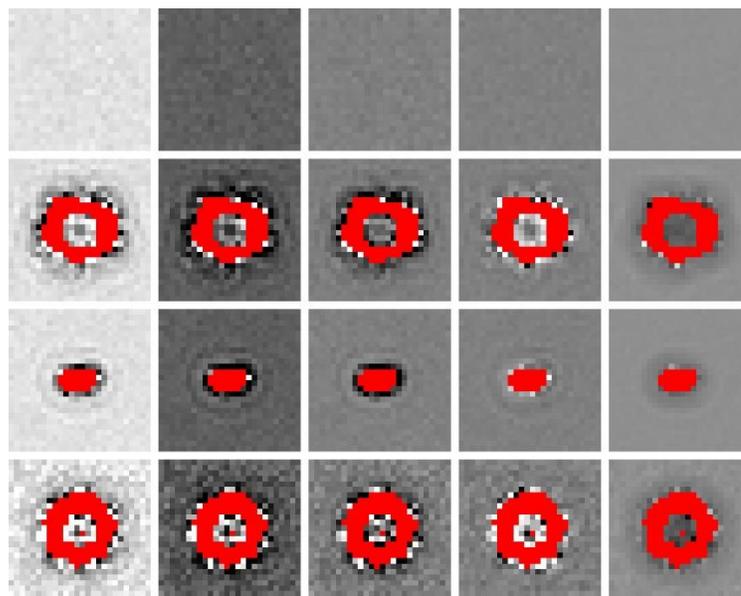


Figure 5.9. The water entropy decompositions near the region of  $\beta$ -cyclodextrin. From top to bottom are pure water, free  $\beta$ -cyclodextrin, free 1-butanol, and  $\beta$ -cyclodextrin-1-butanol complex. From left to right are total entropy, vibrational entropy (translational + vibrational), translational entropy, rotational entropy and conformational entropy. Generally brighter means higher water entropy and vice versa. For each decomposed term the reference is the corresponding pure water entropy. The red areas are regions filled by no water molecules due to the presence of the solute molecules.

For  $\beta$ -cyclodextrin, because the outer surface consists of flexible hydrophilic hydroxyl groups and hydrophobic non-polar carbon rings, it is unable to differentiate the rotational entropy increment or decrement originating from the hydrophilicity or hydrophobicity. However it is clear that the translational entropy decreases significantly due to the hindering of free water diffusion. In the region of the cavity of  $\beta$ -cyclodextrin, the translational water entropy also decreases but not as intuitively losing all translational entropy. This indicates that the water molecules tend to be slowed down instead of totally captured in the cavity. On the other hand the rotational entropy increases on the internal surface of the cavity rather than decrease. This apparently indicates that the internal

cavity of  $\beta$ -cyclodextrin is hydrophobic. Meanwhile in the center of the cavity the rotational water entropy does not show significant change because this area is 2 Å away from  $\beta$ -cyclodextrin. Overall only in the small area in the center 8 Å<sup>3</sup> space the water entropy decreases for roughly 4 J/mol/K and the water entropy in the rest area of the cavity does not change significantly due to the compensation of the translational and rotational entropies. Combining with a slight conformational entropy decrement, the overall water entropy decreases with the presence of  $\beta$ -cyclodextrin.

For 1-butanol, similar behavior of translational water entropy has been observed near the area of the solute. The rotational entropy increase for 2-4 J/mol/K near the carbon chains while it decreases for 3-7 J/mol/K near the hydroxyl group. This clearly corresponds to the facts that hydrophobic surface makes water tumble more and the directional hydrogen bond rigidify the rotation of the water molecules. Similar behavior is also observed in 2-naphthylethanol case Figure 5.10. The naphthyl group is hydrophobic and the water rotational entropy increases for similar amount on its surface. This behavior results in a smaller overall entropy decrease in the case of 1-naphthylethanol and 2-naphthylethanol.

In the complexes, the difference from the free  $\beta$ -cyclodextrin is mainly in the cavity. The water entropy actually increases due to a smaller but more hydrophobic space between the two molecules but because the number of water existing in the space becomes limited, the overall entropy near the region of the complex still decreases although it is not as much as the free  $\beta$ -cyclodextrin.

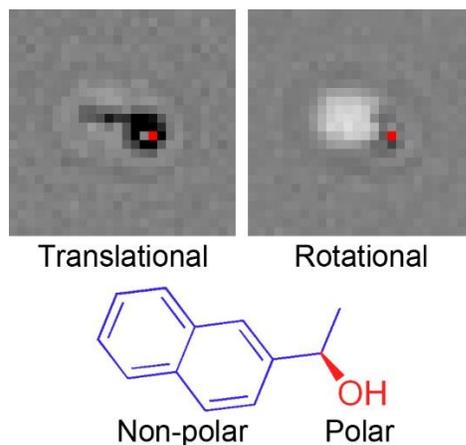


Figure 5.10. The translational and rotational entropies near the region of free 2-naphthylbutyrate. The solute is positioned as the formula below in the two entropy images. The red spots in the images correspond to the hydroxyl group in the molecule.

Combining the effects from free  $\beta$ -cyclodextrin, free ligands, and the complexes, the water entropy changes are summarized in Table 5.1 and Table 5.5. Aspirin, 1-propanol and methyl butyrate have more water entropy gains than the rest ligands. This is because the free aspirin, 1-propanol and methyl butyrate ligands are more hydrophilic and thus they captures nearby water molecules more than the rest ligands. Therefore upon complexation, water molecules gains more entropy by getting released from the free ligands. By comparing 1-butanol and 1-propanol, the only difference comes from the 1 carbon longer carbon chain. Similarly the two naphthylethanols, which also have one hydroxyl group but with a more bulky non-polar group, also have smaller water entropy gains. These all can be explained by the conclusion that the water molecules freed by hydrophilic ligands gain more entropy in the inclusion complexation process with  $\beta$ -cyclodextrin.

To sum up, for the weak binders include 1-propanol, 1-butanol, 1-butanol and methylbutyrate, the desolvation energy dominates the enthalpy change, and therefore the attractions between them and  $\beta$ -cyclodextrin are not strong enough for them to form the complexes for long periods. The driving force of the complexation is actually the entropy change of  $\beta$ -cyclodextrin and the water molecules.  $\beta$ -cyclodextrin gains entropy due to less contact with solvation water and more freedom in internal conformational changes. Water molecules gain entropy due to less restraints from hydrophilic surface exposed to the water molecules. On the other hand for the relatively stronger binders including aspirin, 1-naphthylethanol and 2-naphthylethanol, the enthalpic interactions are stronger than the weak binders, and contributes to the complexation to some degree but still not strong enough for the binding.  $\beta$ -cyclodextrin and water entropy gains also contribute to the complexations to some extent. This kind of behavior is fundamentally different from the rigorous strong binders which totally lose the external entropy and have strong attractions with the host.

In the binding process, the water plays important roles. First it contributes to the repulsive desolvation energy by creating additional excluded volume in the complex. And the water entropy gain from less exposed hydrophilic surface is one of the driving force of the binding of these complexes. The water contribution to the subtle balance between enthalpic repulsion and the entropic attraction also affects the binding equilibrium.

### 5.3.2 Kinetics

From Table 5.2, it is obvious that the  $k_{\text{on}}$  generally matches with the experimental value better than  $k_{\text{off}}$ . This can be explained by the force field.  $k_{\text{on}}$  is determined by the period of free state, which is only governed by the free diffusion and orientation of the two molecules. However  $k_{\text{off}}$  is determined by the period of the bound state, which is governed also by the interactions of the host, guest and the water near binding site. The more sophisticated situation in the bound state needs more accurate description from the force field. Therefore the relatively less accurately described bound state leads to worse evaluation of  $k_{\text{off}}$ . The other reason is from a less complete sampling of the dissociation events. From the period length of bound and free state (Table 5.6), it is safe to assume that longer bound state can be sampled if the MD simulation is further extended. So to obtain a better resolution of the rate constants, longer simulation is necessary.

Table 5.6. The bound and free period lengths of the complexes. All values are in nanosecond.

	1-Propanol		1-Butanol		Methyl butyrate		t-Butanol	
	Bound	Free	Bound	Free	Bound	Free	Bound	Free
1	18.0	58.3	97.7	9.0	52.9	6.8	83.9	6.5
2	1.7	15.2	64.6	90.2	8.5	25.2	1.1	173.9
3	11.3	4.4	1.7	1233.2	29.5	17.2	4.2	848.2
4	7.3	2.4	3.2	326.5	3.4	144.4	1.8	59.6
5	1.5	846.1	1.3	10.5	3.2	139.2	1.8	706.7
6	4.0	310.9	1.4	202.4	1.5	136.4	1.1	208.1
7	2.7	234.9	2.0	423.2	1.1	893.8	3.7	276.3
8	2.6	21.7	3.9	29.1	2.5	7.2	1.8	110.0
9	1.8	179.2			15.9	1.3	192.5	36.6
10	7.5	136.5			10.4	84.4	1.9	41.1
11	1.3	426.8			12.4	14.1	187.1	180.5
12	1.5	202.4			17.2	80.0	1.9	2.6
13					5.2	221.4	4.8	229.3
14					1.0	57.4	1.6	42.3
15					26.6	38.0	22.3	103.6
16					1.8	55.9	1.3	89.2
17					3.0	14.8	2.2	123.6
18					7.6	82.8	3.0	34.7
19					4.5	189.0	4.3	213.2
20					1.2	25.0	1.6	34.0
21					1.6	54.4	2.7	124.4
22					48.4	152.5	4.8	24.4
23					4.1	23.5	37.3	39.8
24					2.4	12.8	1.6	59.0
25					1.8	3.8	1.4	140.5
26					1.8	167.3	1.8	178.7
27					6.0	576.1	1.2	80.0
28							6.7	4.4
29							3.6	173.5
30							1.3	308.3
31							1.2	8.8
32							1.1	249.9
33								
34								
35								
36								
37								
Avg.	5.1	203.2	22.0	290.5	10.2	119.4	18.4	153.5

	1-Naphthylethanol		Aspirin		2-Naphthylethanol	
	Bound	Free	Bound	Free	Bound	Free
1	600.1	25.1	143.1	13.8	794.3	11.6
2	2.2	31.6	37.8	6.7	152.1	22.2
3	65.2	38.3	294.0	35.9	46.4	76.8
4	72.6	49.8	4.1	36.6	106.2	5.3
5	7.2	161.8	21.9	56.4	6.6	6.7
6	2932.5	92.3	102.0	35.7	17.9	13.1
7	1.6	47.4	12.6	30.2	0.0	0.0
8	5.1	84.5	12.3	11.0	2.0	12.4
9	2.7	3.4	467.0	28.7	1.7	11.8
10	5999.4	19.3	27.0	3.7	2.9	14.6
11	6.5	23.5	53.5	44.3	50.0	102.0
12	728.0		2.3	13.9	8.2	52.2
13			149.2	8.7	2.1	15.0
14			1.5	2.1	4.5	37.1
15			8.8	6.3	7.7	88.1
16			84.4	51.5	353.7	92.5
17			1.7	5.0	4.1	17.5
18			3.5	14.4	6.2	125.4
19			10.9	22.9	2.0	50.9
20			1.6	34.8	84.2	37.7
21			3.0	10.5	29.7	341.0
22			164.5	39.1	1.4	164.7
23			3.1	57.3	203.7	12.1
24			3.2	19.8	83.8	7.2
25			8.2	24.7	52.0	5.4
26			4.0	85.2	120.1	34.5
27			4.3	39.1	1878.1	6.2
28			1.2	14.2	2149.9	14.1
29			3.9	119.8	13.5	13.2
30			5.4	16.8	33.1	44.7
31			3.7	7.7	2.1	25.5
32			26.9	25.1	956.3	90.6
33			1.5	7.0	94.1	131.7
34			2.1	45.3	2046.1	
35			396.0	17.4		
36			15.7	5.4		
37			2417.4			
Avg.	868.6	52.5	121.7	27.7	274.0	51.0

The  $k_{\text{on}}$  of all ligands in this study are on the order of  $10^8$  L/mol/s, which are just one order slower than the diffusion controlled binding. This can be supported by the fact that the binding site of  $\beta$ -cyclodextrin is widely exposed except for the blocking from the glucopyranose units on the macrocyclic ring. Two or three of the glucopyranose units on  $\beta$ -cyclodextrin may flip to occupy the cavity of  $\beta$ -cyclodextrin in the free state while in the bound state the rings flip back to the pose when they are perpendicular to the macrocyclic ring. This also indicates that along the binding reaction coordinate, the free energy barrier is not too high. Combining with the fact that there is no enthalpic barrier along the association/dissociation, it is reasonable to assume the energy barrier is some small entropic barrier. Such entropic barrier should consists of a contribution from  $\beta$ -cyclodextrin entropy increment, a contribution from ligand entropic loss and another contribution from water molecules that is involved with  $\beta$ -cyclodextrin conformational changes.

Comparing the weak binder and the relatively stronger binders, the binding paths show slight differences. For the weak binder case, the only possibility of binding pathway is the one in which the ligand randomly hits into the cavity of  $\beta$ -cyclodextrin and gets trapped by the entropic well. After some rearrangement of the conformations, the ligand slides into the cavity and opens up the glucopyranose units. For the stronger binders, there are more possibilities. First they can hit into the cavity and bind simply as the behavior as the weak binders. Second they can stick to the outer surface of  $\beta$ -cyclodextrin due to stronger enthalpic attractions and then slides on the surface of  $\beta$ -cyclodextrin until they find the cavity and opens up the glucopyranose units. Thirdly they can even hit into the cavity of

$\beta$ -cyclodextrin when the glucopyranose unit flips upside down, and make that glucopyranose unit stuck in that inverted conformation. In the third case the binding event usually is shorter than binding with a regular conformation of  $\beta$ -cyclodextrin.

### 5.3.3 Results from q4md-CD Force Field

The thermodynamic results are summarized in Table 5.7, Table 5.8 and Table 5.9. The binding affinities of the complexes are systematically overestimated for 1-2 kcal/mol by the q4md-CD force field. From the results of 1-propanol and 1-butanol, the enthalpy terms are overestimated by roughly 1.5 kcal/mol. On the other hand, the q4md-CD force field underestimates the contribution from the entropy for the three ligands with experimental entropy data. This is related to the rigidity of  $\beta$ -cyclodextrin in the q4md-CD force field.

The kinetic results are shown in Table 5.10. The q4md-CD force field generally overestimates the  $k_{\text{on}}$  of all complexes by one order. However the  $k_{\text{off}}$ , which is governed by the interaction of host and guest and underestimated by the GAFF force field, matches with experimental values reasonably well. This is because the cavity of  $\beta$ -cyclodextrin is more open in the more rigid description and allows faster entrance of the ligands, but the ligand can stay in the binding site stably due to the overestimated attractions.

	1-Propanol	1-Butanol	Methyl Butyrate	t-Butanol	1-Naphthyl Ethanol	Aspirin	2-Naphthyl Ethanol
Calc.	$\Delta G$ -2.13 -0.9 (0.78)	-1.45 -1.07 (0.59)	-4.24 -2.26 (0.65)	-2.54 -3.13 (0.65)	-5.43 -5.69 (0.57)	-6.12 -4.55 (0.64)	-4.30 -4.33 (0.56)
	$\Delta H$ -1.23	-0.39	-1.98	0.59	0.26	-1.57	0.03
	$-T\Delta S$ $\Delta G$	-1.67	-1.99	-2.22	-3.22	-3.74	-3.97
Expt.	$\Delta H$ 1.43	0.69	/	/	/	/	/
	$-T\Delta S$ -2.34	-2.34	/	/	/	/	/

Table 5.7. Calculated and experimental free energies, enthalpies and entropies of binding in q4md-CD force field. The standard deviations of mean of the enthalpy changes are shown in parentheses. All values are in kcal/mol.

	1-Propanol	1-Butanol	Methyl Butyrate	t-Butanol	1-Naphthyl Ethanol	Aspirin	2-Naphthyl Ethanol
Total Enthalpy Change ( $\Delta H$ )	-0.90	-1.07	-2.26	-3.13	-5.69	-4.55	-4.33
Solute-Solute Interaction ( $\Delta H$ )	-13.23	-16.28	-20.23	-17.00	-28.19	-29.47	-27.03
Solute-Water Interaction ( $\Delta H$ )	21.01	26.46	32.70	23.72	43.71	46.62	42.31
Water-Water Interaction ( $\Delta H$ )	-8.68	-11.25	-14.73	-9.84	-21.21	-21.70	-19.60
Desolvation energy ( $\Delta H$ )	12.33	15.21	17.97	13.87	22.50	24.92	22.71
Host Conformational Change ( $\Delta H$ )	-1.67	-2.27	-3.15	-3.13	-3.05	-3.01	-3.00
Guest Conformational Change ( $\Delta H$ )	0.04	0.09	0.18	-0.05	0.20	0.56	0.17
Host-Guest Interaction ( $\Delta H$ )	-11.60	-14.10	-17.26	-13.82	-25.34	-27.03	-24.20

Table 5.8. The decompositions of binding enthalpies of the complexes in q4md-CD force field. All values are in kcal/mol.

	1-Propanol	1-Butanol	Methyl Butyrate	t-Butanol	1-Naphthyl Ethanol	Aspirin	2-Naphthyl Ethanol
Water (-TΔS)	-3.27	-2.20	-3.97	-1.71	-2.06	-3.56	-2.11
cD Internal (-TΔS)	0.44	0.24	0.29	0.54	0.36	0.23	0.15
Ligand Internal (-TΔS)	0.03	-0.17	-0.14	0.00	0.00	-0.06	0.01
Ligand External (-TΔS)	1.57	1.74	1.84	1.76	1.95	1.82	1.99
Total (-TΔS)	-1.23	-0.39	-1.98	0.59	0.26	-1.57	0.03

Table 5.9. The entropy decompositions of the complexes in q4md-CD force field. All values are in kcal/mol.

	1-Propanol	1-Butanol	Methyl Butyrate	t-Butanol	1-Naphthyl Ethanol	Aspirin	2-Naphthyl Ethanol
$k_{on}$	$1.02 \times 10^9$	$1.34 \times 10^9$	$1.51 \times 10^9$	$1.11 \times 10^9$	$5.64 \times 10^8$	$3.35 \times 10^9$	N/A
Calc. $k_{off}$	$1.11 \times 10^8$	$2.80 \times 10^7$	$7.88 \times 10^6$	$5.51 \times 10^6$	$1.26 \times 10^6$	$3.29 \times 10^6$	N/A
K	9.15	47.75	191.91	202.29	448.01	1017.02	N/A
$k_{on}$	$5.10 \times 10^8$	$2.80 \times 10^8$	$3.70 \times 10^8$	$3.60 \times 10^8$	$4.70 \times 10^8$	$7.20 \times 10^8$	$2.90 \times 10^8$
Expt. $k_{off}$	$1.20 \times 10^8$	$3.80 \times 10^7$	$1.28 \times 10^7$	$8.50 \times 10^6$	$4.80 \times 10^5$	$1.30 \times 10^6$	$1.80 \times 10^5$
K	4.20	7.20	29.00	42.60	230.00	553.00	820.00

Table 5.10. Calculated and experimental  $k_{on}$ ,  $k_{off}$  and K from q4md-CD force field. The units of  $k_{on}$  is in  $s^{-1}M^{-1}$ ,  $k_{off}$  in  $s^{-1}$  and K in  $M^{-1}$ .  $\beta$ -cyclodextrin-2-naphthyl ethanol complex was in bound state through the trajectory and therefore no kinetic properties were obtained.

The flexibility change of  $\beta$ -cyclodextrin in free state, complexes with 2-naphthylethanol and t-butanol are shown in Figure 5.11. Compared with the plot for GAFF force field, the free  $\beta$ -cyclodextrin is more rigid and prefers to stay in the crystal like structure rather than glucopyranose flipping conformations. Similar to GAFF, complexation makes the population of  $\beta$ -cyclodextrin shift to more crystal like structures. Unlike the GAFF, the standard deviations of the RMSD plot reveal that the flexibility of  $\beta$ -cyclodextrin decreases from free state (0.481 Å) to complex with 2-naphthylethanol (0.177 Å) and complex with t-butanol (0.178 Å). And this corresponds to thermodynamic results, in which the  $\beta$ -cyclodextrin loses entropy upon complexation with the ligands.

In sum, the  $\beta$ -cyclodextrin in q4md-CD force field is more rigid and behaves more like the crystal structure than that in GAFF. The binding rate is overestimated due to the more open cavity and stronger attraction between the host and guest while the unbinding is comparable to experimental data. The binding enthalpy is overestimated but the binding entropy is underestimated due to the rigid description of the  $\beta$ -cyclodextrin.

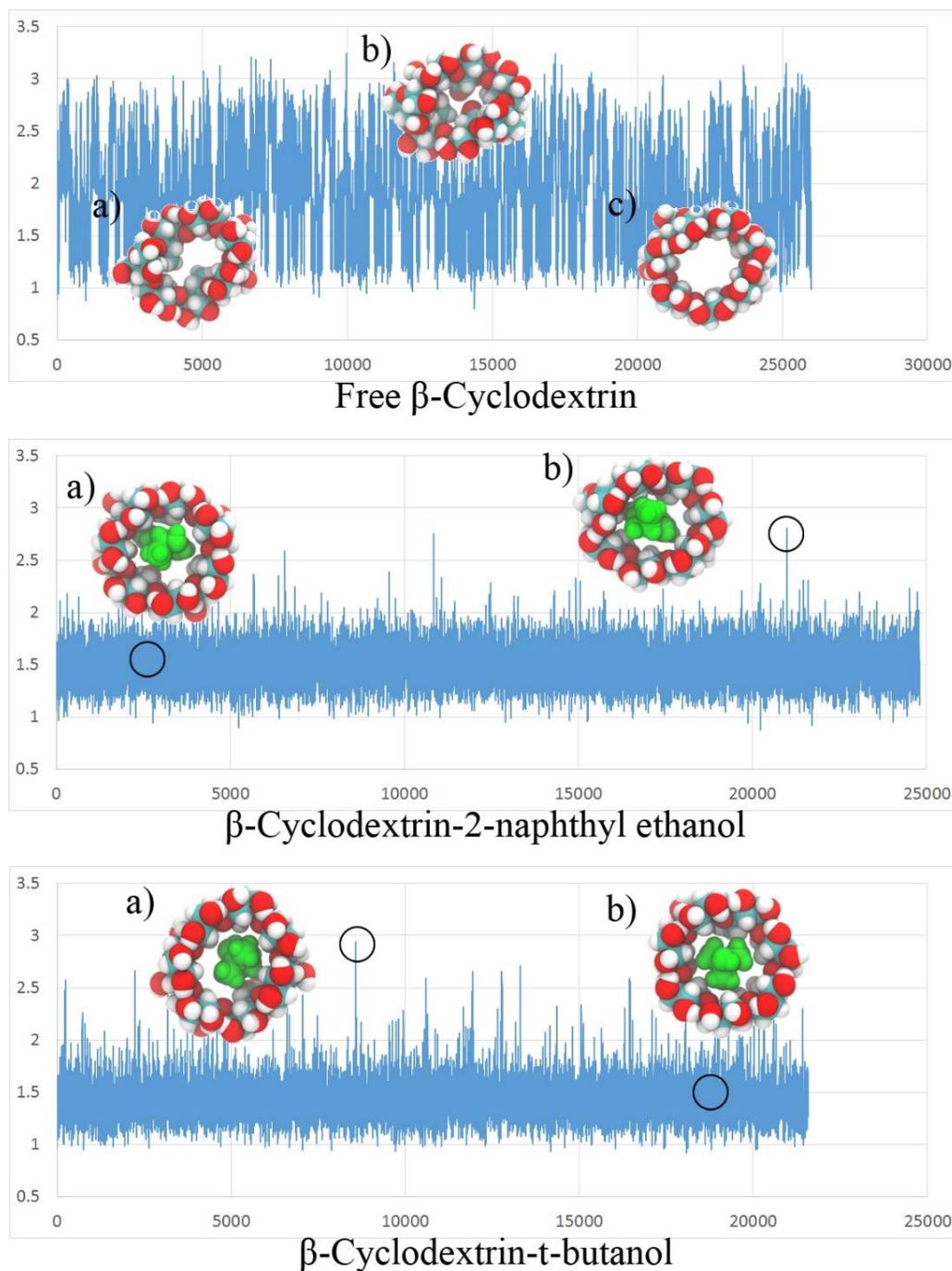


Figure 5.11. The RMSDs in reference with the crystal structure plotted against the frame number in q4md-CD force field. The RMSD is in Å. The representative conformations are shown above the plots and circled on the plots.

## 5.4 Conclusion

In this work a complete protocol for calculating the binding free energy by computing the solute and solvent enthalpy and entropy respectively is presented with  $\beta$ -cyclodextrin-complexes as the studied system. The enthalpy changes of the solute and solvent were calculated from potential energy calculation. The entropy change of the solutes were evaluated by dihedral population analysis and integration of external entropy from the MD trajectory. The solvent entropy was obtained from the grid cell method calculation. The calculated kinetics and thermodynamics results match with the experimental values reasonably well. The weak ligands, 1-propanol, 1-butanol, t-butanol and methyl butyrate were entropically driven to bind with  $\beta$ -cyclodextrin while the stronger ligands, aspirin, 1-naphthylethanol and 2-naphthylethanol were mixed entropy and enthalpy driven binding. The entropy driving forces of these ligands were from water entropy gain and  $\beta$ -cyclodextrin entropy gain, while the desolvation energy was the major reason of the enthalpic repulsion. The desolvation energy was from the excluded volume on binding and the loss of solute-water interactions. Binding pathways of the two types of ligands were observed from the MD trajectories and the corresponding thermodynamic reason were discussed. This provides important clues for relieving the entropic penalty in strong binding drugs and also fundamental understanding of the role of waters in the thermodynamics and kinetics.

## 5.5 References

1. Tummino, P. J., and Copeland, R. A. (2008) Residence Time of Receptor-Ligand Complexes and Its Effect on Biological Function (vol 47, pg 5481, 2008). *Biochemistry* **47**, 8465-8465
2. Copeland, R. A., Pompliano, D. L., and Meek, T. D. (2007) Drug-Target Residence Time and Its Implications for Lead Optimization (vol 5, pg 730, 2006). *Nature Reviews Drug Discovery* **6**, 249-249
3. Schneider, E. V., Boettcher, J., Huber, R., Maskos, K., and Neumann, L. (2013) Structure-Kinetic Relationship Study of CDK8/CycC Specific Compounds. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8081-8086
4. Huang, Y.-m. M., Kang, M., and Chang, C.-e. A. (2014) Switches of Hydrogen Bonds During Ligand-Protein Association Processes Determine Binding Kinetics. *Journal of Molecular Recognition* **27**, 537-548
5. Zwanzig, R. W. (1954) High-Temperature Equation of State by a Perturbation Method .1. Nonpolar Gases. *Journal of Chemical Physics* **22**, 1420-1426
6. Straatsma, T. P., and McCammon, J. A. (1990) Free-Energy Thermodynamic Integrations in Molecular-Dynamics Simulations Using a Noniterative Method to Include Electronic Polarization. *Chemical Physics Letters* **167**, 252-254
7. Izatt, R. M., Pawlak, K., Bradshaw, J. S., and Bruening, R. L. (1991) Thermodynamic and Kinetic Data for Macrocyclic Interaction with Cations and Anions. *Chemical Reviews* **91**, 1721-2085
8. Straatsma, T. P., and McCammon, J. A. (1991) Multiconfiguration Thermodynamic Integration. *Journal of Chemical Physics* **95**, 1175-1188

9. Khavrutskii, I. V., and Wallqvist, A. (2011) Improved Binding Free Energy Predictions from Single-Reference Thermodynamic Integration Augmented with Hamiltonian Replica Exchange. *Journal of Chemical Theory and Computation* **7**, 3001-3011
10. Torrie, G. M., and Valleau, J. P. (1977) Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *Journal of Computational Physics* **23**, 187-199
11. Hainam, D., Hirst, J. D., and Wheatley, R. J. (2012) Calculation of Partition Functions and Free Energies of a Binary Mixture Using the Energy Partitioning Method: Application to Carbon Dioxide and Methane. *Journal of Physical Chemistry B* **116**, 4535-4542
12. Miller, B. R., III, McGee, T. D., Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **8**, 3314-3321
13. Chang, C. E., and Gilson, M. K. (2004) Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *Journal of the American Chemical Society* **126**, 13156-13164
14. Muddana, H. S., and Gilson, M. K. (2012) Calculation of Host-Guest Binding Affinities Using a Quantum-Mechanical Energy Model. *Journal of Chemical Theory and Computation* **8**, 2023-2033
15. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of the American Chemical Society* **112**, 6127-6129
16. Borukhov, I., Andelman, D., and Orland, H. (1997) Steric Effects in Electrolytes: A Modified Poisson-Boltzmann Equation. *Physical Review Letters* **79**, 435-438

17. Boschitsch, A. H., and Fenley, M. O. (2004) Hybrid Boundary Element and Finite Difference Method for Solving the Nonlinear Poisson-Boltzmann Equation. *Journal of Computational Chemistry* **25**, 935-955
18. Rocchia, W., Alexov, E., and Honig, B. (2001) Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *Journal of Physical Chemistry B* **105**, 6507-6514
19. Mondal, J., Morrone, J. A., and Berne, B. J. (2013) How Hydrophobic Drying Forces Impact the Kinetics of Molecular Recognition. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 13277-13282
20. Sasikala, W. D., and Mukherjee, A. (2014) Single Water Entropy: Hydrophobic Crossover and Application to Drug Binding. *Journal of Physical Chemistry B* **118**, 10553-10564
21. Zangi, R., and Berne, B. J. (2008) Temperature Dependence of Dimerization and Dewetting of Large-Scale Hydrophobes: A Molecular Dynamics Study. *Journal of Physical Chemistry B* **112**, 8634-8644
22. Kurisaki, I., Barberot, C., Takayanagi, M., and Nagaoka, M. (2015) Dewetting of Si-Pocket via Water Channel upon Thrombin-Substrate Association Reaction. *Journal of Physical Chemistry B* **119**, 15807-15812
23. Lazaridis, T., and Karplus, M. (1996) Orientational Correlations and Entropy in Liquid Water. *Journal of Chemical Physics* **105**, 4294-4316
24. Lazaridis, T. (1998) Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *Journal of Physical Chemistry B* **102**, 3531-3541
25. Lazaridis, T. (1998) Inhomogeneous Fluid Approach to Solvation Thermodynamics. 2. Applications to simple fluids. *Journal of Physical Chemistry B* **102**, 3542-3550

26. Nguyen, C. N., Young, T. K., and Gilson, M. K. (2012) Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit 7 uril. *Journal of Chemical Physics* **137**
27. Lin, S.-T., Maiti, P. K., and Goddard, W. A., III. (2010) Two-Phase Thermodynamic Model for Efficient and Accurate Absolute Entropy of Water from Molecular Dynamics Simulations. *Journal of Physical Chemistry B* **114**, 8191-8198
28. Lin, S. T., Blanco, M., and Goddard, W. A. (2003) The Two-Phase Model for Calculating Thermodynamic Properties of Liquids From Molecular Dynamics: Validation for the Phase Diagram of Lennard-Jones Fluids. *Journal of Chemical Physics* **119**, 11792-11805
29. Henchman, R. H. (2007) Free Energy of Liquid Water From a Computer Simulation Via Cell Theory. *Journal of Chemical Physics* **126**
30. Gerogiokas, G., Calabro, G., Henchman, R. H., Southey, M. W. Y., Law, R. J., and Michel, J. (2014) Prediction of Small Molecule Hydration Thermodynamics with Grid Cell Theory. *Journal of Chemical Theory and Computation* **10**, 35-48
31. Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., Young, C., Batson, B., Bowers, K. J., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., Ierardi, D. J., Kolossvary, I., Klepeis, J. L., Layman, T., McLeavey, C., Moraes, M. A., Mueller, R., Priest, E. C., Shan, Y., Spengler, J., Theobald, M., Towles, B., and Wang, S. C. (2008) Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Communications of the Acm* **51**, 91-97
32. Shaw, D. E., Grossman, J. P., Bank, J. A., Batson, B., Butts, J. A., Chao, J. C., Deneroff, M. M., Dror, R. O., Even, A., Fenton, C. H., Forte, A., Gagliardo, J., Gill, G., Greskamp, B., Ho, C. R., Ierardi, D. J., Iserovich, L., Kuskin, J. S.,

- Larson, R. H., Layman, T., Li-Siang, L., Lerer, A. K., Li, C., Killebrew, D., Mackenzie, K. M., Mok, S. Y. H., Moraes, M. A., Mueller, R., Nociolo, L. J., Peticolas, J. L., Quan, T., Ramot, D., Salmon, J. K., Scarpazza, D. P., Schafer, U. B., Siddique, N., Snyder, C. W., Spengler, J., Tang, P. T. P., Theobald, M., Toma, H., Towles, B., Vitale, B., Wang, S. C., and Young, C. (2014) Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*, 41-53
33. Arkhipov, A., Shan, Y., Das, R., Endres, N. F., Eastwood, M. P., Wemmer, D. E., Kuriyan, J., and Shaw, D. E. (2013) Architecture and Membrane Interactions of the EGF Receptor. *Cell* **152**, 557-569
34. Pierce, L. C. T., Salomon-Ferrer, R., de Oliveira, C. A. F., McCammon, J. A., and Walker, R. C. (2012) Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation* **8**, 2997-3002
35. Frembgen-Kesner, T., and Elcock, A. H. (2006) Computational Sampling of a Cryptic Drug Binding Site in a Protein Receptor: Explicit Solvent Molecular Dynamics and Inhibitor Docking to p38 MAP Kinase. *Journal of Molecular Biology* **359**, 202-214
36. Filomia, F., De Rienzo, F., and Menziani, M. C. (2010) Insights into MAPK p38 Alpha DFG Flip Mechanism by Accelerated Molecular Dynamics. *Bioorganic & Medicinal Chemistry* **18**, 6805-6812
37. Yang, Y., Shen, Y., Liu, H., and Yao, X. (2011) Molecular Dynamics Simulation and Free Energy Calculation Studies of the Binding Mechanism of Allosteric Inhibitors with p38 alpha MAP Kinase. *Journal of Chemical Information and Modeling* **51**, 3235-3246

38. Hornak, V., Okur, A., Rizzo, R. C., and Simmerling, C. (2006) HIV-1 Protease Flaps Spontaneously Open and Reclose in Molecular Dynamics Simulations. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 915-920
39. Perryman, A. L., Lin, J. H., and McCammon, J. A. (2004) HIV-1 Protease Molecular Dynamics of a Wild-Type and of the V82F/I84V Mutant: Possible Contributions to Drug Resistance and a Potential New Target Site for Drugs. *Protein Science* **13**, 1108-1123
40. Bowman, G. R., Pande, V. S., and Noe, F. (2014) An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. Springer Science+Business Media, USA
41. Dellago, C., Bolhuis, P. G., Csajka, F. S., and Chandler, D. (1998) Transition Path Sampling and the Calculation of Rate Constants. *Journal of Chemical Physics* **108**, 1964-1977
42. Gharib, R., Greige-Gerges, H., Fourmentin, S., Charcosset, C., and Auezova, L. (2015) Liposomes Incorporating Cyclodextrin-Drug Inclusion Complexes: Current State of Knowledge. *Carbohydrate Polymers* **129**, 175-186
43. Wei, G., Dong, R., Wang, D., Feng, L., Dong, S., Song, A., and Hao, J. (2014) Functional Materials From the Covalent Modification of Reduced Graphene Oxide and Beta-Cyclodextrin as a Drug Delivery Carrier. *New Journal of Chemistry* **38**, 140-145
44. Bourkaib, N., Zhou, J., Yao, J., Fang, Z., and Mezghrani, O. (2013) Combination of Beta-Cyclodextrin Inclusion Complex and Self-Microemulsifying Drug Delivery System for Photostability and Enhanced Oral Bioavailability of Methotrexate: Novel Technique. *Drug Development and Industrial Pharmacy* **39**, 918-927

45. Shin, J.-A., Lim, Y.-G., and Lee, K.-H. (2012) Copper-Catalyzed Azide-Alkyne Cycloaddition Reaction in Water Using Cyclodextrin as a Phase Transfer Catalyst. *Journal of Organic Chemistry* **77**, 4117-4122
46. Jiang, L., Deng, M., Wang, Y., Liang, D., Yan, Y., and Huang, J. (2009) Special Effect of beta-Cyclodextrin on the Aggregation Behavior of Mixed Cationic/Anionic Surfactant Systems. *Journal of Physical Chemistry B* **113**, 7498-7504
47. Elk, J. C., and Benjamin, I. (2015) beta-Cyclodextrin at the Water/1-Bromobutane Interface: Molecular Insight into Reverse Phase Transfer Catalysis. *Langmuir* **31**, 5086-5092
48. Giuffrida, S., Ventimiglia, G., Petralia, S., Conoci, S., and Sortino, S. (2006) Facile Light-Triggered One-Step Synthesis of Small and Stable Platinum Nanoparticles in an Aqueous Medium From a Beta-Cyclodextrin Host-Guest Inclusion Complex. *Inorganic Chemistry* **45**, 508-510
49. Mangolim, C. S., Moriwaki, C., Nogueira, A. C., Sato, F., Baesso, M. L., Neto, A. M., and Matioli, G. (2014) Curcumin-Beta-Cyclodextrin Inclusion Complex: Stability, Solubility, Characterisation by FT-IR, FT-Raman, X-Ray Diffraction and Photoacoustic Spectroscopy, and Food Application. *Food Chemistry* **153**, 361-370
50. Marcolino, V. A., Zanin, G. M., Durrant, L. R., Benassi, M. D. T., and Matioli, G. (2011) Interaction of Curcumin and Bixin with beta-Cyclodextrin: Complexation Methods, Stability, and Applications in Food. *Journal of Agricultural and Food Chemistry* **59**, 3348-3357
51. Siro, I., Fenyvesi, E., Szente, L., De Meulenaer, B., Devlieghere, F., Orgovanyi, J., Senyi, J., and Barta, J. (2006) Release of Alpha-Tocopherol From Antioxidative Low-Density Polyethylene Film Into Fatty Food Simulant:

Influence of Complexation in Beta- Cyclodextrin. *Food Additives and Contaminants Part a-Chemistry Analysis Control Exposure & Risk Assessment* **23**, 845-853

52. Hedges, A., Shieh, W., and Ammeraal, R. Removal of Beta-Cyclodextrin From Egg Yolk Material by Hydrolysis Under Acid Conditions with Cyclodextrin Glycosyl Transferase and Amylase, Opt. Also Debranching Enzyme. American Maize Prod Co; Cerestar USA Inc
53. Marchetti, L., and Levine, M. (2011) Biomimetic Catalysis. *Acs Catalysis* **1**, 1090-1118
54. Rekharsky, M. V., and Inoue, Y. (1998) Complexation Thermodynamics of Cyclodextrins. *Chemical Reviews* **98**, 1875-1917
55. Izatt, R. M., Pawlak, K., Bradshaw, J. S., and Bruening, R. L. (1995) Thermodynamic and Kinetic Data for Macrocycle Interaction with Cations, Anions, and Neutral Molecules. *Chemical Reviews* **95**, 2529-2586
56. Yim, C. T., Zhu, X. X., and Brown, G. R. (1999) Kinetics of Inclusion Reactions of Beta-Cyclodextrin with Several Dihydroxycholesterol Ions Studied by NMR Spectroscopy. *Journal of Physical Chemistry B* **103**, 597-602
57. Nilsson, M., Valente, A. J. M., Olofsson, G., Soderman, O., and Bonini, M. (2008) Thermodynamic and Kinetic Characterization of Host-Guest Association Between Bolaform Surfactants and Alpha- and Beta-Cyclodextrins. *Journal of Physical Chemistry B* **112**, 11310-11316
58. Fukahori, T., Kondo, M., and Nishikawa, S. (2006) Dynamic Study of Interaction Between Beta-Cyclodextrin and Aspirin by the Ultrasonic Relaxation Method. *Journal of Physical Chemistry B* **110**, 4487-4491

59. Fukahori, T., Nishikawa, S., and Yamaguchi, K. (2004) Kinetics on Isomeric Alcohols Recognition by Alpha- and Beta-Cyclodextrins Using Ultrasonic Relaxation Method. *Bulletin of the Chemical Society of Japan* **77**, 2193-2198
60. Yamaguchi, K., Fukahori, T., and Nishikawa, S. (2005) Dynamic Interaction Between Alkylammonium Ions and Beta-Cyclodextrin by Means of Ultrasonic Relaxation. *Journal of Physical Chemistry A* **109**, 40-43
61. Fukahori, T., Nishikawa, S., and Yamaguchi, K. (2004) Ultrasonic Relaxation Due to Inclusion Complex of Amino Acid by Beta-Cyclodextrin in Aqueous Solution. *Journal of the Acoustical Society of America* **115**, 2325-2330
62. Nishikawa, S., Yamaguchi, K., and Fukahori, T. (2003) Ultrasonic Relaxation Due to Complexation Reaction Between Beta-Cyclodextrin and Alkylammonium Ions. *Journal of Physical Chemistry A* **107**, 6415-6418
63. Nishikawa, S., Fukahori, T., and Ishikawa, K. (2002) Ultrasonic Relaxations in Aqueous Solutions of Propionic Acid in the Presence and Absence of Beta-Cyclodextrin. *Journal of Physical Chemistry A* **106**, 3029-3033
64. Nishikawa, S., Ugawa, T., and Fukahori, T. (2001) Molecular Recognition Kinetics of Beta-Cyclodextrin for Several Alcohols by Ultrasonic Relaxation Method. *Journal of Physical Chemistry B* **105**, 7594-7597
65. Barros, T. C., Stefaniak, K., Holzwarth, J. F., and Bohne, C. (1998) Complexation of Naphthylethanols with Beta-Cyclodextrin. *Journal of Physical Chemistry A* **102**, 5639-5651
66. Kondo, M., and Nishikawa, S. (2007) Inclusion Kinetics of a Nucleotide Into a Cyclodextrin Cavity by Means of Ultrasonic Relaxation. *Journal of Physical Chemistry B* **111**, 13451-13454

67. Nishikawa, S., Kondo, M., Kamimura, E., and Xing, S. (2007) Ultrasonic Relaxation Associated with Inclusion Complex of Drugs and Beta-Cyclodextrin. *Bulletin of the Chemical Society of Japan* **80**, 694-698
68. Nishikawa, S., and Kondo, M. (2006) Kinetic Study for the Inclusion Complex of Carboxylic Acids with Cyclodextrin by the Ultrasonic Relaxation Method. *Journal of Physical Chemistry B* **110**, 26143-26147
69. Jana, M., and Bandyopadhyay, S. (2013) Molecular Dynamics Study of beta-Cyclodextrin-Phenylalanine (1:1) Inclusion Complex in Aqueous Medium. *Journal of Physical Chemistry B* **117**, 9280-9287
70. Wickstrom, L., He, P., Gallicchio, E., and Levy, R. M. (2013) Large Scale Affinity Calculations of Cyclodextrin Host-Guest Complexes: Understanding the Role of Reorganization in the Molecular Recognition Process. *Journal of Chemical Theory and Computation* **9**, 3136-3150
71. Pacioni, N. L., Pierini, A. B., and Veglia, A. V. (2013) Structural Characterization of N-Methylcarbamate: Beta-Cyclodextrin Complexes by Experimental Methods and Molecular Dynamics Simulations. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **103**, 319-324
72. Koehler, J. E. H., and Grczelschak-Mick, N. (2013) The Beta-Cyclodextrin/Benzene Complex and Its Hydrogen Bonds - a Theoretical Study Using Molecular Dynamics, Quantum Mechanics and COSMO-RS. *Beilstein Journal of Organic Chemistry* **9**, 118-134
73. Sellner, B., Zifferer, G., Kornherr, A., Krois, D., and Brinker, U. H. (2008) Molecular Dynamics Simulations of Beta-Cyclodextrin-Aziadamantane Complexes in Water. *Journal of Physical Chemistry B* **112**, 710-714

74. Weinzinger, P., Weiss-Greiler, P., Snor, W., Viernstein, H., and Wolschann, P. (2007) Molecular Dynamics Simulations and Quantum Chemical Calculations on Beta-Cyclodextrin Spirolactone Complex. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **57**, 29-33
75. Choi, Y. J., Lee, J., Cho, K. W., Hwang, S., Jeong, K., and Jung, S. (2005) Binding Geometry of Inclusion Complex as a Determinant Factor for Aqueous Solubility of the Flavonoid/Beta-Cyclodextrin Complexes Based on Molecular Dynamics Simulations. *Bulletin of the Korean Chemical Society* **26**, 1203-1208
76. Aiquan, J., Xing, Z., Xueming, X., and Zhengyu, J. (2013) Molecular Dynamics Simulations of Cyclodextrin-Cumene Hydroperoxide Complexes in Water. *Computational and Theoretical Chemistry* **1013**, 1-6
77. Chen, W., Chang, C. E., and Gilson, M. K. (2004) Calculation of Cyclodextrin Binding Affinities: Energy, Entropy, and Implications for Drug Design. *Biophysical Journal* **87**, 3035-3049
78. Tidemand, K. D., Schonbeck, C., Holm, R., Westh, P., and Peters, G. H. (2014) Computational Investigation of Enthalpy-Entropy Compensation in Complexation of Glycoconjugated Bile Salts with beta-Cyclodextrin and Analogs. *Journal of Physical Chemistry B* **118**, 10889-10897
79. Nutho, B., Khuntawee, W., Rungnim, C., Pongsawasdi, P., Wolschann, P., Karpfen, A., Kungwan, N., and Rungrotmongkol, T. (2014) Binding Mode and Free Energy Prediction of Fisetin/Beta-Cyclodextrin Inclusion Complexes. *Beilstein Journal of Organic Chemistry* **10**, 2789-2799
80. Sun, T., Shao, X., and Cai, W. (2010) Self-Assembly Behavior of Beta-Cyclodextrin and Imipramine. a Free Energy Perturbation Study. *Chemical Physics* **371**, 84-90

81. Katritzky, A. R., Fara, D. C., Yang, H. F., Karelson, M., Suzuki, T., Solov'ev, V. P., and Varnek, A. (2004) Quantitative Structure-Property Relationship Modeling of Beta-Cyclodextrin Complexation Free Energies. *Journal of Chemical Information and Computer Sciences* **44**, 529-541
82. Bonnet, P., Jaime, C., and Morin-Allory, L. (2002) Structure and Thermodynamics of Alpha-, Beta-, and Gamma-Cyclodextrin Dimers. Molecular Dynamics Studies of the Solvent Effect and Free Binding Energies. *Journal of Organic Chemistry* **67**, 8602-8609
83. Klein, C. T., Polheim, D., Viernstein, H., and Wolschann, P. (2000) A Method for Predicting the Free Energies of Complexation Between Beta-Cyclodextrin and Guest Molecules. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **36**, 409-423
84. Henriksen, N. M., Fenley, A. T., and Gilson, M. K. (2015) Computational Calorimetry: High-Precision Calculation of Host-Guest Binding Thermodynamics. *Journal of Chemical Theory and Computation* **11**, 4377-4394
85. Pedretti, A., Villa, L., and Vistoli, G. (2004) VEGA - An Open Platform to Develop Chemo-Bio-Informatics Applications, Using Plug-in Architecture and Script Programming. *Journal of Computer-Aided Molecular Design* **18**, 167-173
86. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **25**, 1157-1174
87. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven,

- T., Montgomery Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, N. J., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09. Gaussian, Inc., Wallingford, CT, USA
88. Cezard, C., Trivelli, X., Aubry, F., Djedaini-Pilard, F., and Dupradeau, F.-Y. (2011) Molecular Dynamics Studies of Native and Substituted Cyclodextrins in Different Media: 1. Charge Derivation and Force Field Performances. *Physical Chemistry Chemical Physics* **13**, 15103-15121
89. Fenley, A. T., Henriksen, N. M., Muddana, H. S., and Gilson, M. K. (2014) Bridging Calorimetry and Simulation through Precise Calculations of Cucurbituril-Guest Binding Enthalpies. *Journal of Chemical Theory and Computation* **10**, 4069-4078
90. Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of Molecular Graphics & Modelling* **25**, 247-260
91. Doll, J. D., and Dion, D. R. (1976) Generalized Langevin Equation Approach for Atom-Solid-Surface Scattering - Numerical Techniques for Gaussian Generalized Langevin Dynamics. *Journal of Chemical Physics* **65**, 3762-3766
92. Adelman, S. A. (1979) Generalized Langevin Theory for Many-Body Problems in Chemical-Dynamics - General Formulation and the Equivalent Harmonic Chain Representation. *Journal of Chemical Physics* **71**, 4471-4486

93. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics & Modelling* **14**, 33-38
94. Flyvbjerg, H., and Petersen, H. G. (1989) Error-Estimates on Averages of Correlated Data. *Journal of Chemical Physics* **91**, 461-466
95. Irudayam, S. J., and Henschman, R. H. (2010) Solvation Theory to Provide a Molecular Interpretation of the Hydrophobic Entropy Loss of Noble-Gas Hydration. *Journal of Physics-Condensed Matter* **22**
96. Pauling, L. (1935) The Structure and Entropy of Ice and of Other Crystals with Some Randomness of Atomic Arrangement. *Journal of the American Chemical Society* **57**, 2680-2684
97. Chang, C. E., Potter, M. J., and Gilson, M. K. (2003) Calculation of Molecular Configuration Integrals. *Journal of Physical Chemistry B* **107**, 1048-1055

## Chapter 6 PSIM: Pathway Search Guided by Internal Motions

### 6.1 Introduction

Thorough conformational sampling is one important objective of computational chemistry. It is a prerequisite step in a variety of fundamental scientific topics such as protein folding<sup>1-6</sup>, protein kinetics<sup>7-9</sup>, non-bonding interactions<sup>10-12</sup>, and accurate free energy evaluation<sup>13-15</sup>. Furthermore, it also has industrial or pharmaceutical applications like drug design<sup>16-19</sup>, crystal structure prediction<sup>20-23</sup> and etc. Despite its key role, how to conduct conformational sampling efficiently remains as an open question because of the intrinsic difficulties lying in it. Usually undetermined but high energy barrier stops the sampling from hopping between minima in practical computational time scale. Even by focusing on a local region without considering the energy barrier, a thorough sampling requires intensive efforts because it means to capture solute rearrangement, solvent effects and molecular interactions. For large systems like a protein, there are enormous number of local degenerate or non-degenerate thermodynamic states, which also makes a thorough sampling intractable. To perform conformational sampling with limited computational power, compensation is often made by sacrificing the accuracy or resolution of the sampling and this results in loss of desirable details.

So far three major categories of conformational sampling methods have been developed over the past few decades. The first class is based on molecular dynamics. The most straightforward and fundamental method among them is conventional molecular dynamics (MD) simulation. This method can generate physically meaningful trajectories

and reproduce unbiased energetic profiles, but its time scale, even with contemporary graphic processor units (GPU) accelerated techniques or specialized Anton machines<sup>24</sup>,<sup>25</sup>, is still limited to millisecond or microsecond<sup>25-28</sup>. This leads to the invention of many derivatives or specialized protocols of MD. The first branch of methods adds additional forces to guide the system to some target<sup>29,30</sup>, lowers the probability of redundant sampling of the conformational space<sup>31-36</sup>, or exaggerates the concurrent motion of the system<sup>37</sup>. The other group of MD based methods smooths out the details on the potential energy surface (PES) and thus enhance the sampling. The strategy leads to coarse grained description of systems<sup>38</sup>, accelerated MD<sup>39</sup>, GaMD<sup>15</sup>, RaMD-db method<sup>40</sup>, softcore force field description<sup>41</sup>, and many similar methods. Milestoning<sup>42,43</sup>, which utilizes the idea of Markov chain model, enhances sampling and also allows direct calculation of kinetic properties. By considering the behavior of slow events, people realized that a major part of these long events actually consists of a dominant waiting time and relatively rapid sudden transition<sup>44</sup>. This leads to the third group of MD variants, or more exactly, MD protocols, that focus on shortening the waiting time of the slow event by constant restarting the MD from carefully selected conformation on the sampled trajectories. This group includes, progress index based resampling<sup>45</sup>, resampling from free energy basin edges<sup>46</sup>, WExplore<sup>47</sup>, and other resampling methods using various criteria<sup>48-52</sup>. Another accelerating technique roots in increasing the temperature of the MD to boost the rate of high energy events. This results in the multi-ensemble methods, like parallel replica method (PRM)<sup>53</sup>, replica exchange MD (REM)<sup>54,55</sup>, MuSTART MD method<sup>56</sup>. It

combines with the idea <sup>44</sup> in previous group, and leads to transition path sampling <sup>57, 58</sup> and transition interface sampling <sup>59</sup>.

The second category of sampling methods is based on Monte Carlo (MC) sampling <sup>60, 61</sup>. A series of variants of original MC methods have been proposed since 1970s. The past four decades have witnessed force-biased MC <sup>62</sup>, smart MC <sup>63</sup>, hybrid MC (HMC) <sup>64</sup>, J-walking <sup>65</sup>, Simulated tempering <sup>66, 67</sup>, multiple-stage J-walking <sup>68</sup>, parallel tempering <sup>69, 70</sup>, S-walking <sup>71</sup>, flat energy with probability modifier <sup>72</sup>, smart darting MC <sup>73</sup>, C-walking <sup>74</sup> and finite reservoir replica exchange method (FRREM) <sup>75</sup>.

The last category of sampling methods roots deeply in normal mode analysis, which requires no or little dynamics of the system but provides a tremendous reservoir of new conformations efficiently. These methods started from LMOD <sup>76</sup>, in which a minimized conformation is distorted along the low normal modes and then minimized to obtain new conformations. A variants of it came as LLMOD <sup>77</sup>, in which the hessian matrix calculation is avoided by only looking at the low modes implicitly. A Tork method <sup>78</sup> was developed and it builds the hessian matrix eigenvectors and distorts the minimized conformation in internal coordinates to avoid unphysical distortion in Cartesian space. Recently in Low Mode MD method (LMMD) <sup>79</sup>, the mathematical distortion process of LMOD was replaced by a short distortion using low mode guided MD simulation. Minimized RMSD was used as a criterion to select new conformations by distortion along low normal modes <sup>80</sup>. The normal mode analysis was also combined with elastic models and gave some useful coarse grained level methods <sup>81, 82</sup>.

Principle component analysis (PCA), which is conceptually similar with normal mode analysis and quasi-harmonic approximation, also provides an orthogonal set of intrinsic motions borne in the dynamics of systems. This set of orthogonal motions is similar to normal modes in several ways. It is orthogonal and the amplitude of the motion is governed by the corresponding eigenvalue. It tells the intrinsic motions of the system from the most significant to the least significant sorted by the eigenvalues. On the other hand, it differs from normal modes in the sense that the motions are produced from a continuous region traveled by the trajectory instead of the rigorous mathematical motions at one single minimum. This advantage ensures that the motions from PCA can be used to describe further conformational changes of the system than the normal modes do. In the study of proteins, PCA is usually used to analyze the backbone motion by looking at the Cartesian coordinates of the  $\alpha$ -carbon atoms in each residue. In this study, we refer to this as  $\alpha$ -carbon PCA.

Dihedral PCA, which is an internal coordinate version of conventional PCA, has been proved to produce more informative results than its original work<sup>83-88</sup>. More importantly, not like  $\alpha$ -carbon PCA, which is limited to  $\alpha$ -carbon atoms when applied to protein systems, dihedral PCA does not have the limitation on backbone  $\psi$ ,  $\phi$  angles and can be naturally extended to side chain dihedrals so that the full atomic details can be retained in the dihedral PCA results. However, dihedral PCA has two shortcomings, 1) it is unable to reproduce Cartesian motions from the dihedral PCA modes, and 2) it doubles the number of modes than that of the original degrees of freedom. In Dihedral PCA, people have solved the issue arising from the periodicity of dihedral angles using trigonometric

function operations. For an  $N_{\text{dih}}$  dihedral angle system each degree of freedom from the dihedral angle is split into a sin and a cos term, so that the dimension of the covariance matrix becomes  $2N_{\text{dih}}$ , and thus  $2N_{\text{dih}}$  eigenvectors are yielded in the calculation. This  $2N_{\text{dih}}$  eigenvectors are orthogonal in the space of sin and cos but, on the other hand, the  $N_{\text{dih}}$  dihedral variance vectors reproduced by merging the sin and cos terms are no longer orthogonal to each other in the dihedral space. Therefore eigenvectors obtained by this sin and cos approach cannot be used as orthogonal directions for conformational search. This issue is solved by complex dihedral PCA<sup>84</sup>, which uses complex representation in Euler's formula to replace the one to two mapping in sin cos dihedral PCA. However the computation power consumed in the diagonalization of covariance matrix in this approach still doubles.

Meanwhile, by considering the efficiency of existing normal mode based conformational search methods, the most time consuming part of these methods is the acceptance test. Typically a full minimization is performed and conformations within a certain energy threshold are accepted. This is rigorous and efficient for small molecules but for protein-like large molecules, a full minimization is less efficient or sometimes even intractable. To avoid energy evaluation and energy minimization, a collision test like approach has been proposed<sup>89</sup>. In this approach, no energy evaluation is necessary. Instead, only distance based collision test is conducted to simplify the interatomic potential to a steep wall type potential. In this way thousands of conformations can be accepted or rejected per second<sup>90</sup>. Following that idea, methods have been proposed to use collision as the acceptance/rejection test in coarse grained elastic model<sup>91</sup>.

Combining those thoughts, we propose a novel conformational search method in this work. A novel internal PCA is proposed that can reproduce full atomic motions comparable to classical  $\alpha$ -carbon Cartesian PCA. It produces same number of modes as the degrees of freedom. A conformational search algorithm, referred to as PSIM, in which an initial conformation is continuously distorted along the internal PC motions, is developed. A collision based acceptance test is used to ensure physically meaning motions of the system with high efficiency. This method was tested in several real molecular systems.

## 6.2 Method

### 6.2.1 Method Overview

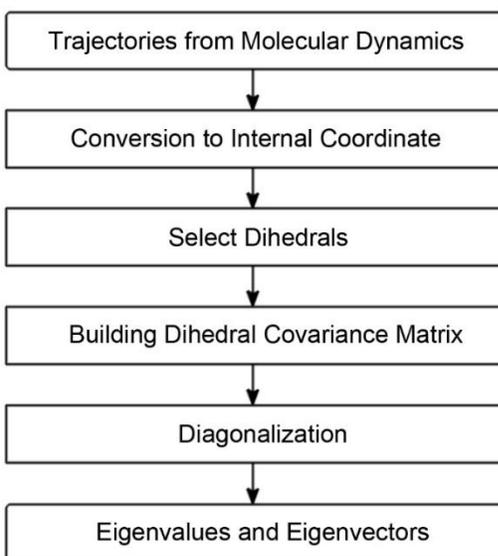


Figure 6.1. The flowchart of internal Principal Component analysis (internal PCA).

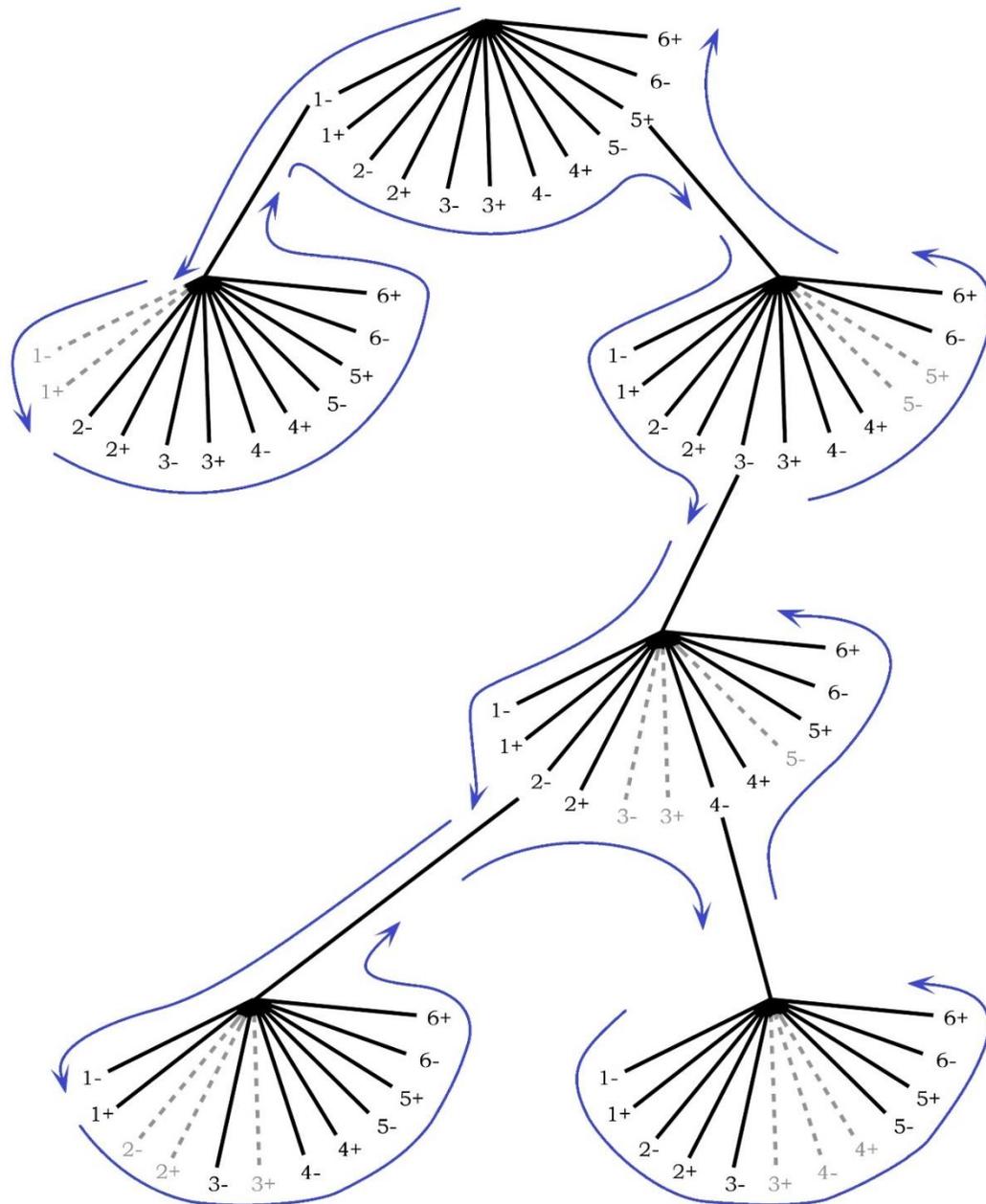


Figure 6.2. A representative systematic search tree of 5 PC modes. Disabled PC modes are rendered in gray dashed lines, while all other PC modes in use are in black solid lines. Only representative sub branches are shown. The blue arrowed curves show a representative flow of the search on the tree diagram.

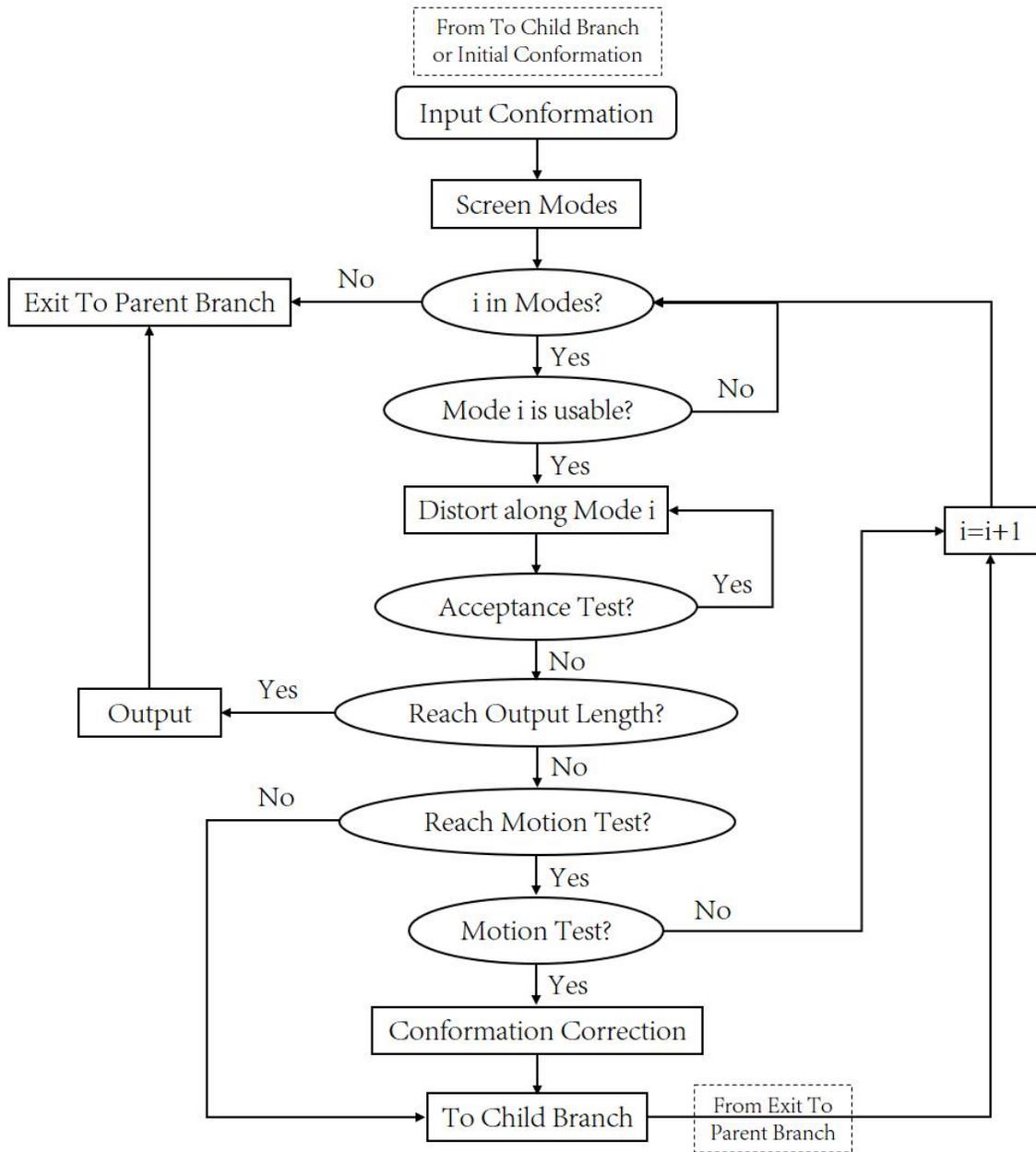


Figure 6.3. The flowchart of each search branch.

PSIM uses internal PCA modes computed from a MD trajectory to guide the searches. Figure 6.1 illustrates the steps which are similar to computing classical PCA in Cartesian coordinate except additional steps needed for use of internal coordinates, *Conversion to Internal Coordinate* and *Building Dihedral Covariance Matrix*. Because a biomolecules are significantly larger than chemical compounds, we developed multi-layer internal coordinates, as detailed in later sections. The overview of the algorithm used in PSIM is shown in Figure 6.2 and Figure 6.3. In this algorithm, the internal PC modes are sorted by the extend the ligand moves instead of eigenvalues. A minimized receptor-ligand complex is used as an initial structure to begin the search. The conformational search is then conducted in a systematic manner (Figure 6.2). First, the initial conformation is distorted stepwise along both positive and negative directions of each PC modes until acceptance test rejects the new conformation. Then, the conformational search continues systematically using the new conformations from two directions of each mode as the starting point, except that the PC modes are screened before used in the sub branches. In the course of conformational search, the modes used and the number of steps of distortion are tracked as the path of the trajectory. When the number of steps exceeds a predefined parameter, a motion test is performed to ensure the guest/ligand moves away from its original position and the host/protein maintains a reasonable conformation while being distorted. A failed motion test backs up the search along the recorded search path. If the motion test is passed a structure correction will be performed to reduce the artifacts induced by the distortion, and the search is then continued from the corrected conformation. The search algorithm dumps distortion path when the total step of

distortion exceeds a predefined path length. The algorithm is exited after the systematic search exhausts the search tree within the predefined path length.

## 6.2.2 Computational Details

### 6.2.2.1 Multi-Layer Internal Coordinate Definition

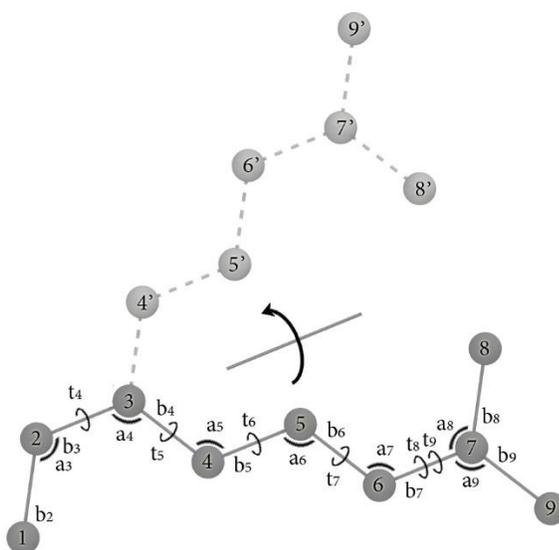


Figure 6.4. Classic internal coordinate representation (Z-matrix) for a small molecule. Atom 1 is at the origin; atom 2 is defined by a bond length ( $b_2$ ) and placed on the z-axis; and atom 3 is defined by a bond length ( $b_3$ ), bond angle ( $a_3$ ) and placed on the x-z plane. Six external coordinates are eliminated, and a molecule of  $N$  atoms has  $3N-6$  internal coordinates. Atoms 1-3 are termed root atoms for this molecule. For atoms  $i > 3$  are defined by ( $b_i, a_i, t_i$ ), where  $t$  are torsion angles. For example, atom 4 is presented by ( $b_4, a_4, t_4$ ). The dashed light gray shows the molecule after rotating the bond between Atom 2 and 3 ( $t_4$ ) for 180 degrees.

Because biomolecules are larger than small molecules, we developed a multi-layer definition of internal coordinates to avoid unrealistic motions produced by using classical internal bond-angle-torsion coordinates (Z-matrix). Classical definition uses a set of root

atoms to begin converting Cartesian coordinates to a Z-matrix, and the position of atom  $i$  depends on other atoms that are bonded in sequence (Figure 6.4)<sup>92</sup>. This creates artifacts that small rotations of some torsion angles close to the root atoms may result unrealistic motions on one end of a molecule as shown in Figure 6.4.

Therefore, we divide a molecule into  $N$  fragments, and use a multi-layer pattern to connect them. The fragments are chosen so that usually an  $\alpha$ -helix or  $\beta$ -sheet is treated as one or two fragments and a loop is divided into fragments with 3-4 residues in each fragment. Each fragment is presented as a small molecule shown in Figure 6.4, with a designated root dihedral angle in the fragment. The portions of a protein which moves in concert is grouped and a root fragment is chosen among the fragments in that group. For example, in the case of a typical kinase, the C-lobe and N-lobe can be treated as two groups of fragments. Every fragment in the same group are connected to the root fragment through the six pseudo degrees of freedom shown in Figure 6.5. Then the groups, if more than one groups exist in a molecule, are connected by linking the root dihedrals in the same way. In systems with multiple molecules, the molecules are then connected in the same way. The BAT coordinate-fragment-group-molecule-system framework defines the multi-layer internal coordinate.

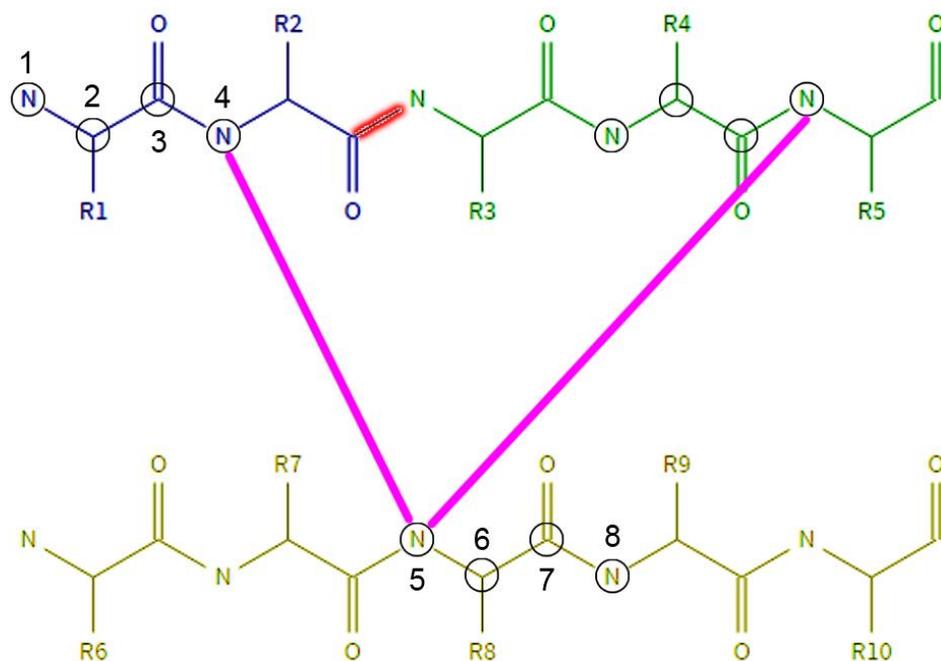


Figure 6.5. Two peptide chains divided into three fragments, in blue, green and yellow. The atoms, N, CA, C, N, in root dihedrals in the fragments are circled. Pseudo bonds connecting the fragments are rendered in purple. The peptide bond (C-N) connecting the blue and green fragments is rendered in red. In the internal coordinates, the peptide bond (in red) connecting blue and green fragments are removed. For this reason, artifacts may arise from the absence of the chemical bond between groups. Instead, the pseudo bonds are used to connect the fragments. For example, between blue and yellow fragments, the N-N (4-5) pseudo bond connects the two fragments, and along the pseudo bond, two pseudo angles, 3-4-5 and 4-5-6, and three pseudo dihedrals, 2-3-4-5, 3-4-5-6 and 4-5-6-7 are also included. These six pseudo degrees of freedom are used as the linker between fragments.

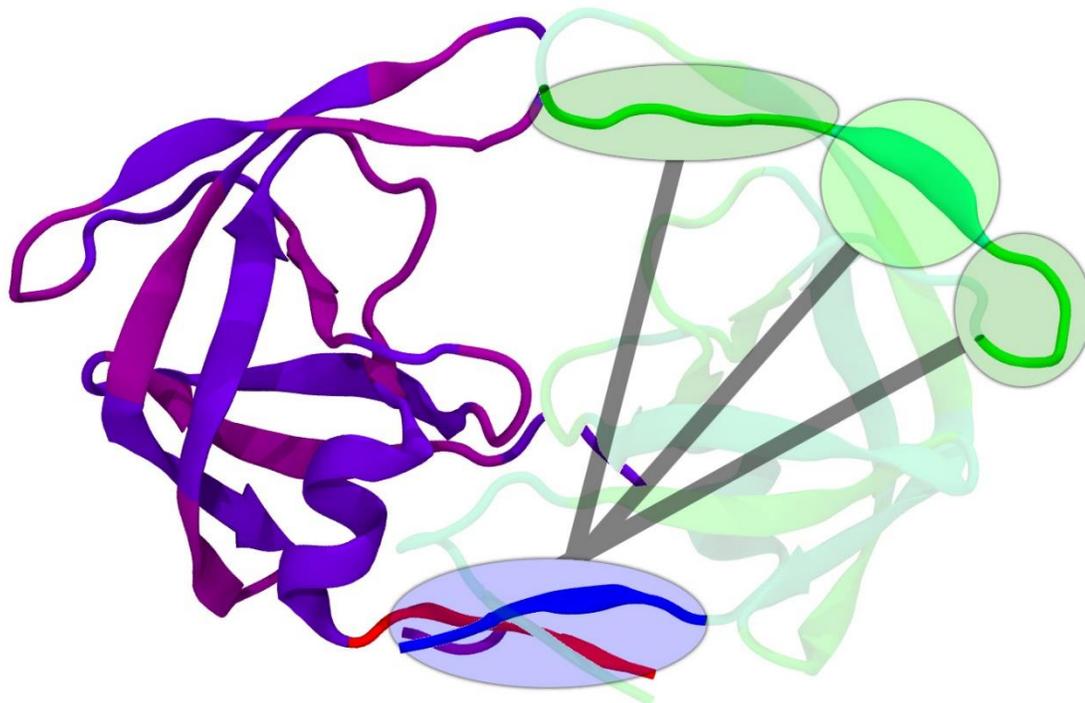


Figure 6.6. An example of fragmentation of HIVp. each monomer was divided into 21 fragments. The primary fragment,  $1_{\text{fragment}}$ , on monomer 1 is colored in blue and on monomer 2 is in red. The rest fragments on monomer 1 are in light and dark green and on monomer 2 are in light and dark magenta. The greenish fragments on monomer 1 are defined to the blue primary fragment, instead of mutually connected. Similarly the magentaish fragments are defined to the red primary fragment. Finally, the blue fragment on monomer 1 is chosen as the starting point of the internal coordinate, and the red fragment on monomer 2 is defined to it.

Here we use HIV-1 protease (HIVp) as an example. HIVp is a homodimer; therefore, we treat the two monomers in exactly the same way. The entire monomer is treated as one group, and this group is divided into 21 fragments, with the root fragment located on the end of the peptide chain. As illustrated in Figure 6.6, three fragments are circled, and they are connected to the root fragment on the same monomer through 6 pseudo degrees of freedom (see details in Suppl Info). The atoms in each fragment are defined as a small molecule in the BAT coordinate. Then the root fragments of the two monomers are connected through the six pseudo degrees of freedom.

Because a molecular system is divided into several fragments and each fragment is directly connected to the root fragment of its group by 6 pseudo degrees of freedom, atoms at the boundaries of two fragments in the same layer are not connected by physical bonds in our multi-layer internal coordinates (Figure 6.5). Therefore, it may produce unrealistic distortion in a principal component (PC) mode. One needs to assign fragments carefully and validate the PC motions beforehand, and we also include quick minimization in our algorithm to reduce the artifacts.

#### 6.2.2.2 Periodicity of Dihedral

Dihedral rotation is periodic. When a dihedral moves from  $179^\circ$  to  $-179^\circ$ , it literally moves  $2^\circ$ . However, the basic arithmetic operation used in computing a covariance matrix results in a large,  $179^\circ - (-179^\circ) = 358^\circ$ , rotation. Therefore, we use trigonometric functions in Eq. 6.1, 6.2 and 6.3 to operate average, addition and subtraction, respectively, when constructing the covariance matrix with internal coordinates. In this

way, one torsion angle is represented by exactly one element in the covariance matrix, in contrast to dihedral PCA<sup>83-88</sup>.

$$\text{Average: } \tan\left(\frac{a+b}{2}\right) = \frac{\sin(a)+\sin(b)}{\cos(a)+\cos(b)} \quad \text{eq. 6.1}$$

$$\text{Addition: } \tan(a+b) = \frac{\sin(a)\cos(b)+\sin(b)\cos(a)}{\cos(a)\cos(b)-\sin(a)\sin(b)} \quad \text{eq. 6.2}$$

$$\text{Subtract: } \tan(a-b) = \frac{\sin(a)\cos(b)-\sin(b)\cos(a)}{\cos(a)\cos(b)+\sin(a)\sin(b)} \quad \text{eq. 6.3}$$

### 6.2.2.3 Internal Coordinate Element Selection

As shown in Figures 6.4, multiple torsion angles such as  $t_8$  and  $t_9$  may be used to present a bond rotation. Some bond rotations, such as methyl rotation and rotating torsion angles in rigid rings, are not important conformation determinants. As a result, we only select torsion angles whose rotation is expected to be important in conformational changes. Peptide bonds in protein backbone are not considered flexible, either. In the internal PCA algorithm, only degrees of freedom in the internal coordinates will be selected into the covariance matrix, i.e. even some of the backbone dihedrals will be ignored based on the definition of internal coordinates, for instance, in Figure 6.5, the dihedrals along the bond at the boundary of two internally connected fragments. Also, all sets of 6 pseudo degrees of freedom used to connect fragments are automatically included in the covariance matrix. Therefore, the covariance matrix does not only contain torsion angle elements but also bond and angle elements.

#### 6.2.2.4 Calculation of Eigenvalues and Eigenvectors

As the classical way of PCA, the selected elements from the internal coordinates are used to construct the covariance matrix. Then the covariance matrix is diagonalized. The resulting eigenvalues and eigenvectors are automatically ranked by the eigenvalues. These are the corresponding PCs of the molecular system. PC modes are normalized before used in the conformational search algorithm.

#### 6.2.2.5 Conformational Search Algorithm

##### 6.2.2.5.1 Internal PC Mode Re-Ranking

PCs with greatest eigenvalues usually identify major motions of a molecule. However, for ligand-receptor motions, large relative motions of the two molecules may not appear in these PCs. To enhance the efficiency for ligand dissociation from the receptors, we re-rank the PCs by a PC mode ranking module. Along the positive and negative directions of the PC modes, the initial conformation of the complex system is distorted for the same extend, and then the PC modes are ranked from greatest to smallest by the distance the ligand travels with the protein/receptor aligned. In this way, the PC modes that move the ligand more significantly have higher priority in the conformational search algorithm.

##### 6.2.2.5.2 Systematic Search

PSIM starts the conformational search from the initial conformation. It first performs the conformational search according to the flowchart shown in Figure 6.3, and then continues the same procedures starting from the conformations yielded from distortions along each

internal PC mode. Until every branches of the systematic search reaches either path output or a dead end where no PC modes can be used, the algorithm ends.

#### 6.2.2.5.3 PC Mode Screening

In the conformational search starting from the initial conformation, both positive and negative directions of all PC modes will be used in conformational distortion. However, it is unnecessary to loop over all PC modes in the searches on subsequent branches. A rule to screen the modes is implemented as schemed in Figure 6.2. First, the mode, which has been used in the parent branch, regardless of the positive or negative direction, is discarded in the distortion. Secondly, the negative (positive) direction of any mode is also discarded after been used for certain number of steps. Any direction of any mode otherwise is considered in the distortion.

#### 6.2.2.5.4 Conformational Distortion

A stepsize is predefined to scale the eigenvectors as one step in the distortion. In each search branch, the algorithm distorts the current conformation ( $X$ ) by converting it to internal coordinate ( $Q=f(X)$ ) using the same multi-layer internal coordinate definition ( $f$ ), and adding the in-using eigenvector ( $dQ$ ) to the internal coordinate for a distorted internal coordinate ( $Q'=Q+dQ$ ), and finally converting it back to Cartesian coordinate ( $X'=f^{-1}(Q')$ ) using reported algorithm<sup>93</sup>. The distortion undergoes the same direction of the in-using mode and an acceptance test is evaluated at each step, until the acceptance test rejects the distorted conformation. New conformations are proceeded to the following modules shown in Figure 6.3.

#### 6.2.2.5.5 Acceptance Test

The acceptance test consists of three parts, 1) collision test, 2) bond stretching test and 3) bond angle vibration test. The collision test is a test that checks the distance of automatically predefined sets of atom pairs against scaled value of the sum of their vdW radii provided in the force field <sup>89</sup>. Any distorted distance lesser than the scaled (COLLISION\_PAIR\_SCALE) sum signals the rejection of the distorted conformation. This is a coarse approximation of the widely used LJ potential which governs the vdW interaction of atoms. The bond stretching and bond angle vibration are designed to ensure the behavior of those bonds and angles, which are not involved in the internal coordinate definition, are realistic to some extent. In the two tests, every bond and angle are calculated and compared to the equilibrium value, and a shrinking or extending greater than a fraction defined by COLLISION\_BOND\_RANGE or COLLISION\_ANGLE\_RANGE, the conformation is rejected.

#### 6.2.2.5.6 Output Trajectories

When the distorted conformation passes the acceptance test, and the accumulated number of steps reaches a predefined output length, a trajectory from the initial conformation to the current distorted conformation along the search path will be outputted.

#### 6.2.2.5.7 Motion Test

An RMSD based motion test is designed to ensure movement of ligand and reject denature of the protein. It is initiated when the accumulated number of steps reaches to a predefined value. In the motion test, three RMSD calculations are performed. First, the

initial and current protein (host) conformations are aligned, and then the RMSD of the two proteins (hosts) is compared against the corresponding RMSD cutoff (RMSD\_NODE\_PROTEIN\_RESTRAIN). If the protein RMSD goes above the cutoff, which means the protein deforms too much, the conformation is rejected. Second, the conformations of protein at previous node and the current node are aligned, and then the RMSD of the two ligands is compared against the corresponding RMSD cutoff (RMSD\_NODE\_LIGAND\_SHIFT). If the RMSD is smaller, in the ligand case, than the cutoff, which means the ligand does not move significantly in the current node, the conformation is rejected. The third test is designed so that a back and forth scenario of the ligand, which can be accepted in the second test, is ruled out. Again, the initial and current conformations of the protein (host) is aligned, and the RMSD of the ligands at the two conformations is compared against another node-linear-accumulative RMSD cutoff (RMSD\_CUM\_LIGAND\_SHIFT  $\times$  node number). If the RMSD is smaller than the current cutoff, which means the ligand does not move as expected, again, the conformation is rejected.

#### 6.2.2.5.8 Conformation Correction

This module is design to reduce the artifacts induced by further distortion of the system. It is performed only after the motion test. An mild minimization is performed with only bond, angle, and improper terms for predefined number of steps, to eliminate the accumulated local artifacts from fragmentation of the molecule but still hold the overall structure of the current conformation.

### 6.2.3 Molecular Systems

The subsection includes simulation configurations such as MD setup and parameters used in PSIM and umbrella sampling. All MD simulations were run with the Amber 14 package<sup>94</sup> and a 2 fs time step. The SHAKE algorithm was used to constrain hydrogen atoms during the simulations. Unless otherwise stated, all systems were heated gradually and equilibrated before the production runs. Parameters used in the PSIM search are detailed in Table 6.1.

System		Alanine Dipeptide	Cyclodextrin 2-naphthyl ethanol	Cryptophane Me <sub>4</sub> N <sup>+</sup>	TRPS G3P
Output Step Number		100	4000	3000	2000
Motion Test Step Number		200	400	200	200
Step Size		0.05	0.02	0.05	0.05
Boost		1	1	20	2
Acceptance Test	Pair Distance	0.8	0.65	0.5	0.5
	Bond Stretch	0.2	0.2	0.3	0.4
	Angle Stretch	0.2	0.2	0.3	0.5

Table 6.1. The details of parameters used in the PSIM search for all systems.

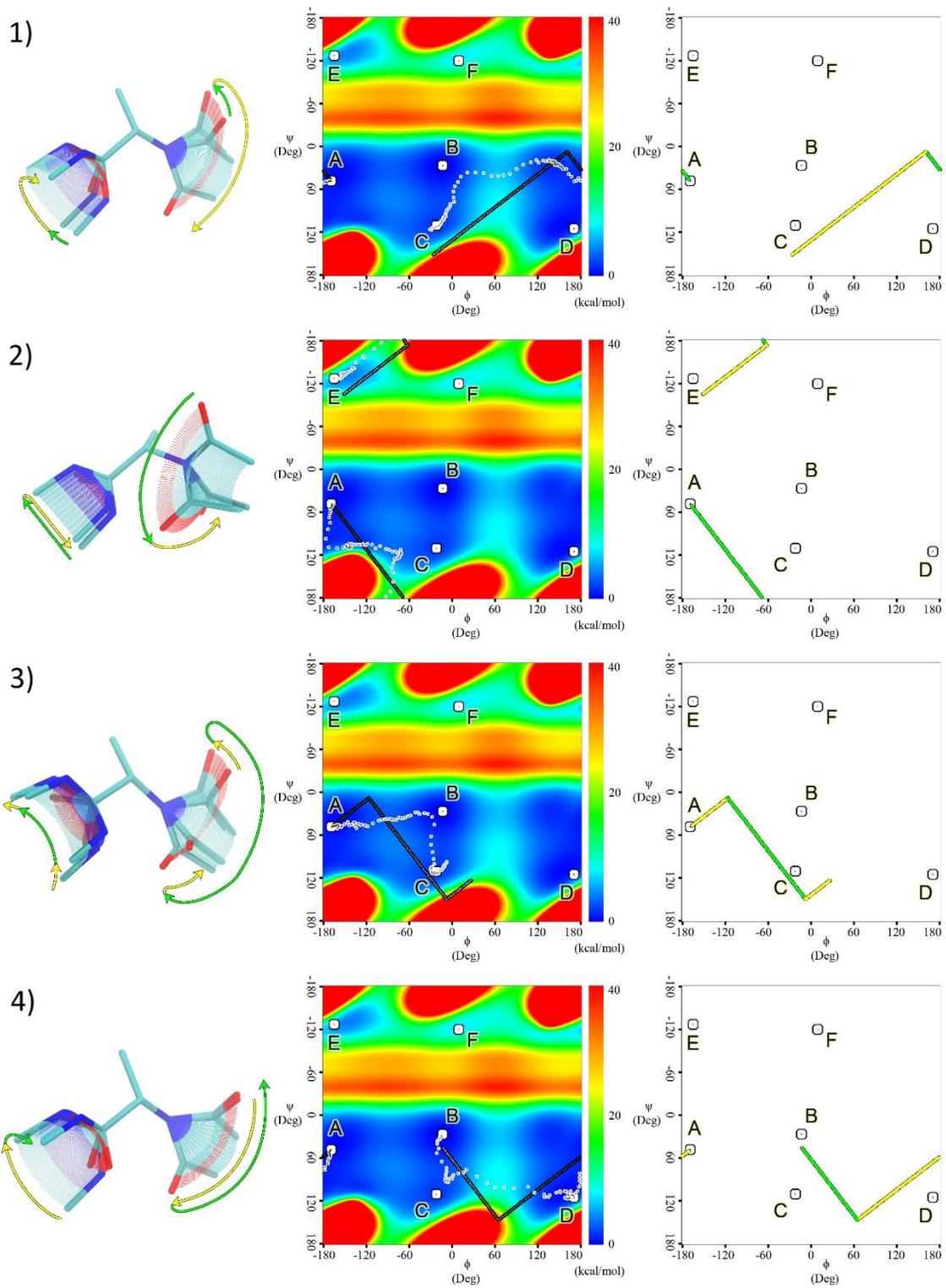
#### 6.2.3.1 Alanine Dipeptide

The alanine dipeptide was parameterized with AMBER FF14SB force field<sup>95</sup>. The molecule was heated from 200 K to 298 K and then followed by a 10 ns molecular dynamics simulation in implicit water by generalized Born (GB) model<sup>96</sup> saved at every 1 ps with AMBER 14 package<sup>94</sup>. We obtained the Internal PCA modes by selecting the  $\phi$  and  $\psi$  angles in the molecule using all 10000 frames from MD. Because this molecule is small enough, the whole molecule was treated as one fragment. The conformational

search was started from minimum A (Figure 6.7) by minimizing a randomly picked MD conformation. The motion test was disabled so that the selection of a ligand is not necessary. The stepsize was set to 0.05. The atom pair distance cutoff was set to 0.8. The angle and bond stretch range was set to 0.2. Output trajectory length was set to 100 steps. Both PC modes were used in the conformational search. The trajectories thus obtained were post processed by BFGS minimization with GB for 50 steps.

PES of alanine dipeptide in aqueous solution was obtained by rotating the  $\phi$  and  $\psi$  angles stepwise ( $0.5^\circ$ ) and calculating the energy of resulting conformer with GB model <sup>96</sup>. The original and minimized transition pathways obtained from PSIM were projected onto the PES (Figure 6.7). Corresponding conformational transitions were also illustrated.

Figure 6.7. The four transition path found by PSIM. The search started from Minimum A, and led the transitions to different minima starting from four directions including the positive and negative of PC mode 1 and 2. The left views of each path show the conformational change of the original path using PC mode 1 (Green) and 2 (Yellow). The middle views of each path are the original path (Black bordered dots) and the minimized path (White bordered dots) on PES of alanine dipeptide. The right views show the search path on PES with PC mode 1 in green and PC mode 2 in yellow.



### 6.2.3.2 HIV Protease

A 250 ns MD simulation saved at every 100 ps of HIVp parameterized with Amber FF99SB force field was used in internal PCA calculation. Each of the two identical monomers was divided into exactly the same 21 fragments according to the secondary structures followed by an error and trial process and refinement (Figure 6.6). Only  $\phi$  and  $\psi$  angles on the backbone were taken into consideration. First three low mode motions were produced. For comparison, a dihedral PCA based on classical internal coordinate definition and an  $\alpha$ -carbon PCA were performed and the first three low mode motions were produced respectively. The RMSFs of internal PCA and classical dihedral PCA were calculated against  $\alpha$ -carbon PCA.

### 6.2.3.3 Beta-Cyclodextrin-2-Naphthyl Ethanol

The structure of  $\beta$ -cyclodextrin was taken from Cambridge Crystallographic Data Centre (CCDC) (PDB ID: WEWTOJ) and the structure of 2-naphthylethanol was drawn manually with Vega ZZ<sup>97</sup>. The ligand was docked into the binding site with Autodock 4.2<sup>98</sup>. The initial structure from docking was solvated with 1736 TIP3P<sup>99</sup> water molecules and then parameterized with GAFF force field<sup>100</sup>. MD simulations were performed with AMBER 14 package<sup>94</sup>. A water equilibrium was run at 298K in NPT ensemble for 1 ns. Then the system was heated up to 298K and a production run of total 3  $\mu$ s was run at 298K in NPT. 15 dissociation events were identified from the trajectory. A 500 ns bound state sub trajectory at the initial stage of whole MD simulation was resaved for internal PCA calculation. In internal PCA, the  $\beta$ -cyclodextrin was treated as a whole

fragment. All 72 PC modes obtained from internal PCA were used in the conformational search. The stepsize used in this example is 0.02, different from all other examples. We collected 20 dissociation pathways from PSIM in 3.1 h.

For the 15 dissociation events from MD simulation, we evaluated the host molecule conformation. We first smoothed the trajectories by averaging forward 100 and backward 100 frames on the concurrent frame throughout the whole trajectory. Then along the smoothed trajectory, we calculated the average planes of each glucose by selecting the atom on the ring and the average plane of the entire host molecule by selecting all 147 atoms. The average planes were computed with multiple regression theory. We calculated the normal vector of the planes, and computed the 7 angles between the 7 normal vectors of the glucose rings and the normal vector of the host molecule. These 7 angles were used as fingerprints of the  $\beta$ -cyclodextrin conformation. We analyzed the conformational changes of the host molecule by comparing the trajectories with the conformational fingerprints. For the 20 dissociation trajectories from PSIM, we calculated the same fingerprints without smoothing the trajectory.

#### 6.2.3.4 Cryptophane-Tetramethyl Ammonium

The bound structure of cryptophane- $\text{Me}_4\text{N}^+$  complex was taken from other works<sup>101</sup>, and then parameterized with CHARMM22 force field<sup>102-104</sup>. MD simulation was performed in the similar procedure as alanine dipeptide, except that the dielectric constant in GB<sup>96</sup> of the solvent was set to 8.42 corresponding to the value of tetrachloroethane, and the production run was 20 ns. In the MD run, no dissociation event happened. We performed

Internal PCA by dividing the cryptophane into six fragments and leaving the ligand as one. In the conformational search, the MAXSTEP was set to 3000. A STEPSIZE scaling factor was set to 0.05. vdW radii scaling factor was set to 0.5. All 63 modes were used for conformational search. Four pathways were found by the search in 8.3 h. The trajectories were then post-processed by BFGS minimizer for 100 steps to remove unphysical clashes, bond stretching and angle bending. The first trajectory was picked as example for illustration purpose. For comparison, one dissociation pathway was also created by pulling the ligand along one dissociation window for 200 steps. At each step, the ligand was pulled for 0.1 Å and the whole structure was minimized by fixing the distance between the centers of masses of the two molecules.

Umbrella sampling of the search path was performed along the reaction coordinates represented by distance between two centers of mass. Because the trajectory from the search algorithm may be sparse for umbrella sampling, interpolation of five frames between any two existing frames was done by using the corresponding PCA mode in the search. From the interpolated trajectory, conformations that are closest to every 0.1 Å center of mass distance from 0 Å to 19.8 Å were picked from the dissociation region, as initial conformations for 0.5 ns MD runs with the center of mass distance fixed. Umbrella sampling was also performed on the dissociation pathway generated by pulling the ligand. In both cases, the distance was saved every 50 ps. The umbrella force constant was 100 kcal/mol/Å<sup>2</sup>. Umbrella sampling data was process by using WHAM<sup>105</sup>.

### 6.2.3.5 Tryptophan Synthase

We resaved the one  $\alpha$ -subunit with G3P from a 10 ns bound state tryptophan synthase-G3P complex MD trajectory parameterized with AMBER FF99SB force field with one copy of both  $\alpha$ -subunit and  $\beta$ -subunit. Then we performed internal PCA on the G3P-tryptophan synthase  $\alpha$ -subunit complex (Figure 6.8). The tryptophan was coincidentally divided into 42 fragments as the procedure described for HIVp. The dihedrals in residue 174 to 193 210 to 217 56 to 67 232 to 242, including sidechain dihedrals were considered in the dihedral covariance matrix, while the rest part of the protein were kept fixed. All dihedrals in the ligand were included except those filtered out as described above. All resulting 254 PCA modes were used for conformational search from a randomly picked bound state from the MD. The MAXSTEP was set to 2000 step, with the STEPSIZE equaled to 0.05 and the vdW radius scaling factor COLLISION\_PAIR\_SCALE equaled to 0.5. Six consecutive search runs were performed until the ligand escaped from the binding site. From each run the resulting path trajectories were examined by the user to select the most plausible trajectory for continuing the next run. In this case the criteria included 1) not significantly deformed protein, 2) more open space for ligand to dissociate, or more specifically, more open Loop 6, 3) no tumbling motion of the ligand in the binding site at the initial stage and 4) ligand not going to the indole tunnel to the  $\beta$ -subunit.

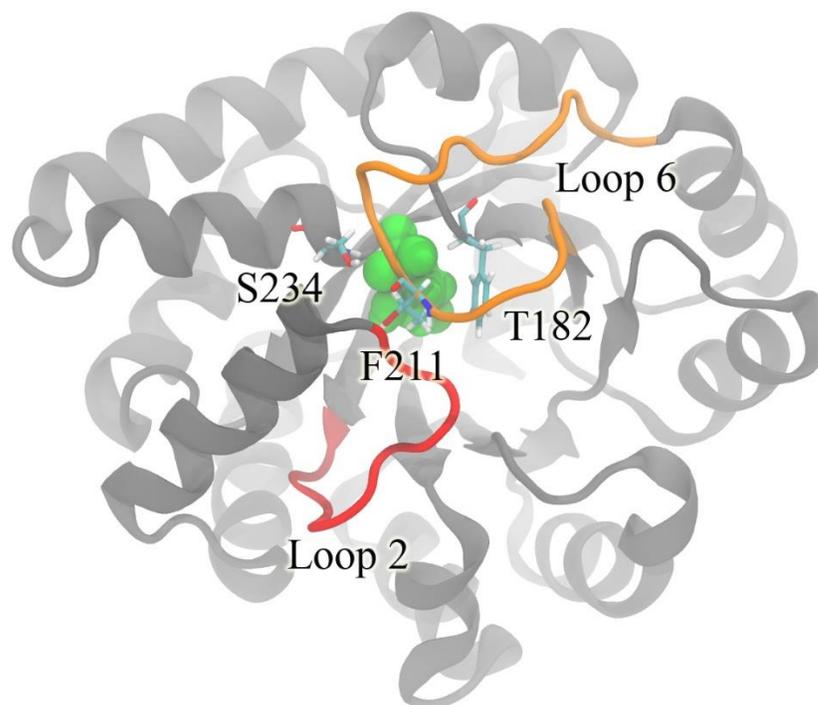


Figure 6.8. Structure of the  $\alpha$ -subunit of tryptophan synthase. Loop 6 and Loop 2 are shown in orange and red. Bound state G3P is shown in green.

### 6.3 Results and Discussion

In the following section, we first illustrate the basic idea of the mechanism of the PSIM method with a typical toy model molecule, alanine dipeptide. Then the performance of multi-layer internal coordinate definition is evaluated by using HIVp as a model system in comparison with  $\alpha$ -carbon PCA and classic internal coordinate definition. The capacity of identifying dissociation pathways is demonstrated with cryptophane- $\text{Me}_4\text{N}^+$  complex which has a slow dissociation kinetics. To show that the dissociation pathway from P is reasonable, the search trajectories are compared to MD conformations with  $\beta$ -Cyclodextrin-2-naphthylethanol as the model system because this system has a fast

kinetics and MD is able to produce multiple dissociation trajectories for it. Finally this method is applied to tryptophan synthase-G3P system to show the full capacity of this method in large scale molecules.

### 6.3.1 Performance of the Multi-Layer-Internal Coordinate Definition

As discussed above, the classical internal coordinate definition leads to artifacts in the PC motion produced by ordinary vector calculation. The multi-layer internal coordinate definition, on the other hand, avoids the accumulated dependence of dihedral angles by breaking the dependence and reconnecting the further dihedrals to the near region to the primary root. In this ways, the artifact in the overall motion has been eliminated. A comparison of  $\alpha$ -carbon PC motion and multi-layer internal definition based internal PC motion is shown in a) and b) Figure 6.9. By observing the corrected internal PC motions, it successfully reproduced the overall motion of the protein backbones, and moreover, it included the side chain details instead of only the backbone. Figure 6.10 shows the RMSF of the first internal PCA modes from multi-layer internal coordinate definition and standard internal coordinate definition in reference to the first mode of  $\alpha$ -carbon PCA. This quantifies the performance of the multi-layer internal coordinate and the artifacts from classical internal coordinate. It can be noticed that even in the multi-layer definition, residue 47, 48 and 49 have significant deviation from the reference (In Figure 6.10). This is because the two monomers share the same multi-layer internal coordinate pattern while the motions in MD are not exactly symmetric. In the flap, both ends of the highly flexible loop are connected to the relatively rigid  $\beta$ -sheet, which makes the fragment motion dependence have duality in this region and results in the deviation.

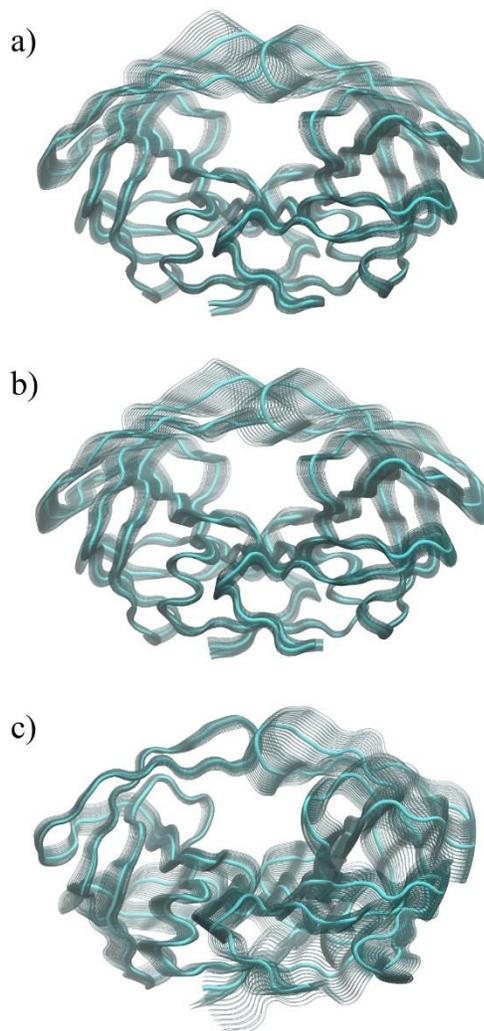


Figure 6.9. A comparison between a)  $\alpha$ -carbon PCA scaled by 2.67, b) dihedral PCA by using multi-level internal coordinate definition scaled by 1.0 and c) dihedral PCA by using standard linear type internal coordinate definition of HIVp scaled by 0.2. In a) and b) the motions of the HIVp have very high similarity, which the two monomers have scissoring motion with the two flaps opening and closing and the bottom part of the HIVp mostly is stationary. However, in c), the left monomer is stationary because the primary root of the internal coordinate is on the left monomer while the right monomer has overall translational and rotational motion because it is defined through the mobile loops of left monomer. The overall asymmetric behavior is the artifact from the accumulated dependence of dihedrals in the internal coordinate.

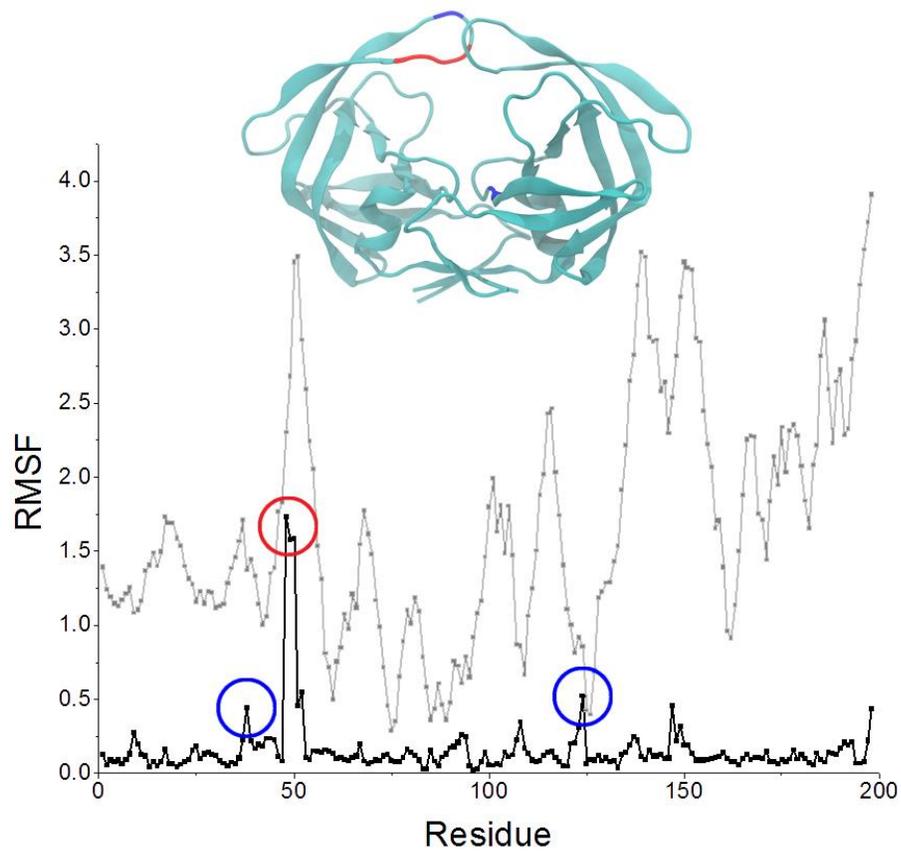


Figure 6.10. The RMSF of first dihedral PCA mode from multi-level internal coordinate definition (black) and standard internal coordinate definition (gray) in reference to the first standard  $\alpha$ -carbon PCA mode. Residues with RMSF greater than  $0.5 \text{ \AA}$  are in blue, and that greater than  $1.5$  is in red, in the top structure and circled in the plot.

By limiting the view to each fragment, the motion is continuous and rigorous. Artifacts in the connection area (bond and angles) between groups can be observed when the motion is exaggerated. This is because there is no real bond or angles defined in the internal coordinates in such multi-layer definition. To minimize this artifact in the algorithm side, it performs minimization in every node test. In the user side, an error and trial approach is necessary in the fragmentation stage. Typically this is done by fragmenting the protein according to secondary structure, conducting internal PCA calculation, visualizing the motions of first few internal PCA modes with greatest eigenvalues and then modifying the root atoms and the fragmentation. The root atoms are recommended to be located on the  $\alpha$ -carbon atoms to minimize the artifacts. One  $\alpha$ -helix or  $\beta$ -sheet may be divided into multiple fragments depending on the length and flexibility and the loops are more often divided into more fragments with two to three residues in one fragment because the loops are more flexible.

### 6.3.2 Pathway of Conformation Transitions for Alanine Dipeptide

Here we first use a model system, alanine dipeptide, to illustrate how the PSIM method searches for new conformations and conformational transition pathways using internal PC modes obtained by MD simulations. We obtained four transition pathways from the PSIM method, starting along four directions, i.e. the positive and negative of the two PC modes. After post-process by mild minimization, more realistic transitions were obtained. The transition pathways and minimized pathways are shown in Figure 6.7. The details of searching modes and number of steps are summarized in Table 6.2.

Path	Mode	Direction	Number of Steps
1	1	+	18
	2	-	82
2	1	-	61
	2	-	39
3	2	+	23
	1	-	62
	2	+	15
4	2	-	56
	1	+	44

Table 6.2. The details of transition pathways. Searching directions are listed in order from top to bottom. Mode in use, its direction and number of steps are listed.

For example in 1) in Figure 6.7, by distorting along Mode 1 positive direction for 18 steps, the search hit into the green (15 kcal/mol) region. The acceptance test rejected further distortion. Therefore the search shifted to Mode 2 negative direction, because Mode 2 positive direction would also hit into the same energy barrier. After distorting for another 82 steps, the accumulative step number reached 100, which was predefined in the input file, and the search stopped near minimum C to output the transition path. The original pathway was straight on the PES because the search goes along exactly one PC mode at a time. It was able to cross the energy barrier in cyan (~7-8 kcal/mol) on its way to minimum C. After mild minimization, the pathway became more realistic; it starts from A, searching for a low energy route to regions near B, and then ends at C. Similarly, in 2), the search started from A along negative of Mode 1. It turns to negative of Mode 1 when the search hits into the high energy barrier in red (40+ kcal/mol), and ends up in E.

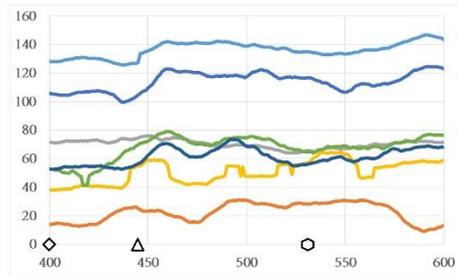
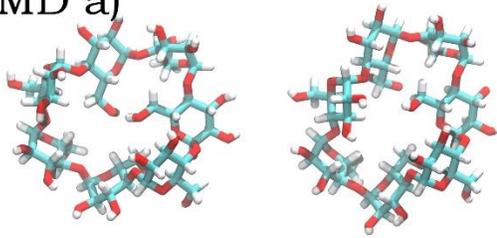
It is noticeable by comparing Pathway 1) and 2), that the acceptance does not rigorously correspond to an energy threshold. In 1), the searching mode shifts from Mode 1 to Mode 2 before the search reaches the green region, while in 2), the searching mode shifts after it hits slightly into the red region. This is because the acceptance is more based on geometrical cutoff rather than energetic cutoff.

### 6.3.3 Comparison with MD Conformations

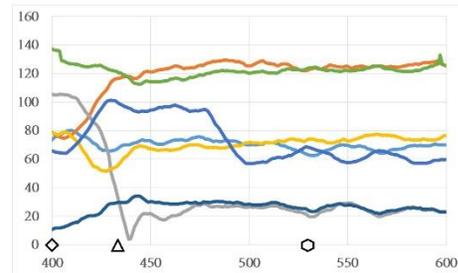
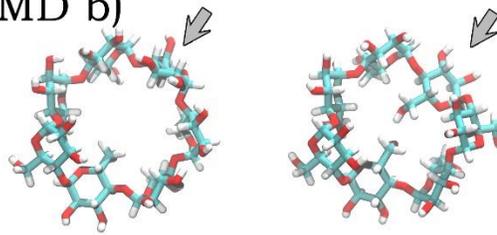
From the fingerprint angles, we classified the dissociation events into three groups, a) no glucose ring flipping, b) one glucose ring flipping and c) multiple glucose ring flipping. Representative dissociation trajectories from MD and PSIM and their plots of fingerprint angles are shown in Figure 6.11.

Figure 6.11. The comparison between MD dissociation events and dissociation events from PSIM. Trajectories are classified by number of glucose ring flipping in  $\beta$ -cyclodextrin, where a) is no significant ring flipping, b) is one ring flipping and c) is multiple ring flipping. The bound and free conformations of  $\beta$ -cyclodextrin are shown on the left and in the middle of the figure. The plots of 7  $\beta$ -cyclodextrin conformation fingerprint angles along trajectories are shown on the right. Corresponding glucose ring are labeled on the conformations using the same color as in the plots. The bound state conformation (rectangle), free state conformation (triangle) and moment of dissociation (hexagon) are indicated on the plots.

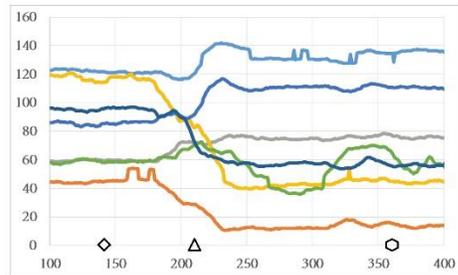
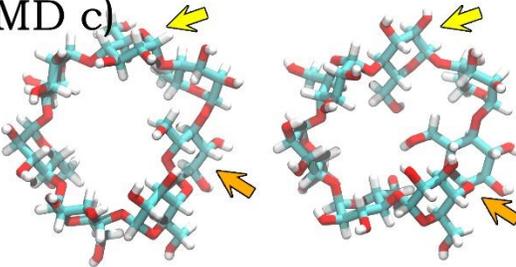
MD a)



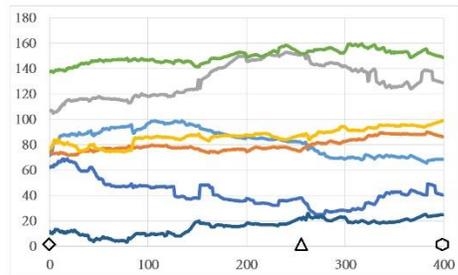
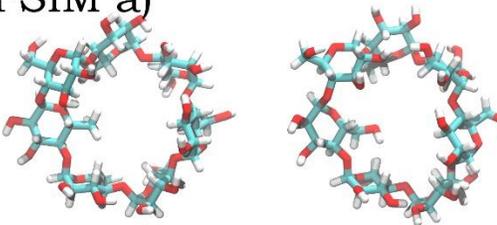
MD b)



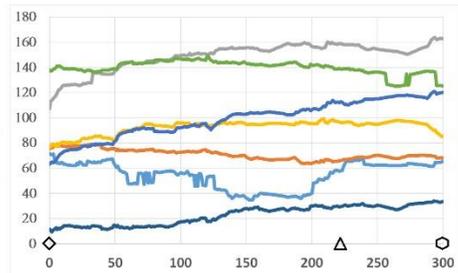
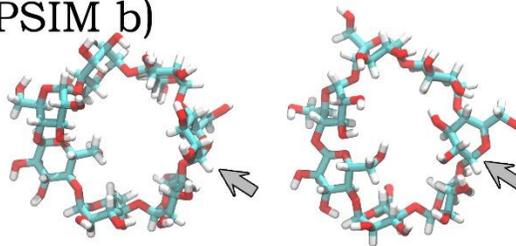
MD c)



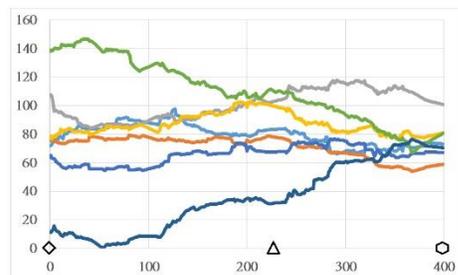
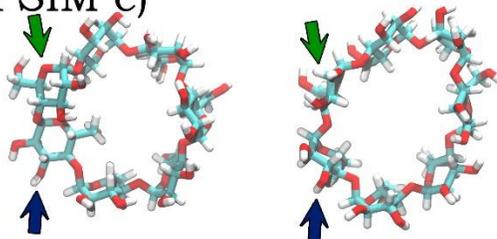
PSIM a)



PSIM b)



PSIM c)



PSIM is able to find dissociation pathways with the glucose rings moving physically as MD does in all three groups, although PSIM is unable to recreate one exactly identical dissociation pathway as the MD simulation. The similarity between PSIM and MD trajectories is because that the motion of molecules in PSIM is under the government of internal principal motions from the MD simulations. In Group a), both MD and PSIM can find dissociation pathways in which the glucose rings do not flip significantly. In addition, in both trajectories, the fingerprint angles change slightly and the cavity of the host molecule shrinks as the ligand dissociates. In Group b), the MD and PSIM starts from exactly the same conformation, where one glucose ring is parallel to the regression plane of the entire molecule while other six are perpendicular to it. In both MD and PSIM, the same glucose ring (arrowed in gray) flips in the dissociation events. In Group c), both MD and PSIM find dissociation pathways, in which multiple glucose rings flip. More interestingly, in the PSIM trajectory, all glucose rings flip to the pose perpendicular to the regression plane of the entire molecule.

The PSIM evolves the trajectories in different way from the MD. In PSIM, the conformations of molecules are moved stepwise, based on internal PC motions and the acceptance test, which is fundamentally different from MD in which the motions are guided by classical mechanics. Therefore, the fingerprint plots from PSIM and MD have distinct behaviors. In MD plots, the fingerprint angles still vibrate significantly even after the trajectories are smoothed. In PSIM, there are local vibrations but the fingerprint angles move generally towards one defined direction without many back and forth motions. However, the conformational transitions are still correlated with the dissociation

event of the ligand. Therefore, the dissociation trajectories from PSIM are useful guidance of the physically meaningful motions of the molecular systems.

#### 6.3.4 Tetramethyl Ammonium Unbinding from Cryptophane-E. Hist: Finding the Pathways and the Corresponding Free Energy Barriers

The cryptophane- $\text{ME}_4\text{N}^+$  complex has a slow dissociation rate constant ( $k_{\text{off}} = 4.81 \times 10^{-2}$  1/s; residence time = 14 s)<sup>106</sup> which is impractical to use classical MD simulations to sample its dissociation pathways<sup>25-28</sup>. Using 63 PC modes constructed from a 20 ns MD simulation for the bound complex, PSIM successfully found 4 dissociation pathways in 8.3 h wall time. The trajectories are similar to each other because of the symmetry of the host molecule. One representative dissociation pathway is shown in Figure 6.12. The two methoxybenzyl groups serve as a gate for ligand binding/unbinding, and the gate is closed when the cryptophane host are in a free state and bound to a guest such as tetramethylammonium. Our search found that the methoxy groups need to rotate away to open the gate during ligand dissociation processes Figure 6.13. We also compare one pathway that PSIM found with one by the hopping minima method and one by simply pulling tetramethylammonium from the pocket (Figure 6.12). All three methods show the gate opening for the ligand to pass through the window. Different from the pulling test that tetramethylammonium is pulled directly toward the solvent, the pathways had tetramethylammonium that kept contacts with the surface of the host molecule. This is more reasonable than the pulling test because tetramethylammonium still gain intermolecular attractions before completely leaving cryptophane. Similar association

pathways can be found using the hopping minima method, which connects multiple local energy minima using normal modes to build the association pathways.

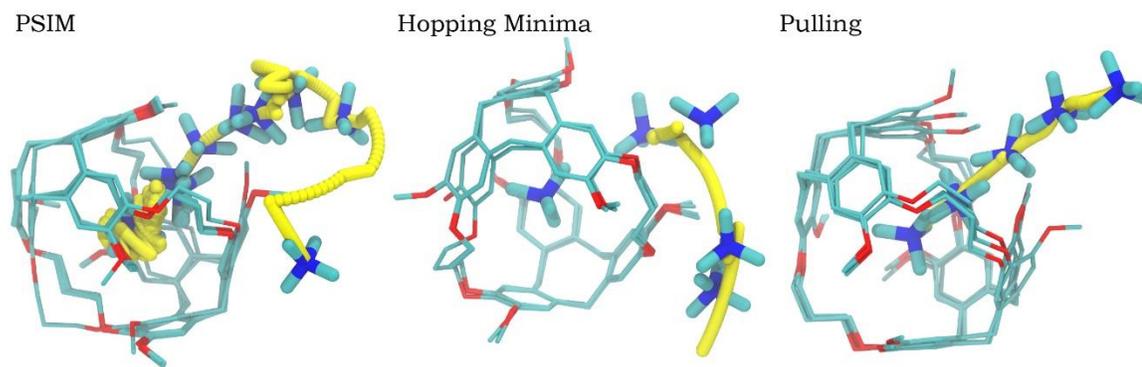


Figure 6.12. The original dissociation pathways found by PSIM, Hopping Minima, and simply pulling the ligand from one direction. The dissociation path is shown in yellow.

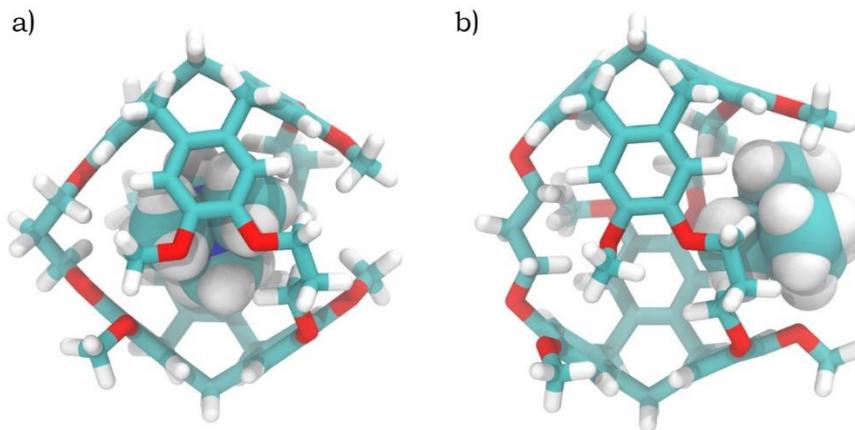


Figure 6.13. The post processed snapshots of crptophane-tetramethyl ammonium in the bound state (a) and the moment when the ligand cross the dissociation window (b). The ligand is shown in vdW representation.

To reveal the free energy barrier for ligand unbinding, umbrella sampling, together with the weighted histogram analysis method (WHAM), were used to plot a potential of mean

force (PMF) profile (Figure 6.14). The initial conformations were chosen based on a dissociation pathway found by PSIM, as detailed in Method. The energy barrier of dissociation  $\Delta G_{d,calc}^\ddagger$  is 16.8 kcal/mol, and the association barrier  $\Delta G_{a,calc}^\ddagger$  is 9.1 kcal/mol. The binding affinity computed from this pathway yields -7.8 kcal/mol, which is in good agreement with experimental measurement  $\Delta G^\circ = -7.3$  kcal/mol<sup>106</sup>. For protein-ligand complex systems, significantly more dissociation pathways are anticipated to find. To thoroughly investigate the barriers which few conformations may be available for computing a free energy profile, methods other than umbrella sampling may need to be utilized.

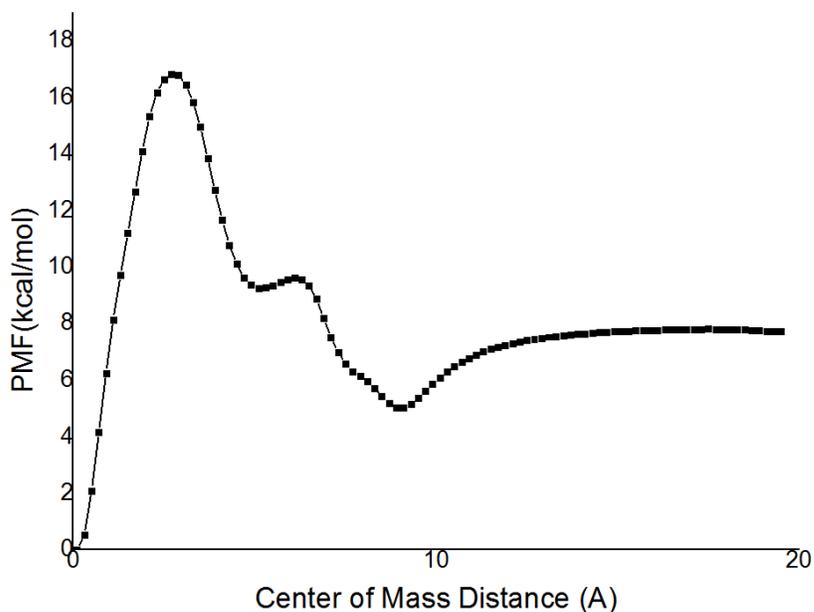


Figure 6.14. This is the free energy plot of the dissociation pathways constructed based on the trajectory from PSIM.

### 6.3.5 Application to Protein-Ligand System

This example is to show the ability of PSIM algorithm to find the dissociation pathway in protein systems. G3P is the byproduct of the chemical reaction happening in the  $\alpha$ -subunit of tryptophan synthase which consists of 266 residues. After the chemical reaction, the major product indole undergoes the transportation through the  $\alpha$ - $\beta$  tunnel into the  $\beta$ -subunit while the G3P is no necessary for further reaction and released into the cell by the opening of Loop 6.

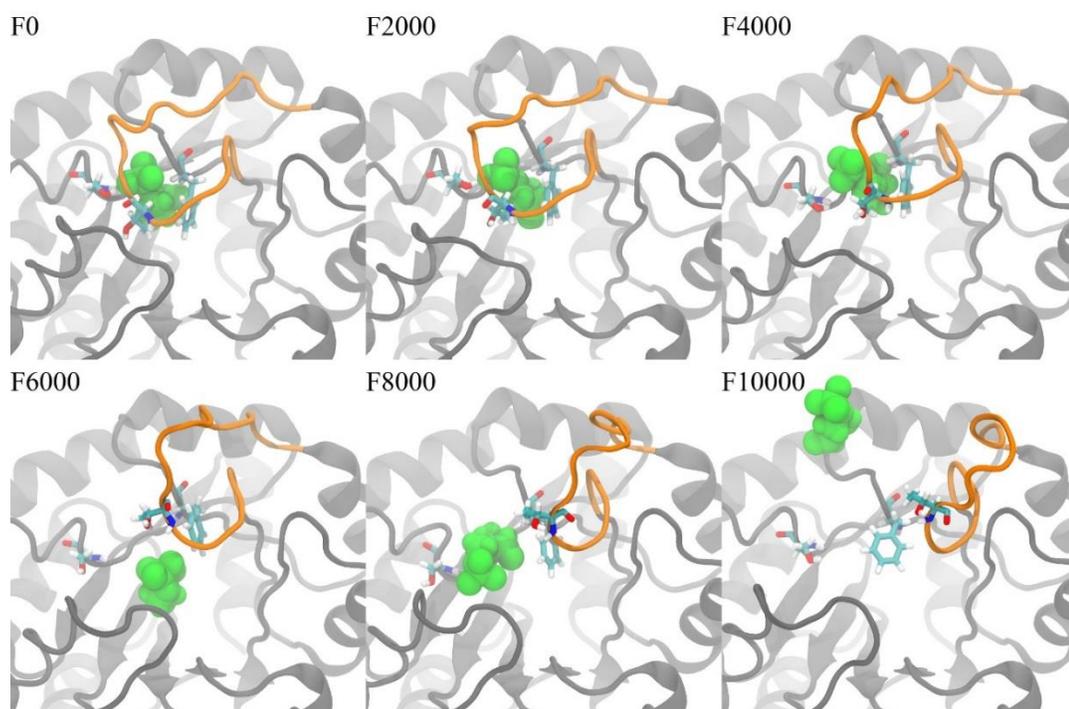


Figure 6.15. Tryptophan  $\alpha$ -subunit-G3P dissociation pathway found by the Internal PCA Search method. The protein is shown in gray with Loop 6 highlighted in orange. The ligand is shown in green. Important residues shown in Figure 6.8 are also shown in this figure. In the process of the dissociation path found by the search, the protein maintains its native conformation except that the Loop 6 moves away from the native state and opens the dissociation pathway for the G3P. The G3P escapes from the space originally hindered by Loop 6.

In the 10 ns MD simulation of tryptophan-synthase-G3P complex, no major conformational change of protein was observed and the ligand remains in the binding site stably without any observable rotation or translation. We applied the conformational search method on the  $\alpha$ -subunit with G3P to illustrate its ability of finding the dissociation pathway. After six continuous runs by starting from selected conformation from the previous run, the G3P was freed from the binding site of  $\alpha$ -subunit of tryptophan synthase. In the process, the protein maintained its native state because first non-relevant portion of the protein was kept fixed in the conformational search and second the RMSD cutoff of the protein disallowed any huge deformation.

The dissociation pathway of G3P is shown in Figure 6.15. Loop 6 moves away from its native state towards the  $\beta$ -subunit, and opens up the dissociation pathway for the G3P with the co-operation of several other  $\alpha$ -helices and loops near the binding site. A closer look reveals that first the T182, which serves as a wedge in the binding site, is taken out with Loop 6, and the movement of residue 180 to 187 on Loop 6 frees some space in the binding site, so that the hydrophobic interaction between the G3P and F211, hydrogen bonding between the G3P and S234 are lost. As the confinement in the binding site gets looser and looser, the G3P becomes more and more disordered with significant tumbling in the binding site. After the Loop 6 fully opens, the G3P stays in the binding site for a short period of time, because of the existence of hydrogen bonding between G3P and the backbone of F211. When the G3P explores the wide-open binding site and breaks the last hydrogen bond, it completely dissociates from the  $\alpha$ -subunit.

This method generates the trajectories in which the ligand moves the most from the very beginning of the search and the relevance of the trajectories generated as time goes decreases generally. Because the possible combinations of the internal PCA modes are innumerable, it is unnecessary to explore all combinations of the modes. It requires the user to examine the resulting trajectories while the search is still ongoing, and shut it down when reasonable trajectories have been found. In this manner, the computational time for the six runs of tryptophan synthase is counted to be 17.05 h for the effective runs, and 146.4 h for all examined trajectories.

#### 6.4 General Comments

This method is a PCA mode distortion based conformational search method and shares some similarity with the normal mode based methods <sup>76-79</sup>. PSIM method uses an orthogonal set of searching directions which characters the intrinsic motions of the system although this set of directions is from PCA mode instead of normal mode. New conformations are also generated by distorting some given conformation assuming that the intrinsic motions lead the system to new physical portions of the conformational space. However there are also several noticeable differences that distinguishes PSIM from the other methods. First this method does not depend on full minimization and yields continuous non-minimized trajectories of paths. This gives it two advantages, i.e. it avoids the computationally expensive minimization routines and thus gains efficiency, and it readily gives the well-connected paths of the dissociation instead of discrete minima to be connected by methods like Markov State Model <sup>107, 108</sup> or Hopping minima <sup>101</sup>. The second

difference is this method is designed for identification of possible dissociation pathways only, rather than a general purpose conformational search methods.

## 6.5 Conclusions

In this work, we presented an internal PCA with multiple-layer internal coordinate definition. Internal PC motions that are comparable to Cartesian PCA are reproduced by using this method. We also proposed a novel PSIM method based on distortion along internal PCA modes and illustrated its capacity of finding dissociation pathways that are physical meaningful with several examples. This method avoids frequent energy evaluation and energy minimization by using a collision test which can be viewed as a simplified potential, and thus gains efficiency in protein-sized complex systems. We discussed practical aspects of the PSIM methods by application to host-guest systems and protein-ligand system.

## 6.6 References

1. Balaraman, G. S., Park, I.-H., Jain, A., and Vaidehi, N. (2011) Folding of Small Proteins Using Constrained Molecular Dynamics. *Journal of Physical Chemistry B* **115**, 7588-7596
2. Lane, T. J., Shukla, D., Beauchamp, K. A., and Pande, V. S. (2013) To Milliseconds and Beyond: Challenges in the Simulation of Protein Folding. *Current Opinion in Structural Biology* **23**, 58-65
3. Best, R. B., Hummer, G., and Eaton, W. A. (2013) Native Contacts Determine Protein Folding Mechanisms in Atomistic Simulations. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 17874-17879
4. Sulkowska, J. I., Noel, J. K., and Onuchic, J. N. (2013) Energy Landscape of Knotted Protein Folding. *Biophysical Journal* **104**, 370A-370A
5. Gelman, H., and Gruebele, M. (2014) Fast Protein Folding Kinetics. *Quarterly Reviews of Biophysics* **47**, 95-142
6. Gnanakaran, S., Nymeyer, H., Portman, J., Sanbonmatsu, K. Y., and Garcia, A. E. (2003) Peptide Folding Simulations. *Current Opinion in Structural Biology* **13**, 168-174
7. Klenin, K., Strode, B., Wales, D. J., and Wenzel, W. (2011) Modelling Proteins: Conformational Sampling and Reconstruction of Folding Kinetics. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1814**, 977-1000
8. Wedberg, R., Abildskov, J., and Peters, G. H. (2012) Protein Dynamics in Organic Media at Varying Water Activity Studied by Molecular Dynamics Simulation. *Journal of Physical Chemistry B* **116**, 2575-2585

9. Frappier, V., and Najmanovich, R. J. (2014) A Coarse-Grained Elastic Network Atom Contact Model and Its Use in the Simulation of Protein Dynamics and the Prediction of the Effect of Mutations. *Plos Computational Biology* **10**
10. Ioannidis, D., Papadopoulos, G. E., Anastassopoulos, G., Kortsaris, A., and Anagnostopoulos, K. (2015) Structural Properties and Interaction Energies Affecting Drug Design. An Approach Combining Molecular Simulations, Statistics, Interaction Energies and Neural Networks. *Computational Biology and Chemistry* **56**, 7-12
11. Paschek, D., Golub, B., and Ludwig, R. (2015) Hydrogen Bonding in A Mixture of Protic Ionic Liquids: A Molecular Dynamics Simulation Study. *Physical Chemistry Chemical Physics* **17**, 8431-8440
12. Zeng, J.-P., Dai, Y., Shi, W.-Y., Shao, J.-L., and Sun, G.-X. (2015) Molecular Dynamics Simulation on the Interaction Between Polymer Inhibitors and Anhydrite Surface. *Surface and Interface Analysis* **47**, 896-902
13. Barducci, A., Bonomi, M., Prakash, M. K., and Parrinello, M. (2013) Free-Energy Landscape of Protein Oligomerization From Atomistic Simulations. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E4708-E4713
14. Pontiggia, F., Pachov, D. V., Clarkson, M. W., Villali, J., Hagan, M. F., Pande, V. S., and Kern, D. (2015) Free Energy Landscape of Activation in a Signalling Protein at Atomic Resolution. *Nature Communications* **6**
15. Miao, Y., Feher, V. A., and McCammon, J. A. (2015) Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **11**, 3584-3595

16. Held, M., and Noe, F. (2012) Calculating Kinetics and Pathways of Protein-Ligand Association. *European Journal of Cell Biology* **91**, 357-364
17. Tiwary, P., Limongelli, V., Salvalaglio, M., and Parrinello, M. (2015) Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E386-E391
18. Cavalli, A., Spitaleri, A., Saladino, G., and Gervasio, F. L. (2015) Investigating Drug-Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Accounts of Chemical Research* **48**, 277-285
19. Schmidtke, P., Javier Luque, F., Murray, J. B., and Barril, X. (2011) Shielded Hydrogen Bonds as Structural Determinants of Binding Kinetics: Application in Drug Design. *Journal of the American Chemical Society* **133**, 18903-18910
20. Wang, Y., Lv, J., Zhu, L., and Ma, Y. (2012) CALYPSO: A Method for Crystal Structure Prediction. *Computer Physics Communications* **183**, 2063-2070
21. Zhu, Q., Oganov, A. R., and Lyakhov, A. O. (2012) Evolutionary Metadynamics: A Novel Method to Predict Crystal Structures. *Crystengcomm* **14**, 3596-3601
22. Abrol, R., Bray, J. K., and Goddard, W. A., III. (2012) Bihelix: Towards de novo Structure Prediction of an Ensemble of G-Protein Coupled Receptor Conformations. *Proteins-Structure Function and Bioinformatics* **80**, 505-518
23. Lyakhov, A. O., Oganov, A. R., Stokes, H. T., and Zhu, Q. (2013) New Developments in Evolutionary Structure Prediction Algorithm USPEX. *Computer Physics Communications* **184**, 1172-1182
24. Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., Young, C., Batson, B., Bowers, K. J., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., Ierardi, D. J., Kolossvary, I., Klepeis, J.

- L., Layman, T., McLeavey, C., Moraes, M. A., Mueller, R., Priest, E. C., Shan, Y., Spengler, J., Theobald, M., Towles, B., and Wang, S. C. (2008) Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Communications of the Acm* **51**, 91-97
25. Shaw, D. E., Grossman, J. P., Bank, J. A., Batson, B., Butts, J. A., Chao, J. C., Deneroff, M. M., Dror, R. O., Even, A., Fenton, C. H., Forte, A., Gagliardo, J., Gill, G., Greskamp, B., Ho, C. R., Ierardi, D. J., Iserovich, L., Kuskin, J. S., Larson, R. H., Layman, T., Li-Siang, L., Lerer, A. K., Li, C., Killebrew, D., Mackenzie, K. M., Mok, S. Y. H., Moraes, M. A., Mueller, R., Nociolo, L. J., Peticolas, J. L., Quan, T., Ramot, D., Salmon, J. K., Scarpazza, D. P., Schafer, U. B., Siddique, N., Snyder, C. W., Spengler, J., Tang, P. T. P., Theobald, M., Toma, H., Towles, B., Vitale, B., Wang, S. C., and Young, C. (2014) Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*, 41-53
26. Arkhipov, A., Shan, Y., Das, R., Endres, N. F., Eastwood, M. P., Wemmer, D. E., Kuriyan, J., and Shaw, D. E. (2013) Architecture and Membrane Interactions of the EGF Receptor. *Cell* **152**, 557-569
27. Decherchi, S., Berteotti, A., Bottegoni, G., Rocchia, W., and Cavalli, A. (2015) The Ligand Binding Mechanism to Purine Nucleoside Phosphorylase Elucidated via Molecular Dynamics and Machine Learning. *Nature Communications* **6**
28. Pierce, L. C. T., Salomon-Ferrer, R., de Oliveira, C. A. F., McCammon, J. A., and Walker, R. C. (2012) Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation* **8**, 2997-3002

29. Schlitter, J., Engels, M., and Kruger, P. (1994) Targeted Molecular-Dynamics - A New Approach for Searching Pathways of Conformational Transitions. *Journal of Molecular Graphics* **12**, 84-89
30. Leech, J., Prins, J. F., and Hermans, J. (1996) SMD: Visual Steering of Molecular Dynamics for Protein Design. *Ieee Computational Science & Engineering* **3**, 38-45
31. Raiteri, P., Laio, A., Gervasio, F. L., Micheletti, C., and Parrinello, M. (2006) Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *Journal of Physical Chemistry B* **110**, 3533-3539
32. Laio, A., Rodriguez-Forteza, A., Gervasio, F. L., Ceccarelli, M., and Parrinello, M. (2005) Assessing the Accuracy of Metadynamics. *Journal of Physical Chemistry B* **109**, 6714-6721
33. Ensing, B., De Vivo, M., Liu, Z. W., Moore, P., and Klein, M. L. (2006) Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions. *Accounts of Chemical Research* **39**, 73-81
34. Bussi, G., Laio, A., and Parrinello, M. (2006) Equilibrium Free Energies From Nonequilibrium Metadynamics. *Physical Review Letters* **96**, 4
35. Barducci, A., Bonomi, M., and Parrinello, M. (2011) Metadynamics. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **1**, 826-843
36. Marinelli, F., and Faraldo-Gomez, J. D. (2015) Ensemble-Biased Metadynamics: A Molecular Simulation Method to Sample Experimental Distributions. *Biophysical Journal* **108**, 2779-2782
37. Wu, X. W., and Brooks, B. R. (2003) Self-Guided Langevin Dynamics Simulation Method. *Chemical Physics Letters* **381**, 512-518

38. Rudd, R. E., and Broughton, J. Q. (1998) Coarse-Grained Molecular Dynamics and the Atomic Limit Of Finite Elements. *Physical Review B* **58**, R5893-R5896
39. Hamelberg, D., Mongan, J., and McCammon, J. A. (2004) Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *Journal of Chemical Physics* **120**, 11919-11929
40. Doshi, U., and Hamelberg, D. (2014) Achieving Rigorous Accelerated Conformational Sampling in Explicit Solvent. *Journal of Physical Chemistry Letters* **5**, 1217-1224
41. Hornak, V., and Simmerling, C. (2004) Development of Softcore Potential Functions for Overcoming Steric Barriers in Molecular Dynamics Simulations. *Journal of Molecular Graphics & Modelling* **22**, 405-413
42. Bello-Rivas, J. M., and Elber, R. (2015) Exact Milestoning. *Journal of Chemical Physics* **142**
43. Faradjian, A. K., and Elber, R. (2004) Computing Time Scales From Reaction Coordinates by Milestoning. *Journal of Chemical Physics* **120**, 10880-10889
44. Chung, H. S., Louis, J. M., and Eaton, W. A. (2009) Experimental Determination of Upper Bound for Transition Path Times in Protein Folding From Single-Molecule Photon-By-Photon Trajectories. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11837-11844
45. Bacci, M., Vitalis, A., and Caflisch, A. (2015) A Molecular Simulation Protocol to Avoid Sampling Redundancy and Discover New States. *Biochimica Et Biophysica Acta-General Subjects* **1850**, 889-902
46. Zhou, T., and Caflisch, A. (2012) Free Energy Guided Sampling. *Journal of Chemical Theory and Computation* **8**, 2134-2140

47. Dickson, A., and Brooks, C. L., III. (2014) WExplore: Hierarchical Exploration of High-Dimensional Spaces Using the Weighted Ensemble Algorithm. *Journal of Physical Chemistry B* **118**, 3532-3542
48. Harada, R., and Kitao, A. (2013) Parallel Cascade Selection Molecular Dynamics (PaCS-MD) to Generate Conformational Transition Pathway. *Journal of Chemical Physics* **139**, 10
49. Harada, R., and Kitao, A. (2015) Nontargeted Parallel Cascade Selection Molecular Dynamics for Enhancing the Conformational Sampling of Proteins. *Journal of Chemical Theory and Computation* **11**, 5493-5502
50. Harada, R., Takano, Y., and Shigeta, Y. (2014) Fluctuation Flooding Method (FFM) for Accelerating Conformational Transitions of Proteins. *Journal of Chemical Physics* **140**
51. Harada, R., Takano, Y., and Shigeta, Y. (2015) Enhanced Conformational Sampling Method for Proteins Based on the TaBoo SeArch Algorithm: Application to the Folding of a Mini-Protein, Chignolin. *Journal of Computational Chemistry* **36**, 763-772
52. Harada, R., Nakamura, T., Takano, Y., and Shigeta, Y. (2015) Protein Folding Pathways Extracted by OFLOOD: Outlier FLOODing Method. *Journal of Computational Chemistry* **36**, 97-102
53. Voter, A. F. (1998) Parallel Replica Method for Dynamics of Infrequent Events. *Physical Review B* **57**, 13985-13988
54. Berg, B. A., and Neuhaus, T. (1991) Multicanonical Algorithms for 1st Order Phase-Transitions. *Physics Letters B* **267**, 249-253
55. Sugita, Y., and Okamoto, Y. (1999) Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **314**, 141-151

56. Yamamori, Y., and Kitao, A. (2013) MuSTAR MD: Multi-Scale Sampling Using Temperature Accelerated and Replica Exchange Molecular Dynamics. *Journal of Chemical Physics* **139**, 11
57. Bolhuis, P. G., Chandler, D., Dellago, C., and Geissler, P. L. (2002) Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annual Review of Physical Chemistry* **53**, 291-318
58. Dellago, C., Bolhuis, P. G., Csajka, F. S., and Chandler, D. (1998) Transition Path Sampling and the Calculation of Rate Constants. *Journal of Chemical Physics* **108**, 1964-1977
59. Swenson, D. W. H., and Bolhuis, P. G. (2014) A Replica Exchange Transition Interface Sampling Method With Multiple Interface Sets for Investigating Networks of Rare Events. *Journal of Chemical Physics* **141**
60. Metropolis, N., and Ulam, S. (1949) The Monte Carlo Method. *Journal of the American Statistical Association* **44**, 335-341
61. Caflisch, R. E. (1998) Monte Carlo and Quasi-Monte Carlo Methods. *Acta Numerica* **7**, 1-49
62. Pangali, C., Rao, M., and Berne, B. J. (1978) Novel Monte-Carlo Scheme for Simulating Water and Aqueous-Solutions. *Chemical Physics Letters* **55**, 413-417
63. Rossky, P. J., Doll, J. D., and Friedman, H. L. (1978) Brownian Dynamics as Smart Monte-Carlo Simulation. *Journal of Chemical Physics* **69**, 4628-4633
64. Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987) Hybrid Monte-Carlo. *Physics Letters B* **195**, 216-222

65. Frantz, D. D., Freeman, D. L., and Doll, J. D. (1990) Reducing Quasi-Ergodic Behavior in Monte-Carlo Simulations by J-Walking - Applications to Atomic Clusters. *Journal of Chemical Physics* **93**, 2769-2784
66. Lyubartsev, A. P., Martsinovski, A. A., Shevkunov, S. V., and Vorontsovveliaminov, P. N. (1992) New Approach to Monte-Carlo Calculation of the Free-Energy - Method of Expanded Ensembles. *Journal of Chemical Physics* **96**, 1776-1783
67. Marinari, E., and Parisi, G. (1992) Simulated Tempering - A New Monte-Carlo Scheme. *Europhysics Letters* **19**, 451-458
68. Matro, A., Freeman, D. L., and Topper, R. Q. (1996) Computational Study of the Structures and Thermodynamic Properties of Ammonium Chloride Clusters Using A Parallel Jump-Walking Approach. *Journal of Chemical Physics* **104**, 8690-8702
69. Hukushima, K., and Nemoto, K. (1996) Exchange Monte Carlo Method and Application to Spin Glass Simulations. *Journal of the Physical Society of Japan* **65**, 1604-1608
70. Geyer, C. J., and Thompson, E. A. (1995) Annealing Markov-Chain Monte-Carlo With Applications to Ancestral Inference. *Journal of the American Statistical Association* **90**, 909-920
71. Zhou, R. H., and Berne, B. J. (1997) Smart Walking: A New Method for Boltzmann Sampling of Protein Conformations. *Journal of Chemical Physics* **107**, 9185-9196
72. Wang, F. G., and Landau, D. P. (2001) Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Physical Review Letters* **86**, 2050-2053

73. Andricioaei, I., Straub, J. E., and Voter, A. F. (2001) Smart Darting Monte Carlo. *Journal of Chemical Physics* **114**, 6994-7000
74. Brown, S., and Head-Gordon, T. (2003) Cool Walking: A New Markov Chain Monte Carlo Sampling Method. *Journal of Computational Chemistry* **24**, 68-76
75. Li, H., Li, G., Berg, B. A., and Yang, W. (2006) Finite Reservoir Replica Exchange to Enhance Canonical Sampling in Rugged Energy Surfaces. *Journal of Chemical Physics* **125**
76. Kolossvary, I., and Guida, W. C. (1996) Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides. *Journal of the American Chemical Society* **118**, 5011-5019
77. Kolossvary, I., and Keseru, G. M. (2001) Hessian-Free Low-Mode Conformational Search for Large-Scale Protein Loop Optimization: Application to c-jun N-Terminal Kinase JNK3. *Journal of Computational Chemistry* **22**, 21-30
78. Chang, C. E., and Gilson, M. K. (2003) Tork: Conformational Analysis Method for Molecules and Complexes. *Journal of Computational Chemistry* **24**, 1987-1998
79. Labute, P. (2010) LowModeMD-Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops. *Journal of Chemical Information and Modeling* **50**, 792-800
80. Marechal, J.-D., and Perahia, D. (2008) Use of Normal Modes for Structural Modeling of Proteins: The Case Study of Rat Heme Oxygenase 1. *European Biophysics Journal with Biophysics Letters* **37**, 1157-1165

81. Kantarci-Carsibasi, N., Haliloglu, T., and Doruker, P. (2008) Conformational Transition Pathways Explored by Monte Carlo Simulation Integrated with Collective Modes. *Biophysical Journal* **95**, 5862-5873
82. Zheng, W. J., and Brooks, B. R. (2006) Modeling Protein Conformational Changes by Iterative Fitting of Distance Constraints Using Reoriented Normal Modes. *Biophysical Journal* **90**, 4327-4336
83. Mu, Y. G., Nguyen, P. H., and Stock, G. (2005) Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins-Structure Function and Bioinformatics* **58**, 45-52
84. Altis, A., Nguyen, P. H., Hegger, R., and Stock, G. (2007) Dihedral Angle Principal Component Analysis of Molecular Dynamics Simulations. *Journal of Chemical Physics* **126**
85. Maisuradze, G. G., and Leitner, D. M. (2007) Free Energy Landscape of a Biomolecule in Dihedral Principal Component Space: Sampling Convergence and Correspondence Between Structures and Minima. *Proteins-Structure Function and Bioinformatics* **67**, 569-578
86. Altis, A., Otten, M., Nguyen, P. H., Hegger, R., and Stock, G. (2008) Construction of the Free Energy Landscape of Biomolecules via Dihedral Angle Principal Component Analysis. *Journal of Chemical Physics* **128**
87. Maisuradze, G. G., Liwo, A., and Scheraga, H. A. (2009) Principal Component Analysis for Protein Folding Dynamics. *Journal of Molecular Biology* **385**, 312-329
88. Riccardi, L., Nguyen, P. H., and Stock, G. (2009) Free-Energy Landscape of RNA Hairpins Constructed via Dihedral Angle Principal Component Analysis. *Journal of Physical Chemistry B* **113**, 16660-16668

89. Cortes, J., Simeon, T., de Angulo, V. R., Guieysse, A. D., Remaud-Simeon, M., and Tran, V. (2005) A Path Planning Approach for Computing Large-Amplitude Motions of Flexible Molecules. *Bioinformatics* **21**, I116-I125
90. Ruiz de Angulo, V., Cortes, J., and Simeon, T. (2005) BioCD : An Efficient Algorithm for Self-Collision and Distance Computation Between Highly Articulated Molecular Models. MIT Press, Robotics: Science and Systems Conference
91. Al-Bluwi, I., Vaisset, M., Simeon, T., and Cortes, J. (2013) Modeling Protein Conformational Transitions by a Combination of Coarse-Grained Normal Mode Analysis and Robotics-Inspired Methods. *Bmc Structural Biology* **13**
92. Miyazawa, T., and Pitzer, K. S. (1959) Internal Rotation and Infrared Spectra of Formic Acid Monomer and Normal Coordinate Treatment of Out-of-Plane Vibrations of Monomer, Dimer, And Polymer. *Journal of Chemical Physics* **30**, 1076-1086
93. Parsons, J., Holmes, J. B., Rojas, J. M., Tsai, J., and Strauss, C. E. M. (2005) Practical Conversion From Torsion Space to Cartesian Space for in Silico Protein Synthesis. *Journal of Computational Chemistry* **26**, 1063-1068
94. Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., York, D. M., and Kollman, P. A. (2015) Amber 2015. University of California, San Francisco

95. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **11**, 3696-3713
96. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *Journal of the American Chemical Society* **112**, 6127-6129
97. Pedretti, A., Villa, L., and Vistoli, G. (2004) VEGA - An Open Platform to Develop Chemo-Bio-Informatics Applications, Using Plug-in Architecture and Script Programming. *Journal of Computer-Aided Molecular Design* **18**, 167-173
98. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry* **30**, 2785-2791
99. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935
100. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **25**, 1157-1174
101. Roberts, C. C., and Chang, C.-e. A. (2013) Ligand Binding Pathway Elucidation for Cryptophane Host-Guest Complexes. *Journal of Chemical Theory and Computation* **9**, 2010-2019
102. Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) Charmm - A Program for Macromolecular Energy,

- Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **4**, 187-217
103. Mackerell, A. D., Wiorkiewicz-Kuczera, J., and Karplus, M. (1995) An All-Atom Empirical Energy Function for the Simulation of Nucleic-Acids. *Journal of the American Chemical Society* **117**, 11946-11975
104. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *Journal of Physical Chemistry B* **102**, 3586-3616
105. Grossfield, A. WHAM: The Weighted Histogram Analysis Method. Version 2.0.9 Ed.
106. Garcia, C., Humiliere, D., Riva, N., Collet, A., and Dutasta, J. P. (2003) Kinetic and Thermodynamic Consequences of the Substitution of SMe for OMe Substituents of Cryptophane Hosts on the Binding of Neutral and Cationic Guests. *Organic & Biomolecular Chemistry* **1**, 2207-2216
107. Swope, W. C., Pitera, J. W., and Suits, F. (2004) Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory. *Journal of Physical Chemistry B* **108**, 6571-6581
108. Swope, W. C., Pitera, J. W., Suits, F., Pitman, M., Eleftheriou, M., Fitch, B. G., Germain, R. S., Rayshubski, A., Ward, T. J. C., Zhestkov, Y., and Zhou, R. (2004) Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and Beta-Hairpin Peptide. *Journal of Physical Chemistry B* **108**, 6582-6594

## **Chapter 7 Conclusion and Future Directions in Protein-Ligand Dissociation**

### **Research**

#### 7.1 Conclusion

Overall, small host-guest and protein-ligand complexes were used to study molecular recognition. Besides that, simulations obtained from these complexes were used to explore the roles of thermodynamics and kinetics. In addition, a novel computational tool was designed, which was inspired from the application and comprehension of existing methods.

To start with, a method composed of an aggressive conformational search and a free energy calculation was used to study the binding affinity of a small bimolecular system, a series of ligand-nanoplate complexes. The secret behind the magic surfactant citrate was unraveled. The calculations support the experimental trends of anisotropic growth of the silver plates from the perspective of ligand-nanoplate interactions and provide insights into the mechanisms of the directional growth. The selectivity of surfactants on the two silver facets was attributed to the partial charge distribution, conformational fitness, surfactant flexibility, and entropic effects. Considering a good ligand should be flexible but not too flexible resulting in severe entropic penalty, two ligands were designed by mimicking the structure of trimesic acid. In this study the limitation of implicit solvent model is demonstrated again and raises the need for a better description of the thermodynamic properties using explicit solvent model.

Continuing the study of binding affinity and moving to the larger scale systems, the binding pose and binding affinity were studied with the TIM44-MitoBloCK-10 system. Due to a larger system size, the aggressive conformational search method used in the previous study is not efficient enough. Instead, a docking and MD approach was adopted and the binding pose was determined which explained the experimental mutation studies. This study demonstrates that the computational methods are able to provide atomistic and energetic insights to complement the experimental studies.

An accurate evaluation of the binding enthalpy and entropy based in explicit solvent model was performed on the  $\beta$ -cyclodextrin complexes. Four trajectories for the complex, host, guest and pure water box were simulated using GPU accelerated MD. The enthalpy calculation was accomplished by potential energy calculation. The solute entropy change was evaluated based population analysis of the dihedral angle while the solvent entropy was calculated using the cell method. Other solvent entropy evaluation methods such as PCF method, IFST, GIST and 2PT were available to be used, but due to the advantages and ease of implementation, the cell method was chosen for this study. Beyond the domain of thermodynamics, the association/dissociation rates were directly extracted from the trajectories of  $\beta$ -cyclodextrin complexes and the kinetic-thermodynamic relationships were discussed. For such weak binding, the driving force is not necessarily the attraction between the host and guest but entropy played an important role. Due to the entropic driven binding process of the weak ligands, the association pathways also differ from the other ligands with strong attraction to the host.

Finally, to sample the dissociation process of protein-ligand systems, a novel conformational search algorithm was developed using a newly derived internal PCA method. The internal PCA method differs from the dihedral PCA through the way it constructs the internal coordinate. Also, it is able to reproduce atomistic motion of the system which resembles the  $\alpha$ -carbon motions in the  $\alpha$ -carbon PCA. Then the internal PCA search method utilizes the orthogonal set of internal PCA modes to search for the dissociation pathways. The efficiency of this method comes from the fact that it does not require full minimization or energy evaluation. Instead a collision test based acceptance test was introduced to ensure physical motions in the pathways. This method provides a new way to sampling the dissociation pathways.

## 7.2 Future Works

### 7.2.1 Cyclin-Dependent Kinase 8 (CDK8) Dissociation Pathway

The cyclin-dependent kinase 8/cyclin C (CDK8/CycC) complex is one kinase system in the CDK family, and regarded as an attractive drug target. Similar to other kinases, this protein has an ATP binding site, which adopts the type I drugs that inhibit kinases non-selectively, and an allosteric binding which allows selectivity due to structural differences among the kinases (Figure 7.1). CDK8 has a DMG motif which functions similarly as the DFG motif in other kinases. However, in a recent structure-kinetic relationship study <sup>1</sup>, it was discovered that the flip of the DMG motif does not affect the residence time <sup>2</sup> of the inhibitors. This drove our interest to investigate the inhibitor dissociation pathway from CDK8. An MD study was performed to investigate the activation loop of CDK8 <sup>3</sup>. Considering the slow dissociation of the target inhibitor (residence time is 1626 min <sup>1</sup>),

conventional MD is no longer capable of sampling the complete dissociation pathway. Therefore, the internal PCA search method can be used to generate an initial guess of the dissociation pathway, and conventional MD can be then used to refine the pathway.

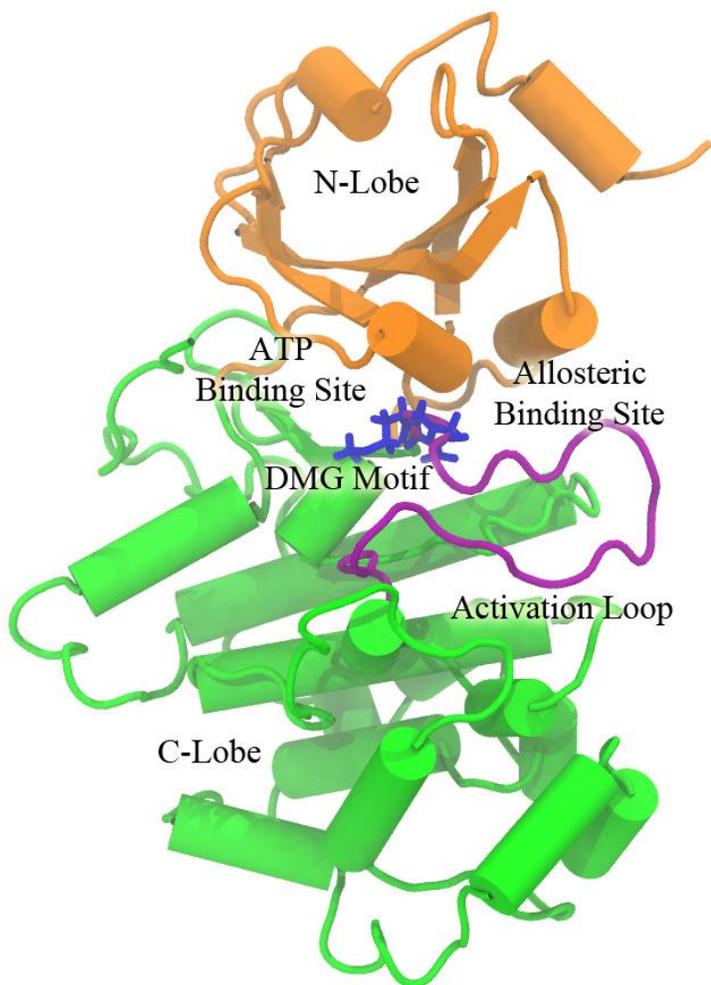


Figure 7.1. The structure of CDK8. The N-lobe is rendered in orange, the C-lobe is rendered in green, the activation loop is rendered in purple, and the DMG motif is rendered in blue. The ATP binding site and the allosteric binding site are on the left and right of the DMG motif respectively.

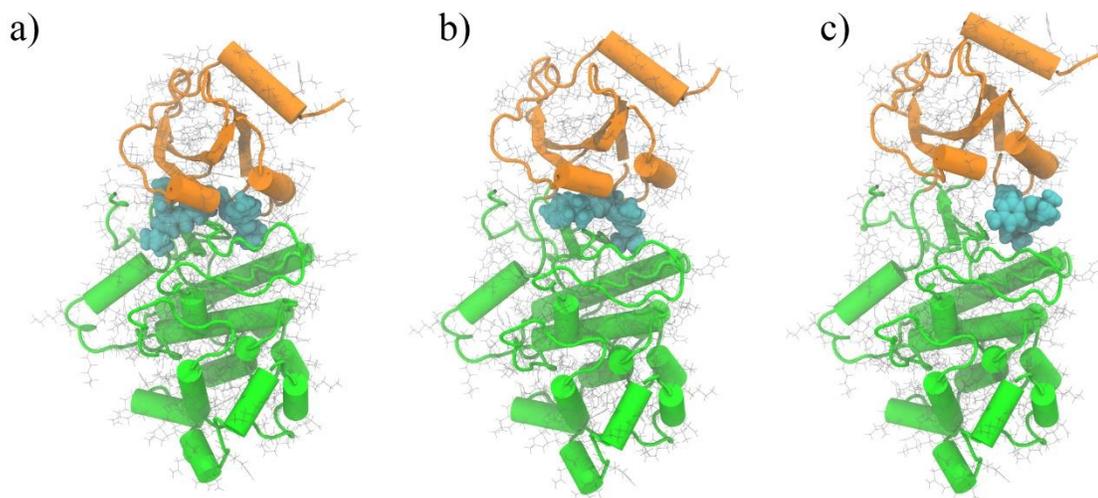


Figure 7.2. The dissociation pathway from internal PCA search. a) is the bound state, b) is the moment when the ligand cross the activation loop, and c) is the final state where the ligand stays in the allosteric binding site. The N-lobe is rendered in orange, the C-lobe is rendered in green, and the ligand is rendered in cyan.

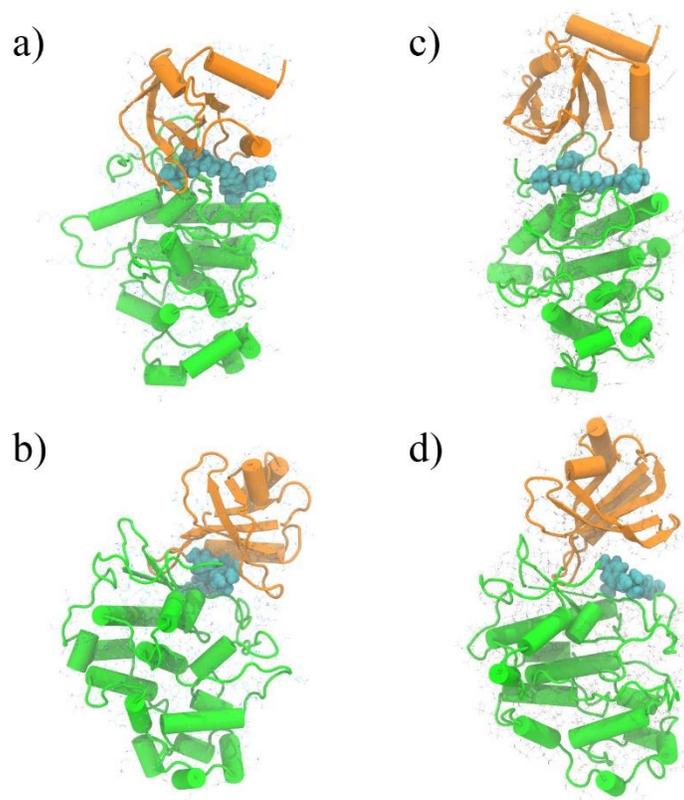


Figure 7.3. The snapshots from MD resampled from the internal PCA search pathway. a) and b) are the intermediate state from two perspectives of view. c) and d) are the breathing conformation of the CDK8 found by conventional MD from two perspectives of view. a) and c) are aligned, and b) and d) are aligned. The color pattern is the same as in Figure 7.2.

In the preliminary results, a dissociation pathway was found by the internal PCA search method (Figure 7.2). In this pathway, the ligand crossed the activation loop (Figure 7.2 b), and sampled the allosteric binding site (Figure 7.2 c). The protein underwent breathing motion, in which the N-lobe moved away from the C-lobe. To refine the pathway, MD simulations were performed on a total of 62 frames. These frames were evenly distributed along the pathway from conformational search. From the simulations, one highly possible intermediate state was identified (Figure 7.3 a and b), in which the

ligand partially dissociates from the binding site and stays on the activation loop while the two lobes of the protein are closed. From the intermediate state, conventional MD was able to sample the breathing motion of the protein (Figure 7.3 c and d).

In the future, further sampling will be continued from the conformation which led the protein to “breath”. In order to sample the complete unbinding event, the sampling will be continued towards bound state from the intermediate state. Another set of MD runs will be set up for the CDK8/CycC complex to investigate the effects from the partner protein.

#### 7.2.2 Improvement of Internal PCA Search Method

Currently this method suffers from two major drawbacks; the acceptance test is efficient but at the same time lacks a description of attraction force between regions of the system; the algorithm uses a single thread minimizer, which slows down the overall efficiency of the method. These allows further improvement of the method.

To address the first drawback, a multi-level model approach will be included in the algorithm. First, a coarse grained version<sup>4</sup> of the model will be constructed and corresponding PCA modes will be calculated. The coarse grained PCA modes will include the motions and amplitude of the motions, therefore these modes can be used to capture the attraction force between regions in the system. For example, regarding the breathing motion of CDK8 described in the previous section, in MD simulation, the hinge between the two lobes of CDK8 bends naturally to open up the binding sites, while in the

internal PCA search, the hinge is stretched exhibiting tension, even though the overall behavior is similar to that of MD (Figure 7.4). With the coarse grained PCA as an auxiliary acceptance test, such artifacts can be further eliminated.

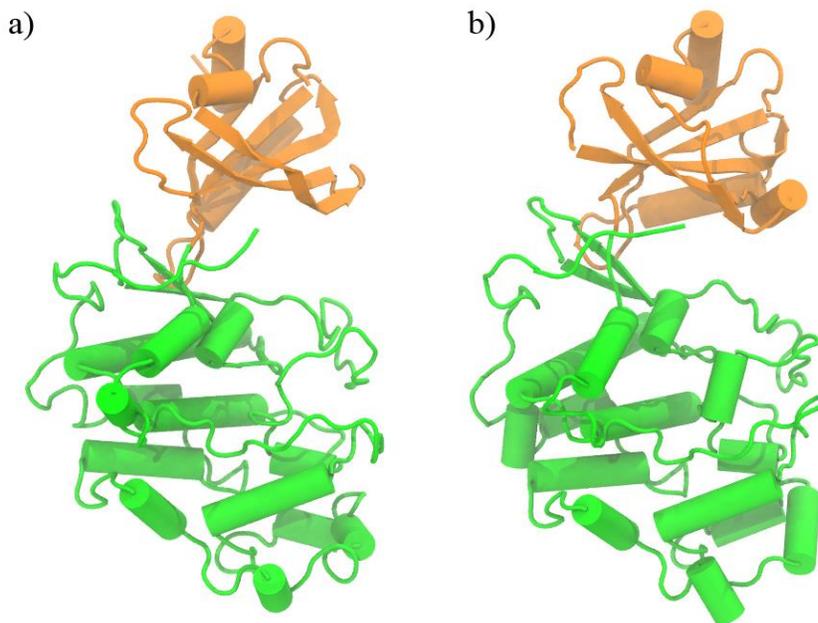


Figure 7.4. The comparison between MD (a) and internal PCA search (b) conformations. Both conformations are for the breathing motion of CDK8. The N-lobe of CDK8 is rendered in orange, and C-lobe in green. The hinge between the two lobes is more stretched in the internal PCA search result but more natural in the MD simulation.

Regarding the second drawback about efficiency, a straightforward solution will be replacing the single thread minimizer to a multi-thread. Such minimizers exist in academically or commercially available packages including NAMD<sup>5</sup> and AMBER<sup>6</sup>, and GPU acceleration of minimization<sup>7</sup> is also available. Therefore this issue is simply a matter of technical implementation.

### 7.3 References

1. Schneider, E. V., Boettcher, J., Huber, R., Maskos, K., and Neumann, L. (2013) Structure-Kinetic Relationship Study of CDK8/CycC Specific Compounds. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8081-8086
2. Copeland, R. A., Pompliano, D. L., and Meek, T. D. (2007) Drug-Target Residence Time and Its Implications for Lead Optimization (vol 5, pg 730, 2006). *Nature Reviews Drug Discovery* **6**, 249-249
3. Xu, W., Amire-Brahimi, B., Xie, X.-J., Huang, L., and Ji, J.-Y. (2014) All-Atomic Molecular Dynamic Studies of Human CDK8: Insight Into the  $\alpha$ -Loop, Point Mutations and Binding with Its Partner CycC. *Computational Biology and Chemistry* **51**, 1-11
4. Tozzini, V. (2005) Coarse-Grained Models for Proteins. *Current Opinion in Structural Biology* **15**, 144-150
5. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781-1802
6. Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., York, D. M., and Kollman, P. A. (2015) AMBER 2015. University of California, San Francisco

7. Friedrichs, M. S., Eastman, P., Vaidyanathan, V., Houston, M., Legrand, S., Beberg, A. L., Ensign, D. L., Bruns, C. M., and Pande, V. S. (2009) Accelerating Molecular Dynamic Simulation on Graphics Processing Units. *Journal of Computational Chemistry* **30**, 864-872