

APPLICATION OF FREE ENERGY METHODS TO DRUG DISCOVERY

By

Lin Song

A DISSERTATION

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

Chemistry—Doctor of Philosophy

2020

ABSTRACT

APPLICATION OF FREE ENERGY METHODS TO DRUG DISCOVERY

By

Lin Song

With the increasing power of computers, computational studies have become more and more significant in drug discovery. High binding free energy is one of the major requirements for an effective drug molecule, hence much effort has been spent to develop fast and accurate computational free energy methods. In this thesis, different free energy methods, *i.e.* umbrella sampling, thermodynamic integration, and double decoupling method, are applied to different systems related to drug discovery. For the first study, umbrella sampling studies are performed to calculate the absolute binding free energies of host-guest systems which serve as great model systems to assess free energy methods due to the small size of the systems, *etc.* We find that benchmarking our method with known systems can significantly improve the results for the unknown systems: the overall RMSE of the binding free energy versus experiment is reduced from 5.59 kcal/mol to 2.36 kcal/mol. The source of error could be from the un-optimized force constants used in umbrella sampling (hence possibly poor window overlaps), as well as force field, sampling issues, *etc.* Our results ranked 4th best in the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL6) blind challenge. For the second study, GPU accelerated thermodynamic integration (GPU-TI) is used to compute the relative binding free energies of a protein-ligand dataset originally assembled by Schrödinger, Inc. The calculations of relative binding free energies between different ligands are the typical process in the lead optimization of computer-aided drug discovery. In our study using GPU-TI from AMBER 18 with the AMBER14SB/GAFF1.8 force field, we obtained an overall MUE of 1.17 kcal/mol and an overall RMSE of 1.50 kcal/mol for the

330 perturbations contained in this data set. They are comparable to the overall MUE of 0.9 kcal/mol and RMSE of 1.14 kcal/mol using their GPU free energy code (FEP+) and the OPLS2.1 force field combined with the REST2 enhanced sampling by Schrödinger, Inc. Notably, after we published our work, several other research groups reported their benchmarking results on the other free energy software using the same dataset.

The third study of this thesis focuses on modeling the thermodynamics of transition metal (TM) ions binding to a protein. TM ions are very common in biology and are important in drug discovery as well, because many TM ions are in the active site of the protein where the inhibitors bind, for example, the histone deacetylase. While the structural details of TMs bound to metalloproteins are generally well understood via experimental and computational means; studies accurately describing the thermodynamics of TM ion binding are less common. Herein, we demonstrate that we can obtain accurate structural and absolute binding free energies of Co^{2+} and Ni^{2+} to the enzyme glyoxalase I (GlxI) using an optimized 12-6-4 (m12-6-4) potential. Optimizing the 12-6-4 potential to accurately model the interactions between the TMs and the binding site residues, as well as protonation state changes associated with TMs (un)binding, are found to be crucial. Given the success of this study, we are now in a position to explore more complicated processes associated with TM-based drug discovery.

This dissertation is dedicated to my parents and my church family (LCCC), who always love, guide, and support me.

ACKNOWLEDGEMENTS

First and foremost, I thank the Lord Jesus Christ, God's one and only Son, for saving me, loving me and carrying the cross for me. Thanks for never forsaking me nor leaving me, through good times and bad times. The Lord's people in Lansing Chinese Christian Church has been supporting me, guiding me and loving me through all these years as well. Pastor Peng and Elder Kris Wang have been great elders, shepherds, who taught me God's word and lead me by their servanthood example. Thanks for all the love, all the patience, and kindness, all the one-on-one talks, the encouragements when I was down, the many prayers that only the Lord knows and remembers. Thank you for the love from your family as well: Pastor Peng's wife Christine, Kris's wife Beiying, and your lovely children of course. Thanks Kris and Beiying for treating me like your own child. I hope and pray that the Lord continue to bless you and your family, and every brother and sister in LCCC. Paul Ding and sister Qianxi are a couple in LCCC that I have known since I prayed the sinner's prayer. Thanks for loving me and showing me brotherhood since the beginning of my pilgrim. I really enjoyed the time when Paul was our cell group leader, and will always miss the fellowship in those days. May your children grow in the Lord and may your family become a blessing to a lot of great fellowship in the Lord. Sister Shuangwen is the Lord's angel who always takes care of me, especially in my difficult times. Shuangwen is an elder sister in LCCC, but she is always joyful, energetic, full of passion and compassion. Thanks for the encouraging words from the Lord, the fellowship in my difficult times, and the prayers. Sister Huei is another angel from the Lord, who is very gentle and warm. Thanks for your support and love in my difficult times. Bin is a post-doc at MSU and a great brother in LCCC. Thanks for your support and for welcoming me to live in your apartment for free. Too many brothers and sisters carried me through

all these years and loved me because of the love of the Lord. Dora, Zhongguang, Li Xie and Qingwei, Tsun-en Lu, Calvin Chin, Yizhou, Tong Wu, Xiaochen Yuan, Chen Du and Mike, Chuanyin, Jason, Donghao, Zhen, Qiaoqin, Tony and Wanyu, Lihua, Xiongzhi, and Chunzi, *etc.* The list can go on and on. Thank you all for your love because of the Lord. May the Lord bless you and through you bless many others. Special thanks to Nick Settington who also led me to the Lord and shared the Gospel with me. It's great to see you once in a while on campus or at URC.

I would like to thank my parents, who gave me life and raised me, love me, educate me, and care about me more than care about themselves. Though I have been one of the best students at school, I was not a lovely boy, nor a lovely teenager. I had a very bad temper. Thanks for putting up with me. Thanks to mom for accompanying me through my whole life, and to dad for working hard to support the family. Thanks to all my other family members as well: the siblings from my mom's side, and my dad's side, my brothers and sisters, cousins, grandpas, and grandmas. Special thanks to my grandmother, who passed away about 12 years ago. Thanks for loving me and taking care of me before I entered high school.

I would like to thank my advisor, Kennie. Thanks for providing the grant for me to do research, providing research projects, spending time and effort to discuss with me about my research. Thanks for your insightful and broad thoughts and ideas in research. Also, thanks for your patience and encouragement when I made mistakes. I'm grateful for Kennie's help and care for me through all these years, especially during my difficult times. I can easily fail my Ph.D. study if I were not in Kennie's lab. Thanks for being a great advisor. I would like to thank my other committee members as well: Prof. Katharine Hunt, Prof. Robert Cukier, Prof. Benjamin Levine, Prof. Robert Hausinger. Thank you for spending time to discuss with me about my projects, writing

recommendation letters for me, collaborations, and the other supports in my research career.

I would like to thank my lab mates and collaborators. Especially I would like to thank Pengfei, who was a senior Ph.D. student in Merz lab. Pengfei has taught me almost everything about AMBER. At the beginning of my Ph.D. study, Pengfei welcomed me to participate in the metal ion parameterization projects and also helped me to do my own projects afterward. Pengfei's attitude towards research, life, and people, has greatly benefited me since we were roommates for more than two years and we go to the lab and go to LCCC together. Thanks for caring about me through all these years, even after you moved to other cities. Pray that you will continue to be a blessing to people around you. I'd like to thank Lili for teaching me how to do QM/MM MD simulations. Thanks Nupur for involving me in multiple projects and helping me to publish my first paper. Thanks Jun for the patient explanation about machine learning and for the company since we entered the Merz lab about the same time. Thanks John for discussions that helped me to derive some equations that turn out to be useful in my final project. Thanks Darrin and Tai-sung for collaborating with me.

I would like to thank my roommate, Yuelin, who helped me a lot in my daily life. Thanks for taking me to the hospital at midnight, sharing great food and snacks, talking with me when I was bored, *etc.* Thanks for the company. Thanks to my other friends, colleges, and teachers as well.

Finally, I would like to thank the computing support from the high-performance computing center at MSU and all the resources the department and the university provides.

TABLE OF CONTENTS

LIST OF TABLES.....	x
LIST OF FIGURES	xii
KEY TO ABBREVIATIONS.....	xv
CHAPTER 1: INTRODUCTION.....	1
1.1 General Introduction.....	1
1.2 Umbrella Sampling and WHAM	4
1.3 Alchemical Free Energy Methods in Computer Aided Drug Design.....	6
1.3.1 Theoretical Foundation	6
1.3.2 Enabling Technologies.....	7
1.3.3 Early FEP studies	10
1.3.4 Modern FEP in CADD.....	13
1.3.5 From retrospective to prospective	14
1.3.6 Modern Status	18
1.4 Metal Ion Force Field: the 12-6-4 Lennard Jones Nonbonded Model.....	21
CHAPTER 2: UMBRELLA SAMPLING STUDIES ON HOST-GUEST SYSTEMS.....	24
2.1 Methods.....	24
2.2 Systems: Benchmark and Test Systems.....	27
2.3 Results and Discussion	29
2.3.1 Benchmark Systems.....	29
2.3.2 Test Systems	34
2.3.3 Lesson Learned	39
2.3.1.1 The Two Faces of CB8.....	39
2.3.1.2 Lesson Learned From CB8-Ligand 5	42
2.4 Conclusions.....	44
CHAPTER 3: GPU TI STUDIES ON PROTEIN-LIGAND SYSTEMS.....	46
3.1 Methods.....	46
3.1.1 System Preparation	46
3.1.2 TI Simulation: The One-step Protocol.....	47
3.2 Results and Discussion	49
3.2.1 Overall Results.....	49
3.2.2 Uncertainty Estimate	53
3.2.3 The “Problematic Cases”	54
3.2.4 The Three-step Protocol	56
3.2.5 Discussion.....	56
3.3 Conclusion	57
CHAPTER 4: THERMODYNAMICS OF TRANSITION METAL ION BINDING TO PROTEINS	59

4.1	Introduction.....	59
4.2	Methods.....	63
4.2.1	<i>Optimization of the 12-6-4 Potentials</i>	63
4.2.1.1	<i>PMF calculations</i>	64
4.2.2	<i>Binding Free Energy Calculations</i>	66
4.2.2.1	<i>System Preparation</i>	66
4.2.2.2	<i>Free MD simulation</i>	66
4.2.2.3	<i>Energy Calculations</i>	67
4.2.2.3.1	<i>Loss of metal ion</i>	67
4.2.2.3.1.1	<i>The DDM theory and procedure</i>	67
4.2.2.3.1.2	<i>Hydration free energy (ΔGHFE0)</i>	68
4.2.2.3.1.3	<i>Step1 of Figure 22(b) (ΔG10)</i>	69
4.2.2.3.1.4	<i>Step2 of Figure 22(b) (ΔG20)</i>	70
4.2.2.3.1.5	<i>Step3 of Figure 22(b) (ΔG30)</i>	70
4.2.2.3.1.6	<i>Restraint set-up</i>	72
4.2.2.3.2	<i>Subsequent protonation</i>	72
4.2.2.3.2.1	<i>System preparation</i>	73
4.2.2.3.2.2	<i>TI simulations</i>	73
4.3	Results and Discussion.....	74
4.3.1	<i>Optimization of the 12-6-4 Potentials</i>	74
4.3.2	<i>Geometries of Metalloproteins</i>	75
4.3.3	<i>Binding Free Energy Calculations</i>	76
4.3.4	<i>Apo State Discussion</i>	81
4.4	Conclusions.....	84
APPENDICES.....		86
APPENDIX: Tables.....		87
APPENDIX: Figures.....		123
BIBLIOGRAPHY.....		129

LIST OF TABLES

Table 1. Root mean square deviation (RMSE) of the 4 repeated PMF simulations of each ligand based on the average value for both benchmark system and test system of OAs and CB8..	30
Table 2. Calculated average binding free energy from PMF simulations versus experiment for the benchmark OAs. In total, there are 6 ligands for both OA and TEMOA (see Figure 7).	31
Table 3. Calculated average binding free energies from PMF simulations along with the experimental values for the benchmark CB8 systems (see Figure 7).	33
Table 4. Summary of the binding free energies for the test OA systems (see Figure 6).	35
Table 5. Summary of the scaled (equation 12) binding free energies for the test OA systems (see Figure 6).	36
Table 6. Summary of the binding free energies for the test CB8 system obtained from PMF simulations (see Figure 6).	37
Table 7. Summary of the scaled (equation 13) binding free energies for the test CB8 systems (see Figure 6).	38
Table 8. Summary of the final binding free energies for the benchmark CB8 system.	40
Table 9. Summary of the final binding free energies for the test CB8 systems (see Figure 6).	41
Table 10. Summary of the MUE and RMSE of the eight systems based on $\Delta\Delta G$ values directly obtained from FEP or TI calculations.	50
Table 11. Summary of the MUE and RMSE, R^2 and Kendall's tau coefficient (τ) of the eight systems based on cycle closure $\Delta\Delta G$ values.	51
Table 12. R^2 and Kendall's tau coefficient for the correlation between predicted binding free energies and experimental data for the eight systems; τ represents the Kendall's tau coefficient.	52
Table 13. Estimate of the uncertainty of the calculations.	54
Table 14. The distance, angle and dihedral restraints for the three sets of DDM calculations on the GlxI-Ni ²⁺ system. See Figure 22 for the info of the restraints and protein atoms A, B, and C. Nomenclature :7(THR)@C means backbone carbonyl carbon from number 7 residue, which is a THR amino acid (Threonine).	72
Table 15. The optimized α_0 and C_4 terms. α_0 is the polarizability for atom type of “nd” and “o” for imidazole and acetate, respectively.	74

Table 16. The coordinating bond distances. The simulated bond distance is the average bond distance from the 300 ns free MD simulations.	76
Table 17. Summary of ΔG_A values of the GlxI-Co ²⁺ system by the double decoupling method (DDM). For each system, nine runs were performed using different restraint strength. Set 1, Set 2 and Set 3 are the three sets of DDM calculations starting with the structure and velocity from the last snap-shot of the 100 ns, 200 ns, 300 ns free MD simulations, respectively....	78
Table 18. Summary of the overall averaged ΔG (in kcal/mol) results.....	81
Table 19. Summary of the overall averaged ΔG (in kcal/mol) results considering two possible <i>apo</i> states.....	84
Table 20. The $\Delta\Delta G$ values directly obtained from the TI calculations as well as the cycle-closure $\Delta\Delta G$ values as mentioned in section 3.2.1.....	87
Table 21. $\Delta\Delta G$ results for the “problematic cases” obtained from TI calculations using the protocol mentioned in section 3.2.3.	98
Table 22. $\Delta\Delta G$ results for the “JNK1” system using the three-step protocol mentioned in section 3.2.4.....	101
Table 23. Summary of the 330 perturbations based on size, ring changes, <i>etc.</i>	102
Table 24. The ΔG values obtained in pKa calculations mentioned in section 4.2.2.3.2.....	120
Table 25. The binding free energies for Ni ²⁺ and Co ²⁺ to imidazole and acetate via PMF calculations mentioned in section 4.2.1.1	121
Table 26. Summary of ΔG_A values of the GlxI-Ni ²⁺ system by the double decoupling method (DDM). For each system, nine runs were performed using different restraint strength. Set 1, Set 2 and Set 3 are the three sets of DDM calculations starting with the structure and velocity from the last snap-shot of the 100 ns, 200 ns, 300 ns free MD simulations, respectively..	121

LIST OF FIGURES

Figure 1. Left: blue and grey curve represents the PMF and the exact opposite biasing potential. Right: the red line represents the biased distribution, in this case, it is equal along the reaction coordinate.....	5
Figure 2. Left: blue and red curve represents the PMF and the proper but arbitrary biasing potential. Middle: the biased distribution histograms. Right: the red line represents the unbiased distribution obtained from WHAM calculation.....	6
Figure 3. The thermodynamic cycle; R is the receptor, L and l are two different ligands, RL represents L bound to R, and Rl represents l bound to R.	12
Figure 4. Selected non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs) from the work of Jorgensen and co-workers.	17
Figure 5. Left: the bonded model; middle: the cationic dummy atom model; right: the non-bonded model. The metal ion is colored in light blue, and the coordinating atoms are colored in red. The dummy atoms in the cationic dummy atom model are in white.....	23
Figure 6. Structures of host OA, TEMOA and their guest molecules. a) Side view and top view of each host. Carbon, nitrogen, oxygen, hydrogen atoms are colored cyan, blue, red, white, respectively; b) the 8 common ligands for OA and TEMOA; c) the 11 ligands for CB8. Protonation states are indicated in the figure.....	28
Figure 7. The structures of the ligands in the benchmark systems. a) Left panel is the 6 common ligands for OA and TEMOA; b) Right panel is the 12 ligands for CB8. Protonation states are indicated in the graph.....	29
Figure 8. Correlation between binding free energies obtained from PMF simulations and experiment.....	32
Figure 9. Correlation between binding free energies calculated from PMF simulations and experiment.....	33
Figure 10. Comparison of the errors of the PMF obtained values and scaled values for the test OAs. On the X axis: 1-8 represents ligands 1-8 binding to OA, while 8-16 represents ligands 1-8 binding to TEMOA.....	36
Figure 11. Comparison of the errors of the PMF values and scaled values for the test CB8. X axis: 1-11 represents ligands 1-11 (see Figure 6).....	39
Figure 12. Correlation between binding free energies calculated from PMF simulations and experiment.....	40

Figure 13. Comparison of the errors of the PMF values and scaled values for the test set for CB8. X axis: 1-11 represents ligands 1-11 (see Figure 6).	42
Figure 14. a) Structure of ligand 5 bound to CB8 system. Carbon, nitrogen, oxygen, hydrogen atoms are colored cyan, blue, red, white, respectively; b) free energy profile CB8-ligand 5 reverse PMF. the reaction coordinate is the distance between center of mass of CB8 and the N2 atom of ligand 5; c) the transition structures from the global minimum to the local minimum for b). CB8 is colored orange; d) free energy profile CB8-ligand 5 reverse PMF. the reaction coordinate is the distance between center of mass of CB8 and the N3 atom of ligand 5; c) the transition structures from the global minimum to the local minimum for d). CB8 is colored orange.....	44
Figure 15. Thermodynamic cycle used for the calculation of the relative binding free energy between protein-ligand system A and protein-ligand system B.	49
Figure 16. Correlation between predicted binding free energies and experimental data for the eight systems.....	52
Figure 17. Correlation between the predicted binding free energies and experimental data for the eight systems studied herein. X axis: Experimental ΔG (kcal/mol); Y axis: Predicted ΔG (kcal/mol). τ is the Kendall's tau coefficient.....	53
Figure 18. Correlation between predicted binding free energies and experimental values for a null model, which has all the $\Delta\Delta G$ set to 0 kcal/mol. X axis: Experimental ΔG (kcal/mol); Y axis: Predicted ΔG (kcal/mol).	58
Figure 19. Left: The binding site structure of GlxI in Co^{2+} bound (<i>holo</i>) and <i>apo</i> form. The Co^{2+} (pink) and its coordinating residues (two units each of HIS, GLU and water) are shown in a ball and stick representation. Right: scheme of calculating the binding free energy of Co^{2+} . HID: neutral form of HIS that is protonated at the δ nitrogen; HIP: +1 charged HIS that is protonated both at both the δ and ϵ nitrogens.	63
Figure 20. Comparison of potential of mean force (PMF) profiles for the default 12-6-4 and optimized m12-6-4 pairwise parameters for the Co^{2+} ion interacting with imidazole and acetate.	65
Figure 21. Free energy profiles calculated with the default and the optimized alpha values for Ni^{2+} acetate and Ni^{2+} imidazole complexes.....	65
Figure 22. (a) Scheme of the DDM method. Dummy atom is an atom that has no interaction with the surroundings, so it can be viewed as the metal ion in gas phase. (b) Scheme of calculating the ΔG_{PO} . Dashed lines mean that the metal ion is restraint to the binding site by restraining to three of the protein atoms through distance, angle and dihedral restraints.....	68
Figure 23. Geometries of Co^{2+} bound protein obtained after 300 ns MD simulations with (a) default 12-6-4 and (b) m12-6-4 potential aligned with crystal structure (light orange). The RMSD measurements are based on the side chain of the two HIS, two GLU, two water molecules along with the metal ion.....	75

Figure 24. The distance, angle and dihedral restraints for the three sets of DDM calculations on the GlxI-Co ²⁺ system. r0, θ_0 , ϕ_0 is the equilibrium distance, angle and dihedral values. ‘:7(THR)@C’: backbone carbonyl carbon of number 7 residue, which is a THR (Threonine) amino acid. Set 1, Set 2 and Set 3: the three sets of DDM calculations starting with the structure and velocity from the last snapshot of the 100 ns, 200 ns, 300 ns free MD simulations, respectively.....	77
Figure 25. Distributions of the selected distance, angle and dihedral in the free MD simulations for Glx-Co ²⁺ . The averaged values were used as the equilibrium distance, angle and dihedral values for the three set of DDM calculations for the default and optimized 12-6-4 potential.	77
Figure 26. Distributions of the selected distance, angle and dihedral in the free MD simulations for Glx-Ni ²⁺ . The averaged values were used as the equilibrium distance, angle and dihedral values for the three set of DDM calculations for the default and optimized 12-6-4 potential.	78
Figure 27. Scheme for computing the free energy change (ΔG_B) associated with the protonation of the HIS6 based on a pK _a shift calculation. represents the protein. ^a approximate value, see reference. ²⁵⁸	80
Figure 28. Upper panel: geometries of apo protein obtained after 100 ns MD simulations with different HIS protonation states aligned with the <i>apo</i> crystal structure (light orange); lower panel: RMSD of the heavy atoms of the two HIS and the two GLU in the binding site comparing to the apo crystal structure (PDB ID: 1fa6).	82
Figure 29. The calculated pKa of the HIS6 and HIS212 at both the ϵ and the δ positions. pKa ^C is the value calculated by TI using method similar to Figure 27; pKa ^H is the value estimated by H++ server.	83
Figure 30. (a) The snapshot after 100 ns MD simulation with the two HIS residues being HIP6_HID212, the blue ball represents the Na ⁺ ion that came into the binding site; (b) the bottom part is more open to solvent, and the red circled hydrogen is the hydrogen on the δ nitrogen position of HIS212.....	83
Figure 31. The perturbation graph plotted based on the work of Wang et.al ⁸ for the GPU TI study in Chapter 2.....	123
Figure 32. Geometries of Ni ²⁺ bound protein obtained after 300ns MD simulations with (a) default 12-6-4 and (b) m12-6-4 potential aligned with crystal structure (light orange). The RMSD measurements are based on the side chain of the two HIS and the two GLU, plus the metal ion and the two water molecules.....	128

KEY TO ABBREVIATIONS

ABFE	Absolute binding free energy
ADME	Adsorption, distribution, metabolism, excretion
AMBER	Assisted Model Building with Energy Refinement
BAR	Bennett acceptance ratio
CADD	Computer-aided drug design
DDM	Double decoupling method
FEP	Free energy perturbation
GAFF	General AMBER force field
GBSA	Generalized Boltzmann surface area
GlxI	Glyoxalase I
GPU	Graphics processing unit
HFE	Hydration free energy
HID	Neutral Histidine residue protonated at δ position
HIE	Neutral Histidine residue protonated at ϵ position
HIP	Protonated Histidine residue
HIS	General name for Histidine residue
IOD	Ion-oxygen distance
LJ	Lennard-Jones
LOMAP	Lead optimization mapper
MC	Monte Carlo
MD	Molecular dynamics

MM	Molecular mechanics
MUE	Mean unsigned error
NPT	Constant number, pressure, and temperature
NVT	Constant number, volume, and temperature
PBSA	Poisson-Boltzmann surface area
PDB	Protein databank
PME	Particle mesh Ewald
PMF	Potential of mean force
QM	Quantum mechanics
RBFE	Relative binding free energy
REM	Replica exchange method
RESP	Restrained electrostatic potential
REST	Replica exchange with solute tempering
RMSD	Root mean square deviation
RMSE	Root mean square error
SMIRNOFF	SMIRKS Native open force field
SP	Standard precision
TI	Thermodynamic integration
TM	Transition metal
US	Umbrella sampling
vdW	van der Waals
WHAM	Weighted histogram analysis method

CHAPTER 1: INTRODUCTION

1.1 General Introduction

High binding free energy is one of the major requirements for an effective drug molecule along with factors associated with ADME/tox (adsorption, distribution, metabolism, excretion, and toxicity) considerations.¹⁻⁷ Because of this much effort has been expended to develop fast and accurate computational free energy methods to predict protein-ligand binding free energies⁸⁻¹³ with the result being the generation of many approaches with a wide range of effectiveness.¹⁴⁻⁴²

One category of free energy methods is the end-point methods, which usually decompose free energy into contributions from enthalpy and entropy.^{14-24, 41} Among these methods, the Molecular Mechanics-Poisson-Boltzmann/Surface Area (MM-PBSA) and the Molecular Mechanics-Generalized Born/Surface Area (MM-GBSA) are widely used.¹⁴⁻¹⁷ In these two approaches, explicit molecular dynamic (MD) simulations on the ligand-bound state or both the bound and unbound state are performed and post-processed to obtain the enthalpy difference between the bound and unbound state. The entropy change is estimated by normal mode analysis or other approaches.²³ Combining the solvation free energy obtained from PB or GB equation, one can estimate the binding free energy. Numerous approximations are made in end-point methods, which can result in significant uncertainties in the enthalpy, entropy, and the solvation free energy term. Another category of free energy methods is the alchemical methods, which define an alchemical pathway to permute one ligand to another. They have been of significant interest for many decades but in recent years, with the advent of advanced technologies to enhance sampling and force field representations, they have gained renewed favor. The theoretical groundwork for these approaches was laid out by Kirkwood, Zwanzig and Bennett among others.²⁵⁻³¹ Thermodynamic Integration (TI),²⁶ Free Energy Perturbation (FEP),²⁸ and Bennett Acceptance Ratio (BAR)^{25, 27} are all widely

used to perform relative binding free energy (RBFEE) calculations as well as absolute binding free energy (ABFE) calculations.

Pathway methods serve as another type of free energy methods and they are usually used to calculate ABFEs.^{32-38, 40} By definition, the ligand is pulled out of and/or pushed into the binding site, and the reversible work of the pulling and/or pushing process is computed to estimate the ABFEs. Within this context, many methods have been developed, including umbrella sampling (US)⁴³, the Jarzynski method⁴⁴, and metadynamics⁴⁵, *etc.* They can be categorized into non-equilibrium methods⁴⁶⁻⁵⁰ and equilibrium methods.^{40, 51-53} The steered molecular dynamics⁵⁴ is one of the non-equilibrium methods and it is based on the Jarzynski approach⁴⁴. The equilibrium methods require the system to achieve equilibrium during the pulling or pushing process. The US method is one of the widely-used equilibrium methods.

In the following section (section 1.2), the US method is introduced in more detail. In section 1.3, the evolution of alchemical free energy methods in computer-aided drug design (CADD) is reviewed. In Chapter 2, the US method is applied to compute the ABFEs of host-guest systems in a blind challenge, *i.e.* the Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL6) challenge. Host-guest systems serve as great model systems for the community to test and improve the free energy methods, due to several advantages including small size hence longer simulations are achievable, no slow-timescale motions of hosts, *etc.* In these US studies, we found large systematic errors, and by performing benchmark calculations on known systems and by scaling the obtained binding free energies, the errors can be largely reduced, especially when the benchmark systems and the unknown systems consist of congeneric ligands. The source of error could be from the force constant used in US, window overlap issues as well as force field, sampling issues, *etc.* The results we obtained by scaling ranked top for one of the host-guest systems and

ranked 4th overall for both systems. In Chapter 3, we present the alchemical free energy studies on a large protein-inhibitor data set covering 8 protein systems, 199 ligands, and 330 perturbations. We computed the RBFs using graphics processing unit accelerated thermodynamic integration (GPU-TI) on the data set originally assembled by Schrödinger, Inc. Using their GPU free energy code (FEP+) and the OPLS2.1 force field combined with the REST2 enhanced sampling approach, these authors obtained an overall MUE of 0.9 kcal/mol and an overall RMSE of 1.14 kcal/mol. In our study using GPU-TI from AMBER (version 18) with the AMBER14SB/GAFF1.8 force field but without enhanced sampling, we obtained an overall MUE of 1.17 kcal/mol and an overall RMSE of 1.50 kcal/mol for the 330 perturbations contained in this data set. A more detailed analysis of our results suggested that the observed differences between the two studies arise from differences in sampling protocols along with differences in the force fields employed. Future work should address the problem of establishing benchmark quality results with robust statistical error bars obtained through multiple independent runs and enhanced sampling, which is possible with the GPU-accelerated features in AMBER.

Metal ion modeling in biology is of great significance in CADD because more than 25% of proteins need metal ions, and many of the proteins have metal ions in the active site where the drug ligands bind, for example, the histone deacetylase (HDAC). In section 1.4, we introduce a metal ion model that was developed recently in our lab, *i.e.* the 12-6-4 Lennard Jones (LJ) nonbonded model. The previous works in our lab have shown its ability to reproduce both thermodynamics and structure properties of metal ion interacting with water⁵⁵⁻⁵⁸ and small ligands⁵⁹. In Chapter 4, we show that the 12-6-4 LJ nonbonded model could be extended to modeling transition metal (TM) ions in proteins, which are critical in biology but very challenging to model. The thermodynamics of TM ions (Co²⁺, Ni²⁺) binding to glyoxalase I (GlxI) as well as the structural features are reproduced

using the optimized 12-6-4 (m12-6-4) potential. Critically, this model simultaneously reproduces the solvation free energy of the individual TM ions, the thermodynamics of TM ion-ligand coordination as well as the thermodynamics of TM ion binding to a protein active site, unlike extant models. We find the incorporation of the thermodynamics associated with protonation state changes for the TM ion (un)binding to be crucial. The high accuracy of m12-6-4 potential in this study presents an accurate route to explore more complicated processes associated with TM-based drug design in metalloprotein platforms.

1.2 Umbrella Sampling and WHAM

Umbrella sampling method divides the physical pathway into successive simulation windows with a biasing force restraining the ligand to a certain position in each window. With the biased distribution data from each window a potential of mean force (PMF) of binding or unbinding is constructed and used to estimate the binding free energy. The weighted histogram analysis method (WHAM)^{51, 60} is usually used to construct the PMF, as well as BAR and FEP.^{33, 36-37, 40, 61} If the biasing potential is exactly the negative of the PMF, see Figure 1, then the simulations would give an equal distribution along the reaction coordinate. However, the PMF is not known *a priori*. In practice one would start with some proper but arbitrary biasing potential, then with the information of the biasing potential and the biased distribution, WHAM could be used to construct the PMF, see Figure 2. The harmonic potentials (colored in red) are used to bias the simulation in order to simulate conformations of higher free energy. The biased distribution is given in the middle. To convert the biased distribution to unbiased distribution, WHAM divides each simulation into M bins. Assuming the total number of simulations is S , for bin j of simulation i , the unbiased distribution p_j^0 can be calculated by the following equations,

$$p_j^0 = \frac{\sum_{i=1}^S n_{ij}}{\sum_{i=1}^S N_i * f_i * c_{ij}} \quad (1)$$

$$f_i^{-1} = \sum_{j=1}^M c_{ij} * p_j^0 \quad (2)$$

$$c_{ij} = \exp\left(-\frac{V_i(x_j)}{k_B T}\right) \quad (3)$$

where n_{ij} is the number of data points in bin j of simulation i , N_i is the number of total data points in simulation i , f_i is the free energy shift from simulation i , c_{ij} is a constant calculated from the biasing potential in bin j of simulation i , *i.e.* $V_i(x_j)$, k_B is the Boltzmann constant and T is the simulation temperature. With the biased distribution data, n_{ij} can be obtained, and with the predefined $V_i(x_j)$, p_j^0 can be solved from equations 1 and 2 iteratively until convergence is reached for f_i . With the unbiased distribution, the PMF can be constructed. For WHAM to work properly, the biasing potential for each window should be properly selected or optimized so that the biased histograms overlap between neighboring windows and the histogram for each window is centered at the restrained position.

Figure 1. Left: blue and grey curve represents the PMF and the exact opposite biasing potential. Right: the red line represents the biased distribution, in this case, it is equal along the reaction coordinate.

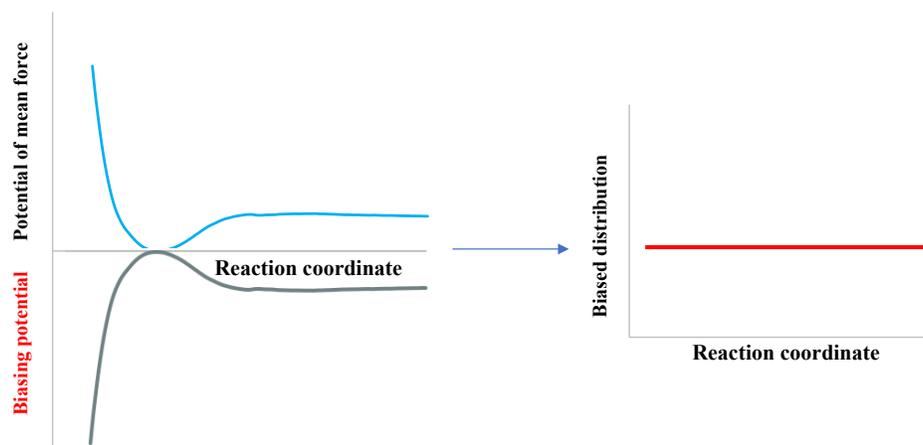
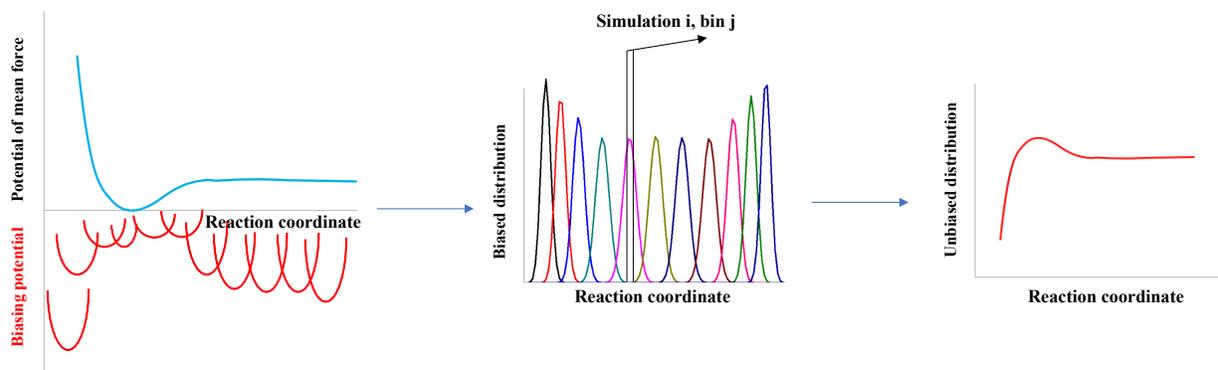


Figure 2. Left: blue and red curve represents the PMF and the proper but arbitrary biasing potential. Middle: the biased distribution histograms. Right: the red line represents the unbiased distribution obtained from WHAM calculation.



1.3 Alchemical Free Energy Methods in Computer Aided Drug Design

1.3.1 Theoretical Foundation

Albeit it was in the 1980s when the free energy perturbation (FEP) method came to the attention of the computer aided drug design (CADD) field, its theoretical foundation was laid 30 years earlier. During his theoretical exploration of the properties of nonpolar gases, Zwanzig derived the master equation for FEP,²⁸ though studies employing thermodynamic perturbation theory can be traced back to R.E. Peierls in the 1930s.⁶²⁻⁶³ The inherent beauty of the FEP equation (see equation 4) lies in the fact that it allows for the computation of the free energy difference between two ensemble states by sampling the configurations of one state and then calculating the potential energy for both states in the sampled configurations. $\langle \rangle_0$ represents taking the Boltzmann average at state 0. On the face of it the use of this master equation appears straightforward, but the convergence of FEP calculations is not easily achieved, especially for systems where large changes result in ΔA values that are a few times $k_B T$. However, the idea of using a coupling parameter, described by Kirkwood in 1935²⁶, greatly benefited FEP calculations because the perturbations between neighboring states become much smaller. Briefly, a series of intermediate states are

introduced via a coupling parameter λ , and the summation of all the ΔA s between the neighboring states results in the total ΔA . A typical form of the potential function used to define the intermediate states as a function of λ is a linear combination of the potential of the reference state U_0 with that of U_1 (see equation 5). Another way to efficiently estimate free energy differences is the Bennett acceptance ratio (BAR), which was developed by Bennett in 1976.²⁵ By analogy to the FEP equation, he derived equation 6, where a weighting function (w) was introduced. By minimizing the expected square error of the free energy difference, Bennett obtained the optimal weighting function leading to equation 7 that can be solved by iterative trials on ΔA . Different from Zwanzig's equation, the BAR analysis requires sampling configurations at both states. The other conventional method, thermodynamic integration (TI), was predicated on Kirkwood's work on the theory of liquids.⁶⁴ It requires the calculation of the Boltzmann averaged potential energy derivative at each intermediate state λ (see equation 8).⁶⁵

$$\Delta A = A_1 - A_0 = -k_B T \ln \left\langle \exp\left(-\frac{U_1 - U_0}{k_B T}\right) \right\rangle_0 \quad (4)$$

$$U_\lambda = \lambda U_0 + (1 - \lambda) U_1 \quad (0 \leq \lambda \leq 1) \quad (5)$$

$$\Delta A = A_1 - A_0 = -k_B T \ln \frac{\langle w \exp(-\beta U_1) \rangle_0}{\langle w \exp(-\beta U_0) \rangle_1} \quad (6)$$

$$\Delta A = A_1 - A_0 = -k_B T \ln \frac{\left\langle \frac{1}{1 + \exp(-\beta(U_1 - U_0) + \beta(A_1 - A_0))} \right\rangle_0}{\left\langle \frac{1}{1 + \exp(\beta(U_1 - U_0) - \beta(A_1 - A_0))} \right\rangle_1} \quad (7)$$

$$\Delta A = A_1 - A_0 = \int_0^1 \left\langle \frac{dU_\lambda}{d\lambda} \right\rangle_\lambda d\lambda \quad (8)$$

1.3.2 Enabling Technologies

Based on the theoretical foundation, performing free energy calculations requires generating configurations and evaluating the corresponding potential energies. Hence, the development of

simulation techniques and force fields served as key factors for the implementation of free energy methods. One major sampling technique is Monte Carlo simulation. Briefly, a Monte Carlo simulation starts with a given configuration and either accepts or rejects the next proposed configuration in a way that the final sampled configuration recovers the true distribution of the actual ensemble. The Monte Carlo method can be traced back to Enrico Fermi, although the earliest published working examples were performed by Ulam and Metropolis.⁶⁶⁻⁶⁸ The first paper that applied the Monte Carlo method to molecular studies was published in 1954, and Monte Carlo simulations on proteins were reported in the 1970s.⁶⁹⁻⁷⁰ Another major sampling technique is molecular dynamics (MD) simulations. MD simulation involves sampling configurations along a time axis. In 1964, Rahman published the landmark MD simulation on liquid argon, where he used a Lennard-Jones potential to describe the argon atoms. With the development of force fields suitable for biomolecules⁷¹, the first MD simulation on a protein⁷² was reported in 1977 by Karplus, McCammon and Gelin, who developed the simulation program based on Levitt and Lifson's original code⁷³. Berendsen, in the same year, developed the "SHAKE" algorithm⁷⁴ that constrains the high frequency heavy atom-hydrogen bonds, and this contribution made it possible to increase the time step used in MD simulations by a factor of two allowing for more extensive sampling given available computer resources.

The development of biomolecular force fields was based on the pioneering work of the Lifson laboratory.⁷¹ Based on the potential energy function (see equation 9), early force fields and programs were developed in the 1980s, including CHARMM⁷⁵ by Karplus *et al*, AMBER⁷⁶⁻⁷⁷ by Kollmann *et al*, OPLS⁷⁸ by Jorgensen *et al*, and GROMOS⁷⁹ by Berendsen, van Gunsteren, *et al*. Briefly, the bond, angle, dihedral, charge and Lennard-Jones (LJ) parameters were optimized to reproduce either quantum mechanical (QM) or experimental properties. Overviews of the current

state-of-the art in force fields as well as detailed historical perspectives can be found in a number of reviews.⁸⁰⁻⁸¹

$$U = \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedral}} K_\phi (1 + \cos(n\phi - \delta)) + \sum_{\text{impropers}} K_{\text{imp}} (\varphi - \varphi_0)^2 + \sum_{\text{nonbond}} \left(\epsilon_{ij} \left[\left(\frac{R_{\text{min}}^{ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{\text{min}}^{ij}}{r_{ij}} \right)^6 \right] \right) + \frac{q_i q_j}{\epsilon r_{ij}} \quad (9)$$

Another important enabling technology for free energy calculations in CADD was the development of robust and accurate water models. Both absolute and relative binding free energies of a drug is affected by solvent; hence, the accurate modeling of water and the solvation of small molecules is crucial. Implicit water models such as Poisson-Boltzmann⁸²⁻⁸⁴ or Generalized Born⁸⁵⁻⁸⁶ models can be used to model the effect of solvent on drug binding, but many water-receptor or water-drug interactions are specific, necessitating the need to use explicit water molecules to obtain the highest quality results. For example, the displacement of active site water can play outsized roles in free energy prediction.⁸⁷⁻⁸⁸ Depending on how many additional charged dummy atoms are utilized, water models can be divided into three-site, four-site, five-site and six-site. The earliest water models were the BF water model⁸⁹ described by Bernal and Fowler in 1933, the BNS water model⁹⁰ by Ben-Naim and Stillinger in 1972, and the ST2 water model⁹¹ by Rahman and Stillinger in 1974. Jorgensen has been a pioneer in developing practical and accurate water models for use in molecular simulations. He developed the TIPS model⁹² (transferable intermolecular potentials) to create solvent models for water, alcohol and ether simulations. Reasonable structural and energetic results for both gas-phase dimers and pure liquids were achieved through MC simulations. Also in 1981, Berendsen and coworkers devised another water model they called the SPC⁹³ (simple point charge) water model. The parameters were fitted to reproduce both the experimental interaction energy and the pressure of the liquid at 300K. Jorgensen continued to develop and revise the TIPS water model creating the TIPS2, TIP3P and TIP4P water models.⁹⁴⁻⁹⁵

These water models were all compared to a range of experimental structural and thermodynamic properties which served as a foundation for the future evaluation of water models. Berendsen evolved his SPC model, by including the self-energy correction, creating the SPC/E water model⁹⁶. To date, the TIP3P and SPC/E water models are the most widely used, but several water models have been developed over the years. For those interested in further details, several reviews are extant that summarize the current state of water model development.^{82, 97-100}

1.3.3 Early FEP studies

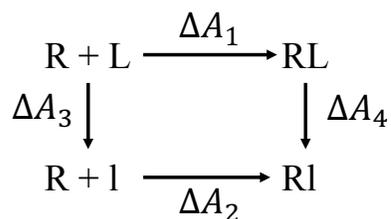
With all the enabling techniques in place applications of free energy methods could begin in earnest. Postma, Berendsen and Haak in 1982,¹⁰¹ reported free energy calculations on the formation of a cavity in water. Soon after this work, Jorgensen performed FEP calculations on representative molecules in 1985.¹⁰² He computed the relative free energies of hydration of methanol and ethane in dilute solution, and the result was, excitingly, in good accord with experiment. This hinted at the potential of FEP calculation in CADD, since relative solvation free energies plays a major role in determining the relative binding free energy of two ligands at a common receptor site. The optimized potential functions for water⁹⁵, ethane¹⁰³ and methanol¹⁰⁴, of course, were critical to the success of this work. Besides the application of FEP on relative hydration free energy calculations, Jorgensen also applied FEP to construct a potential of mean force (PMF) profile for the S_N1 reaction (disassociation of (CH₃)₃CCl),¹⁰⁵ opening the door to the study of solvation effects on reactive processes. Kollman and co-workers implemented FEP into their MD simulation program, *i.e.* AMBER, in 1985. Their FEP results on the perturbation of methanol to ethane, hydronium to ammonium, glycine to alanine and alanine to phenylalanine in water also showed reasonable agreement with available experimental data.¹⁰⁶

The early application of FEP calculations explored the relative solvation free energy of

organic molecules (or ions) and ion-association reactions. These studies showed the potential of FEP on different systems. However, a major contribution that enabled the wide application of FEP to CADD is the realization of the power of the thermodynamic cycle to simplify the calculations for relative drug binding calculations. The upper and lower process shown in Figure 1 represent the absolute binding free energy for ligand L and ligand I to a receptor site R, respectively. These calculations are still challenging to carry out and are quite computationally expensive, so a simplification was needed. In ligand optimization efforts, simple modifications to a scaffold are typically made (for example a -H is replaced by -CH₃), so in the end what is important is the relative free energy change upon these chemical mutations. By connecting the upper and lower processes of Figure 3 we find that $\Delta\Delta A = \Delta A_1 - \Delta A_2$ can be obtained by the alchemical transformations represented by the two vertical processes yielding $\Delta\Delta A = \Delta A_3 - \Delta A_4$. Hence, the relative binding free energy of the two ligands can be obtained through FEP calculations on the two alchemical processes: $R + L \rightarrow R + I$ and $RL \rightarrow RI$. Transforming the problem in this manner converts two computationally expensive calculations into two computationally tractable FEP calculations. Tembe and McCammon first noted the concept of the thermodynamic cycle in their paper “Ligand-receptor interactions” published in 1984.¹⁰⁷ In this paper, they designed a model system and used the thermodynamic cycle coupled with FEP calculations to compute the $\Delta\Delta A$ of two “ligand” atoms. Their result was in close agreement with the $\Delta\Delta A$ obtained by directly calculating the ΔA_1 and ΔA_2 free energies using the umbrella sampling method. McCammon and co-workers also applied “the thermodynamic cycle-perturbation method” to compute the relative hydration free energy of Cl⁻ and Br⁻ ions¹⁰⁸, and showed good agreement with experiment. More importantly this approach was applied to the computation of the relative binding free energy of Cl⁻ and Br⁻ ions to the organic receptor SC24¹⁰⁹ and also showed good agreement with experiment.

This work, published in 1986 by Lybrand, McCammon, and Wipff, demonstrated, for the first time, the applicability of FEP on host-guest systems via a thermodynamic cycle. It's worth noting that the work by Jorgensen in 1985 on relative hydration free energy of methanol and ethane¹⁰² also utilized the concept of thermodynamic cycle. By considering the ligands as rigid ligands, *i.e.* having no intramolecular degrees of freedom, the relative solvation free energy was obtained by perturbing one ligand to the other in aqueous solution, leaving out the perturbation in the gas phase because its value was zero by design. Besides making these calculations more tractable, the thermodynamic cycle concept also affords error cancellation that improves the quality of the computed free energies.¹¹⁰⁻¹¹³

Figure 3. The thermodynamic cycle; R is the receptor, L and l are two different ligands, RL represents L bound to R, and Rl represents l bound to R.



Besides computing the relative free energies, the thermodynamic cycle was also utilized to compute absolute binding free energies. In his work, Kollman calculated the free energies of association of nucleic acid bases *in vacuo* by disappearing one of the nucleic acid bases using FEP. Combining this with the hydration free energy of both bases, he was able to obtain the absolute binding free energy of the two bases in water via a thermodynamic cycle.¹¹⁴ Jorgensen also utilized the thermodynamic cycle and derived the so-called double decoupling method (DDM). Briefly, FEP calculations on the two alchemical processes, *i.e.* $RL \rightarrow R$ and $L \rightarrow 0$, were performed to obtain the absolute binding free energies of the methane dimer.¹¹⁵ DDM has been further

developed^{40,116} and provides the community an alternative approach to obtain the absolute binding free energies for drug bound to receptor. Kollman and Jorgensen's work represent the earliest efforts to use FEP to obtain absolute binding free energies.

1.3.4 Modern FEP in CADD

As highlighted above, the work of McCammon and coworkers on halide ions binding to the SC24 host-guest system¹⁰⁹ demonstrated the applicability of FEP on host-guest systems via a thermodynamic cycle. In the same year, McCammon and co-workers extended their study to enzyme-inhibitor systems. Their work, "Dynamics and Design of Enzymes and Inhibitors" received by JACS in January 1986, was the first modern FEP study in CADD.¹¹⁷ In this study, McCammon computed the relative binding free energy of p-fluorobenzamidine and benzamidine binding to trypsin, and the relative binding free energy of benzamidine to native and mutant trypsin. The computed $\Delta\Delta A$ values, though having somewhat large uncertainties due to the short simulations, agreed with experimental values, demonstrating the applicability of "the thermodynamic cycle-perturbation method" in CADD. The simulations were performed using the GROMOS program. In late 1986, with the FEP method implemented in the AMBER program, Kollman studied a pair of inhibitors (*i.e.* phosphoramidate and phosphonate ester) binding to thermolysin.¹¹⁸ The computed $\Delta\Delta A$ values, 4.21 ± 0.54 kcal/mol, agreed well with the experimental value of 4.10 kcal/mol. For each perturbation, the inverse transformation was also carried out to ensure the convergence of the computed results. The small uncertainties (13% of $\Delta\Delta A$) and close agreement with experiment was encouraging. Importantly, in both the studies of McCammon and Kollman, the relative solvation free energy of the inhibitors contributed significantly to the computed $\Delta\Delta A$, which further illustrates the importance of accurately describing the interactions between the organic molecules and water. Subsequently, Kollman and co-workers extended their

study to the effect of site-specific mutagenesis on enzyme catalysis, in addition to ligand binding.¹¹⁹ They computed the relative activation free energy of a tripeptide substrate by native and mutant subtilisin, as well as the relative binding free energy. They correctly predicted the small difference in the binding free energy, as well as the substantial difference in the activation free energy, before the experimental result was disclosed. After the initial studies in the 1980s, FEP studies in CADD expanded in 1990s, with studies on *Rhizopus* pepsin, Chymotrypsin, Elastase, HN-1 Protease, Carbonic Anhydrase II, Dihydrofolate Reductase, and T4 Lysozyme being reported.¹²⁰⁻¹²¹ Although, MC simulation was successful in early FEP studies of host-guest systems and relative solvation free energy calculations, its usage in protein simulation was limited due to issues surrounding sampling efficiency. A small trial move of the backbone angle in one protein region could result in a large movement of a remote site, hence, the overall acceptance rate will be small. The first two modern MC FEP studies in CADD were performed by Jorgensen in 1997.¹²²⁻¹²³ Good agreement with the experiment was achieved, but both studies kept the protein backbone fixed.

1.3.5 From retrospective to prospective

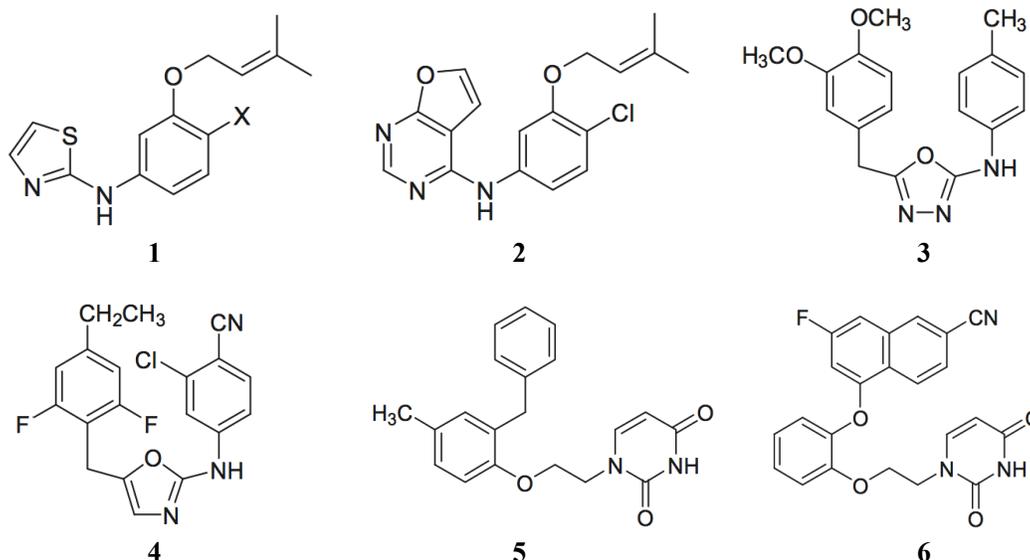
Even though the early work of McCammon, Kollman, Jorgensen, *etc.* showed the power of free energy methods in CADD, they were retrospective studies where the experimental values was known beforehand. In CADD applications the experimental result is not known, so in order for these methods to find a niche in CADD it was essential to demonstrate that it could make accurate prospective predictions. In 1989, Merz and Kollman first utilized FEP for prospective study in CADD.¹¹ For the first time, the binding free energy of a new inhibitor was predicted and subsequently validated by experiment. The protein they studied was thermolysin and the inhibitors had the general structure of carbobenzoxy-Gly^p(X)-Leu-Leu, where X = NH, O, and CH₂. The NH

and O compounds were studied in earlier work¹¹⁸, but their experimental binding free energies were known beforehand. For the new inhibitor, *i.e.* the CH₂ compound, Merz and Kollman correctly predicted that it had similar binding free energy as the NH compound (0.0 kcal/mol computed *versus* -0.3 kcal/mol experiment), which surprisingly was larger than the binding free energy for the O compound. It is worth mentioning that they also explored the influence of charge choice for the drugs and key torsional parameters on the computed $\Delta\Delta A$ values. Optimization of torsional parameters and charge sets is still an ongoing effort for improving the precision of FEP in CADD.⁸ Two years later, in 1991, Kollman and coworkers performed the second prospective FEP study in CADD.¹²⁴ They studied peptide inhibitors binding to HIV-1 protease while the experimental determination of the $\Delta\Delta A$ values was ongoing. The predicted $\Delta\Delta A$ for the S and R diastereomers of the inhibitor, JG365, was consistent with experiment.

To date, a wide range of prospective CADD studies has been reported. It is not the aim of the current viewpoint to exhaustively enumerate all extant studies; nonetheless, we'll highlight a few examples. Starting in the early 21st century, Jorgensen and coworkers have been utilizing FEP to guide the design of non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs), and they also measured the activities (EC50) and cytotoxicities (CC50) experimentally to verify predictions from FEP calculations.¹²⁵⁻¹³⁹ In 2006, starting with the lead compound **1** (Het-NH-PhX-U), see Figure 4, MC/FEP guided the design of several NNRTIs that have similar activities and cytotoxicities as two FDA approved NNRTIs, *i.e.* efavirenz and etravirine, via permutations of the Het and X motifs in **1**.¹³¹⁻¹³⁴ One of the novel NNRTIs is compound **2**, which has an EC50 of 5 nM and CC50 of 17 μ M. The potency was greatly improved from the initial lead compound **1**, which has a 10 μ M EC50. Starting in 2008, important mutations were included to assay the activity of the NNRTIs, such as Y181C. The best of the previous compounds were found to have low

activities or were inactive towards these viral mutations. However, Jorgensen and coworkers were able to use MC/FEP to guide their discovery on new NNRTIs,^{128, 137} starting from a new lead (compound **3**) discovered in 2007, which has the U-Het-NH-PhX structural motif.¹³⁶ The lead compound **3** was a false positive compound obtained from virtual screening, but perturbations on the U, Het and X groups lead to a true positive compound **4**, which has 1.3 nM and 6.9 nM EC50 against wild type and the Y181C mutant transcriptase. In 2009, starting with a lead compound (compound **5**) that was screened targeting Y181C, MC/FEP guided them to promising compounds that are active against both wild type and important mutants, *i.e.* Y181C and K103N/Y181C, via perturbations in the substitution pattern, linker region, and the heterocycle. One example is compound **6**, which has an EC50 of 1.1 nM, 8.0 nM, 6.0 nM for the wild type, Y181C mutant and K103N/Y181C mutant transcriptase, and its CC50 was greater than 100 μ M. Again significant improvement from compound **5** was achieved, which has an EC50 of 4.8 μ M for the wild type and no activity for the two mutations. Through years of effort, MC/FEP has helped Jorgensen and coworkers to improve the EC50's against NNRTIs from the μ M level to nM level for both wild type and important mutant transcriptase. Preclinical trials on mice showed great potency of the designed NNRTIs.¹²⁵ The MC/FEP calculations described above were performed with the MCPRO program.

Figure 4. Selected non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs) from the work of Jorgensen and co-workers.



Recently, using the same program, Lovering and co-workers were able to predict the most active lead compound among the 17 potential drug targets for the spleen tyrosine kinase, and further optimization of the lead compound leads to nM cellular activity.¹⁴⁰ The FEP+ program from the Schrödinger, Inc. has also been employed for prospective studies in CADD. It not only gave ~1 kcal/mol error but also reliably predicted true positive compounds for two prospective drug design projects.⁸ As a final example, recently, Janssen Pharmaceuticals employed the FEP+ to study β -secretase 1 (BACE1) system.¹⁴¹ Modifications on the scaffold and P3 pocket substituent of the inhibitor were evaluated by FEP, and the resulting inhibitors showed nM activity experimentally. Importantly, the average mean unsigned error (MUE) between FEP and experiment was very small: 0.35 ± 0.13 kcal/mol using their computational protocol.

1.3.6 Modern Status

It has been 34 years since the first application of FEP in CADD¹¹⁷, and notable developments have been achieved, enabling the use of free energy methods in CADD. Several major developments include sampling technology improvements, force field improvements, and in the ease of use. In order to achieve convergence and accuracy in FEP calculations, sampling of all the relevant conformations is very important. However, due to the large barriers between different conformations, sampling is often incomplete on the time scale of a typical FEP calculation, hence very often the results depend on the initial structures. One approach to enhance sampling is to couple FEP with the replica exchange method (REM). Temperature REM involves running replicas of MD or MC simulations with different temperatures in parallel and exchanging conformations of a pair of neighboring replicas according to detailed balance considerations.¹⁴² Recently more efficient replica exchange methods have been developed and coupled with FEP.¹⁴³⁻¹⁴⁵ The replica exchange with solute tempering (REST1 and REST2) methods divide the potential energy of a protein-ligand system into three components, *i.e.* the internal energy of the protein, the interaction energy between the protein and water, and the interaction of the water molecules with each other. With different scales designed for the three components of the potential energy of the neighboring replica, rigorous derivation can lead to an acceptance probability (for the exchange of configurations between the two replicas) that does not depend on the water molecules. As a result, fewer replicas for the whole simulation are required, and, hence, sampling is more efficient. For systems where conformational changes were not captured by standard FEP calculations, FEP/REST2 has been shown to resolve the sampling issue and accurately reproduced the experimental $\Delta\Delta A_s$.¹⁴⁵ Another approach to enhance sampling is to use graphical processing units (GPU). Recently both FEP and TI were implemented on GPUs, and with the concomitant two

orders of magnitude speedup, more sampling is achievable.¹⁴⁶⁻¹⁴⁸ With the power of GPUs, FEP and TI calculations on large data sets can be quickly and routinely performed, thereby, enabling the fast identification of potential drug ligands using CADD.^{8, 149-151}

Another major development is in the force fields. General force fields for drug-like organic molecules were first developed the early 21st century, including GAFF¹⁵² for AMBER in 2004 and CGenFF¹⁵³ for CHARMM in 2010. Recently the GAFF 1.8 force field combined with the AMBER protein force field FF14SB was tested on a large data set with 330 perturbations using the AMBER GPU TI implementation.^{149, 151} The overall MUE for the computed $\Delta\Delta A$ s compared to experiment is comparable with the results obtained by the GPU FEP/REST2 calculations with the OPLS2.1 force field.⁸ For the GPU FEP/REST2 calculations, the torsional and covalent parameters were extensively trained (against more than 10,000 representative organic compounds). As was discovered previously, key torsional parameters are important to FEP calculations.¹¹ Continued efforts on optimizing the OPLS and AMBER/GAFF force fields have been reported. By refitting the peptide dihedral parameters along with improvements in ligand charge models *et. al*, Harder and co-authors developed the OPLS3 force field¹⁵⁴ based on the OPLS2.1 force field. Recently, further extensive optimization on the ligand torsion types and optimization of ligand partial charge assignment evolved the OPLS3 force field to the OPLS3e force field.¹⁵⁵ Both OPLS3 and OPLS3e force fields were tested using GPU FEP calculations on the above-mentioned data set, and both showed improvements, with the MUE being 0.77 kcal/mol and 0.78 kcal/mol, respectively. Better correlation with experiment is achieved with the OPLS3e force field, with the R^2 being 0.61. GAFF2¹⁵⁶ was developed based on GAFF, and re-parameterization of the van der Waals, the bond, angle and dihedral parameters were expected to improve the force field, but a robust test of GAFF2 using AMBER GPU TI calculations has not been reported yet. The above-mentioned force fields

use atom types to describe different atoms in different chemical environments, and the different nonbonded and bonded parameters are then assigned based on atom types. Recently, a new force field format, *i.e.* the SMIRKS Native Open Force Field (SMIRNOFF) format was implemented to develop an open force field model.¹⁵⁷ With the SMIRNOFF format, force field parameters were assigned via direct chemical perception instead of based on atom types. The new format has more flexibility and simplicity and can avoid some problematic cases for the atom-type-based force fields.

Last, but not the least, ease of use is another major development. Historically, setting up free energy calculations of all types can be tricky, time consuming and mistake prone, so a key step to make these approaches more accessible was to simplify setting them up, running the simulations and then analyzing the results. For example, in the lead optimization stage, the number of potential ligands considered could range into the hundreds of compounds. However, it turns out that performing all $\Delta\Delta A$ calculations is slow and unnecessary. Liu and coauthors developed the lead optimization mapper (LOMAP) that can automatically design efficient perturbation paths.¹⁵⁸ LOMAP sets up perturbations between ligands based on similarity, and it also connects any two ligands with a minimal number of perturbations, thereby, reducing the error of each perturbation and the chance of error propagation. It also puts every ligand in at least one perturbation cycle so that the thermodynamic cycle closure can be used to validate the accuracy of the calculations. LOMAP was implemented into the Schrodinger FEP+ software and greatly facilitated FEP calculations on large data sets.^{8, 154-155} Besides FEP+, other software, such as Sire¹⁵⁹ and Gromacs¹⁵⁰, are also able to set up automatic FEP workflows. Above all, with the improvement in sampling, force field accuracy, and ease of use, free energy methods are becoming more and more useful in CADD campaigns.

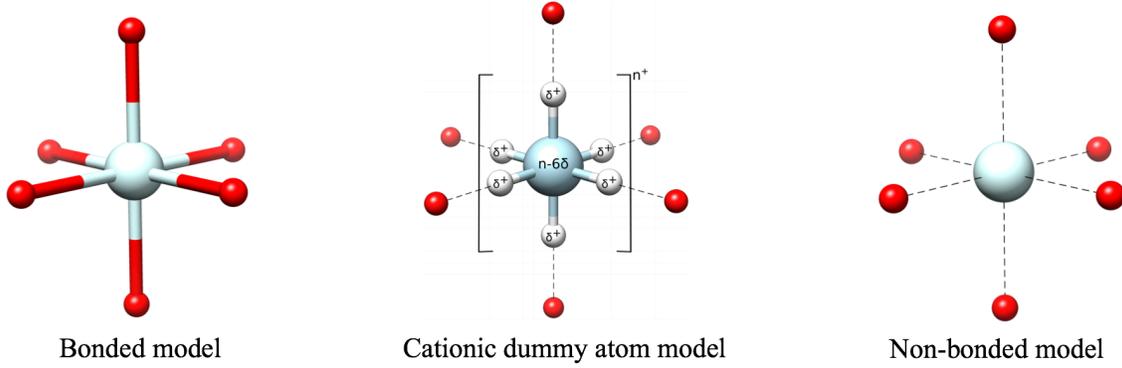
1.4 Metal Ion Force Field: the 12-6-4 Lennard Jones Nonbonded Model

Simultaneous reproducing the thermodynamics (e.g., the hydration free energy) and structural features (e.g., the ion-oxygen distance) of metal ions in the aqueous phase is a challenge⁵⁸. The challenge is even bigger for modeling metal ions in more complicated systems such as metalloproteins. A variety of force field models have been developed for metal ion modeling, such as the bonded model, the nonbonded model, the cationic dummy atom model, the Drude oscillator model¹⁶⁰⁻¹⁶¹, the fluctuating charge model,¹⁶² and ReaxFF¹⁶³, *etc.*. A comprehensive review concerning metal ion modeling has been recently provided by Li and Merz¹⁶⁴.

The bonded model is widely used for metalloprotein modeling¹⁶⁵⁻¹⁷⁰, see Figure 5. In this model, the metal ion is covalently bonded to the coordinating residues. The bond, angle, dihedral, van der Waals (vdW), and electrostatic interactions are described by classical terms. Different schemes have been developed to parameterize these terms¹⁶⁴. Although the bonded model can reproduce experimentally determined structures of metal sites, it cannot simulate ligand exchange and switches in coordination number (CN), which are crucial for modeling catalytic metal centers as well as metal transport. One alternative model is the cationic dummy model proposed by Åqvist and Warshel¹⁷¹⁻¹⁷², see Figure 5. In this model, dummy particles are distributed around the metal ion with a predefined geometry. These dummy particles are covalently bonded to the metal ion, while the rigid cation-dummy particle framework interacts with the coordinating residues through vdW and electrostatic interactions. However, only a few metal ions have been parametrized for this model¹⁷³⁻¹⁷⁷, and it can be biased by the predefined geometry. For example, it may underestimate the interactions between the metal ion and specific residues because of the inconsistency between the distribution of dummy particles and the coordination sphere¹⁷⁸.

The nonbonded model is another widely used model for metal ions, see Figure 5. In this model, the metal ion is represented by a soft sphere that usually has an integer charge and interacts with the environment through vdW and electrostatic interactions. The 12-6 Lennard-Jones (LJ) potential¹⁷⁹ is the most widely used potential to describe the vdW interactions, although the Born-Mayer potential may be used instead¹⁸⁰. Recently, Li, Merz, and co-workers parameterized various (mono-, di-, tri-, tetra-valent) ions for the 12-6 LJ nonbonded model in conjunction with several explicit water models by targeting the experimental hydration free energies (HFEs) or the ion-oxygen distances (IODs)^{56, 58, 181}. Meanwhile, they also found that when a metal ion has a charge of +2 or higher the 12-6 LJ nonbonded model is not able to reproduce its experimental HFE and IOD simultaneously. Moreover, Li and Merz proposed that this deficiency originated from the overlook of ion-induced dipole interaction in the 12-6 model. To solve this problem, they proposed adding a C_4 term to the conventional 12-6 model to take this interaction into account¹⁸². The new model was named as the 12-6-4 LJ nonbonded model and it can reproduce both the experimental HFE and IOD simultaneously for various metal ions¹⁸². Furthermore, the 12-6-4 model can also simulate ion-ligand interactions. For example, by optimizing the C_4 term between metal ion and ligands, Sengupta *et al.* showed that the 12-6-4 model can accurately simulate the chelate effect between a metal ion and organic ligands in the aqueous environment, capturing both the thermodynamic and structural features simultaneously^{56, 59}. In addition, the 12-6-4 model can also accurately model the interactions between nucleic acids and metal ions after parameter optimization¹⁸³. Finally, inspired by the 12-6-4 model, Liao *et al.* added the C_4 term into the cationic dummy atom model and parameterized the new model for several divalent and trivalent metal ions¹⁸⁴.

Figure 5. Left: the bonded model; middle: the cationic dummy atom model; right: the non-bonded model. The metal ion is colored in light blue, and the coordinating atoms are colored in red. The dummy atoms in the cationic dummy atom model are in white.



The 12-6-4 LJ nonbonded potential form is as follows:

$$U_{ij}(r_{ij}) = \frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} - \frac{C_4^{ij}}{r_{ij}^4} + \frac{e^2 Q_i Q_j}{r_{ij}} \quad (10)$$

where e represents the charge of the proton, Q_i and Q_j are partial charges of atoms i and j . The electrostatic interaction between atoms i and j is represented by the Coulomb pair potential, while the van der Waals interaction is represented by the classic Lennard-Jones (12-6) potential plus an extra r^{-4} term. The C_4 terms between ions and water were parameterized in previous studies by Li et.al⁵⁵⁻⁵⁷. The C_4 terms between ions and other ligands can be optimized based on the following equation:

$$C_4(\text{atom type}) = \frac{C_4(\text{H}_2\text{O})}{\alpha_0(\text{H}_2\text{O})} \times \alpha_0(\text{atom type}) \quad (11)$$

where α_0 is an atom type dependent polarizability.

CHAPTER 2: UMBRELLA SAMPLING STUDIES ON HOST-GUEST SYSTEMS

This chapter is drawn from the peer-reviewed publication with the title of “Detailed potential of mean force studies on host–guest systems from the SAMPL6 challenge” in the *Journal of computer-aided molecular design* authored by Lin Frank Song, Nupur Bansal, Zheng Zheng, and Kenneth M. Merz. Nupur Bansal performed docking of the benchmark systems.

2.1 Methods

Overview. First, benchmark systems with known experimental binding free energies were constructed. Then PMF studies were performed to calculate the host-guest binding free energies. For reference, we identify the values obtained in this step as obtained from ‘PMF’ simulations. Afterwards we used linear regression to fit the ‘PMF’ values to experiment to obtain the correlation equation. Finally, after performing PMF studies on the test systems, which formed the SAMPL6 blind challenge, we were able to predict the binding free energies for the test systems by solving the equation obtained in the previous step. For reference, we identify these values as the ‘scaled values’. We initially submitted the scaled values to the SAMPL6 challenge where we performed well. Our overall performance on both host-guest systems ranked as the 4th best, with a RMSE of 2.36 kcal/mol, which was only 0.62 kcal/mol greater than the top rank. Notably, for the OAs system, we performed the best over all the submissions, and the 0.95 kcal/mol RMSE is largely within experimental error. For the CB8 system, we performed the third best among all the MD based methods. The RMSE was 3.51 kcal/mol, which was mainly caused by two significant outliers, which will be further discussed below. Even so, it was <1 kcal/mol greater than the top submission, which has a RMSE of 1.92 kcal/mol.

PMF Studies. For the test systems, all the initial structures were taken from the SAMPL6 distribution, with the ligand already docked into the pocket. For the benchmark systems, the

ligands were docked into the receptors using Glide¹⁸⁵⁻¹⁸⁷ from the Schrodinger suite 2015-4 with the standard precision (SP) methodology. For both the test and benchmark systems, the host charges were AM1-BCC charges¹⁸⁸⁻¹⁸⁹ as supplied by the SAMPL6 organizers who used OpenEye's QUACPAC toolkit, and we calculated the AM1-BCC charges for the ligands using AmberTools16.¹⁹⁰ GAFF1.8¹⁵² was used to describe the atom types and generate the parameters for all the receptors and ligands. TIP3P⁹⁵ water was used with Na⁺ and Cl⁻ ions added to neutralize the system and to maintain the corresponding experimental ion strength. The water box dimensions were around 40 Å * 40 Å * 40 Å. Three and six pairs of Na⁺ and Cl⁻ ions were added to maintain the experimental ionic strength of 10 mM and 25 mM using a Na₃PO₄ buffer, respectively. All simulations were performed using AMBER16-CUDA. SHAKE⁷⁴ was used to constrain bonds involving hydrogen atoms. The particle mesh Ewald (PME) method¹⁹¹⁻¹⁹² was used to treat the long-range electrostatic interactions and the non-bonded cutoff was set to 12 Å.

OAs. A dummy particle that has no charge and van-der Waals properties was bound in the cavity with six carbon atoms at the bottom of the cavity through harmonic distance restraints of 64 kcal/(mol*Å²). The restraints were chosen so that it does not affect the dynamics of the OAs. The distance between the dummy atom and one atom of the ligand was chosen as the reaction coordinate of the PMF and one-dimensional US simulations were performed to generate the PMF. Prior to the US simulations, starting with the initial topology, a total of 100000 minimization steps were performed on the simulation box with the ligand restrained at its initial position. Afterwards the box was heated up from 0 K to 300 K in 1 ns under NVT condition (constant number of atoms, constant volume and constant temperature), after which a 1 ns NPT (constant number of atoms, constant pressure and constant temperature) simulation at 300 K and 1 bar was performed to equilibrate the system. Starting with the equilibrated structure, US simulations were performed.

The US windows were spaced every 0.2 Å, with 2ns NPT simulation for each window. The last snapshot of the previous window was used as the starting structure for the next window. Starting from its initial position, two separate series of US simulations were performed: the ligand was pushed to aqueous solution until the reaction coordinate reached 20Å, as well as pushed into the binding pocket until the reaction coordinate reached 0Å. The dummy particle-ligand restraint was set to 16 kcal/ (mol·Å²) for all US simulations. The step size of the US simulations was 2fs, and the reaction coordinate distance was recorded every 100 steps, resulting in 10000 data points recorded in each window. The first 200ps simulation of each window was discarded for PMF generation. WHAM was used to generate the PMFs. For each ligand, 4 repeated PMF simulations were performed with different initial velocities and an average binding free energy and root mean square deviation were calculated. PMF studies were performed on both the test and the benchmark systems. The speed of Amber16-CUDA is around 100-120 ns/day, hence, it took less than 2 days to finish each PMF run on a K80.

CB8. The procedure was similar to that of the OAs except for some minor differences. First, instead of using the distance between an anchored dummy atom and the ligand as the reaction coordinate, the distance between the center of mass of equatorial carbon atoms and the ligand was used. Second, 15 Å was set as the upper limit of the reaction coordinate distance for most cases, since the ligands were already pushed in solution at this point. Moreover, since the ligand could dissociate from both directions of CB8, we pulled the ligand out from both directions and the lower binding free energy value was reported as the PMF obtained value. Finally, the US windows were spaced every 0.1 Å for some distance intervals to ensure enough sampling, and 4ns NPT simulations were performed for each window with the first 1ns simulations discarded for PMF generation. Again, the above PMF studies were performed on both the test systems and the

benchmark systems. The speed of Amber16-CUDA is around 100-120 ns/day, hence, it took around 4 days to finish each PMF run on a K80.

2.2 Systems: Benchmark and Test Systems

Test systems. All the structures of the hosts and the guests are shown in Figure 6. OA is a basket-like molecular container that is water soluble due to the eight carboxylate groups, four at each end of the molecule. The two ends have different widths, one being around 10Å and the other being around 3Å. The binding pocket is around 10Å in depth, and it is hydrophobic, defined by eight aromatic faces. TEMOA is almost the same as OA except it has four additional methyl groups around the rim of the binding pocket, which may affect the ligand binding affinity by constricting the binding pocket entrance. In this test system, there are 8 common guest ligands for OA and TEMOA, and they are all carboxylic acids with relatively small sizes. CB8 is a member of the Curcubit[n]uril family, which consists of a series of macrocyclic molecules that are comprised of glycoluril (=C₄H₂N₄O₂=) monomers linked by methylene bridges (-CH₂-). As the names indicates, CB8 has eight glycoluril units. Though the oxygen atoms are located along the edges, CB8 has a uniform negative electrostatic potential within the cavity and the portal area, which benefits cationic guests binding to the pocket.¹⁹³ In this test system, we have 11 guests in total. Unlike the OA guests, the size and flexibility of the CB8 guests vary. Notably, all of them have one cationic group, namely -NR₄⁺ (R can be H or alkyl groups), except ligand 4 (see Figure 6), which has three cationic groups.

Benchmark systems. For OAs, the benchmark system were taken from the SAMPL5 main challenge. In total, there are six ligands in the system for both OA and TEMOA. The experimental binding free energies range from -2.38 kcal/mol to -9.38 kcal/mol. For CB8, since there was insufficient data in the previous SAMPL challenges, we turned to other experimental studies.¹⁹³⁻

¹⁹⁴ In total there are 11 ligands in the CB8 benchmark system, and the binding free energies range from 7.96 -kcal/mol to -15.97 kcal/mol. The benchmark systems were constructed in a way that ligands with the same functional groups as the test systems were included while insured that we had a range of binding free energies. All the ligand structures for the benchmark set are shown in Figure 7.

Figure 6. Structures of host OA, TEMOA and their guest molecules. a) Side view and top view of each host. Carbon, nitrogen, oxygen, hydrogen atoms are colored cyan, blue, red, white, respectively; b) the 8 common ligands for OA and TEMOA; c) the 11 ligands for CB8. Protonation states are indicated in the figure.

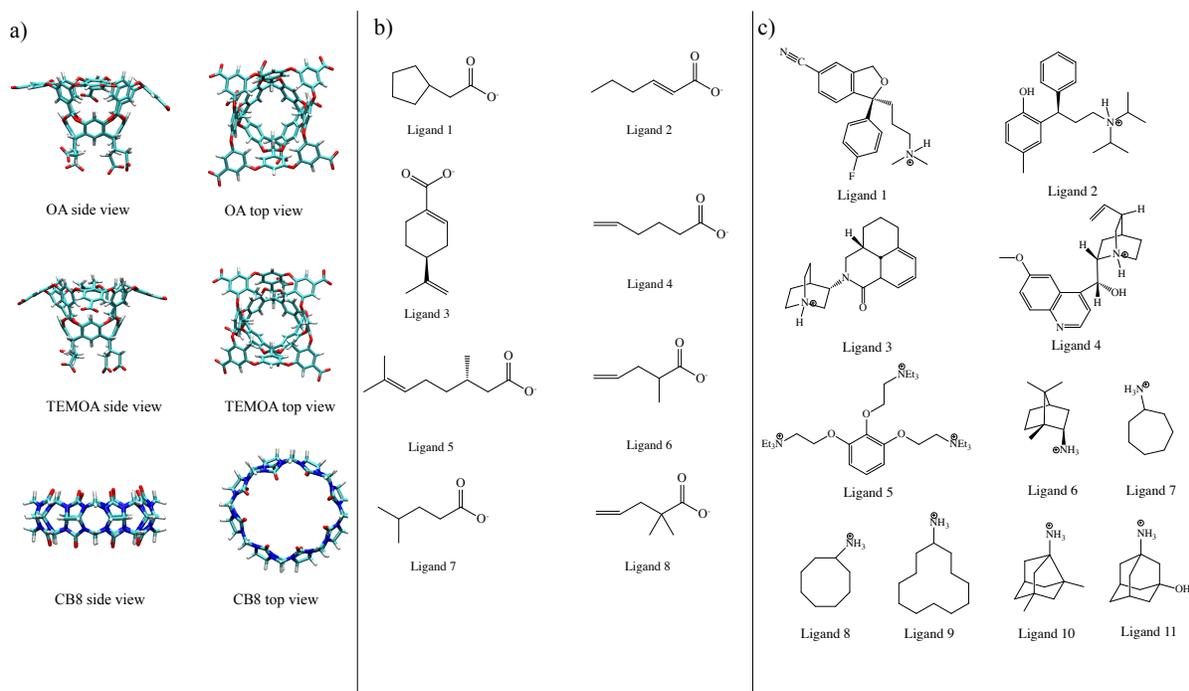
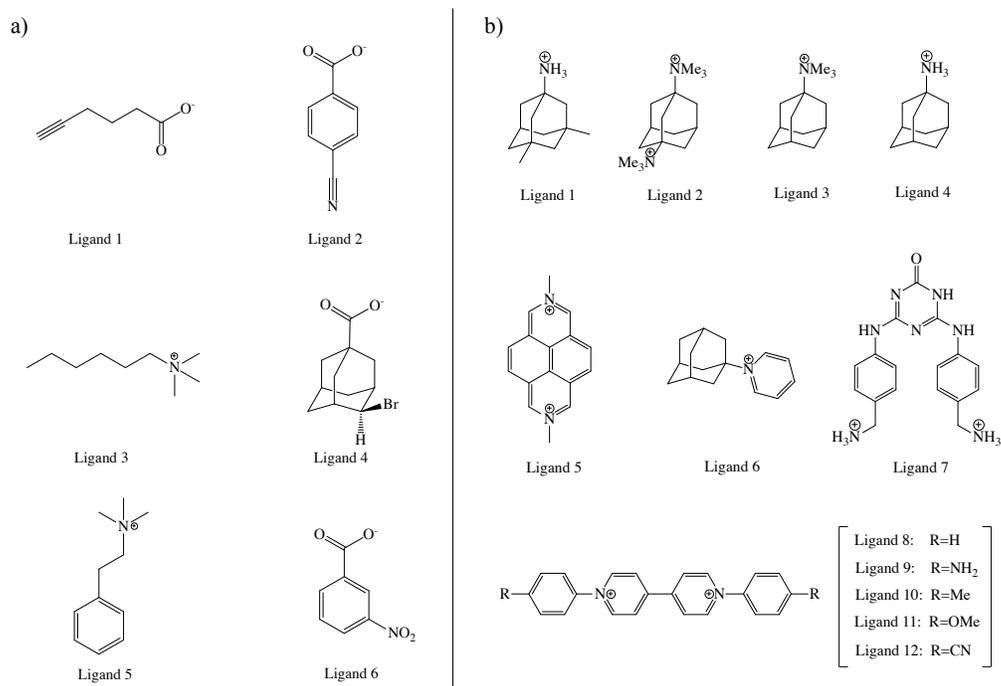


Figure 7. The structures of the ligands in the benchmark systems. a) Left panel is the 6 common ligands for OA and TEMOA; b) Right panel is the 12 ligands for CB8. Protonation states are indicated in the graph.



2.3 Results and Discussion

2.3.1 Benchmark Systems

OAs. As mentioned in the methods section, for each guest, four PMF simulations were performed. The averaged binding free energies were calculated from the PMFs and summarized in Table 2, along with the corresponding experimental values. One interesting thing to note is that all the ligands have similar binding affinities to both OA and TEMOA except for ligand 4, which is 7 kcal/mol more favorable for OA binding. As seen from the table, our PMF simulations are overestimating the binding free energies. The average ‘PMF obtained values’ are around 2 to 3 kcal/mol lower (absolute values) than the experimental values. The overall mean unsigned error (MUE) is 2.85 kcal/mol and the RMSE is 2.96 kcal/mol. This indicated that there is some systematic error in our calculations, which may be associated with the sampling or with the force field, or a combination of both. Interestingly, the correlation between the PMF derived binding

free energies and experiment is very good (see Figure 8). Using a linear regression, we obtain an R^2 of 0.846 and the fitted line was used to scale subsequent PMF derived free energies of binding (see Figure 8). It has the form:

$$\text{Experiment} = 0.7265 * \text{PMF} + 0.6944 \quad (12)$$

Overall, the PMF simulations were well converged: as you can see from Table 1, the RMSEs for the four independent runs for each ligand are within 1.4 kcal/mol, with most being within 0.8 kcal/mol. Our results for ligand 4 also agree well with the values calculated using PMF simulations in a previous paper from our group⁴²: -12.87 kcal/mol vs -14 kcal/mol for OA and -4.84 kcal/mol vs -4.45 kcal/mol for TEMOA.

Table 1. Root mean square deviation (RMSE) of the 4 repeated PMF simulations of each ligand based on the average value for both benchmark system and test system of OAs and CB8.

RMSE (kcal/mol)			
Benchmark system	OA	TEMOA	CB8
Ligand 1	0.49	0.64	0.76
Ligand 2	0.08	1.03	0.24
Ligand 3	0.21	0.57	0.64
Ligand 4	0.74	1.36	0.49
Ligand 5	0.07	0.19	0.45
Ligand 6	0.40	1.04	0.29
Ligand 7			1.43
Ligand 8			1.07
Ligand 9			0.53
Ligand 10			0.68
Ligand 11			1.38
Ligand 12			0.97
Test system	OA	TEMOA	CB8
Ligand 1	0.30	0.21	1.82

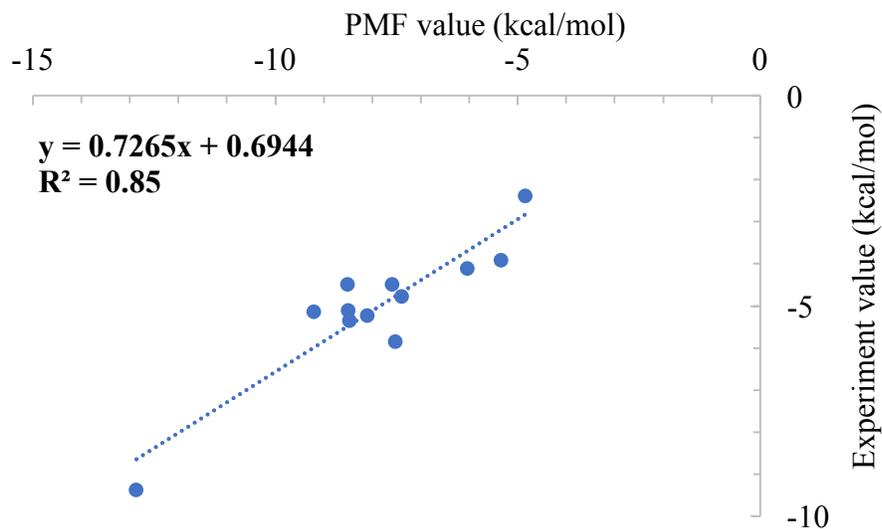
Table 1 (cont'd)

Ligand 2	0.07	0.46	0.85
Ligand 3	0.43	0.39	2.74
Ligand 4	0.18	0.56	0.31
Ligand 5	1.08	0.69	2.35
Ligand 6	0.10	0.47	1.04
Ligand 7	0.16	0.17	0.37
Ligand 8	0.23	0.17	0.23
Ligand 9			1.30
Ligand 10			0.50
Ligand 11			0.20

Table 2. Calculated average binding free energy from PMF simulations versus experiment for the benchmark OAs. In total, there are 6 ligands for both OA and TEMOA (see Figure 7).

	OA			TEMOA		
	PMF (kcal/mol)	Scaled (kcal/mol)	Experiment (kcal/mol)	PMF (kcal/mol)	Scaled (kcal/mol)	Experiment (kcal/mol)
Ligand 1	-8.10	-5.19	-5.23	-8.47	-5.46	-5.36
Ligand 2	-7.59	-4.82	-4.49	-9.21	-6.00	-5.15
Ligand 3	-7.40	-4.68	-4.78	-7.53	-4.78	-5.85
Ligand 4	-12.87	-8.66	-9.38	-4.84	-2.82	-2.38
Ligand 5	-6.04	-3.69	-4.12	-5.35	-3.19	-3.91
Ligand 6	-8.50	-5.48	-5.12	-8.51	-5.49	-4.49
	PMF (kcal/mol)			Scaled (kcal/mol)		
Overall MUE	2.85			0.51		
Overall RMSE	2.96			0.62		

Figure 8. Correlation between binding free energies obtained from PMF simulations and experiment.



CB8. The strategies for our CB8 simulations were similar to those used for the OAs except for some minor differences highlighted in the methods section. The binding free energies obtained from our PMF simulations relative to experiment are shown in Table 3. Similar to the OA benchmark system, the PMF simulations overestimated the binding free energies. The overestimate for the CB8 ligands varied more broadly, ranging from 3.1 kcal/mol to 8.88 kcal/mol. The overall mean error is 5.96 kcal/mol and the RMSE is 6.23 kcal/mol. A linear regression analysis was used to fit the calculated PMF values to experiment values and as for the OAs we found the correlation between them to be very good with an R^2 of 0.83. In order to prospectively scale binding affinities the following equation was used:

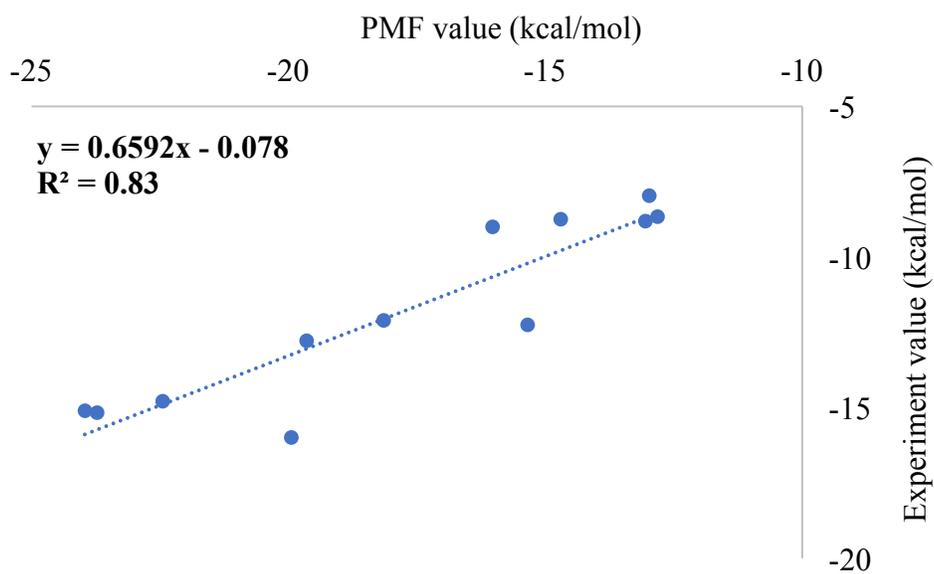
$$Experiment = 0.6592 * PMF - 0.078 \quad (13)$$

Overall, the PMF simulations were well converged: as you can see from Table 1, the RMSEs for the four independent runs for each ligand are within 1.43 kcal/mol, with most being within 0.8 kcal/mol.

Table 3. Calculated average binding free energies from PMF simulations along with the experimental values for the benchmark CB8 systems (see Figure 7).

	CB8		
	PMF (kcal/mol)	Scaled (kcal/mol)	Experiment (kcal/mol)
ligand 1	-19.94	-13.22	-15.97
ligand 2	-23.72	-15.71	-15.16
ligand 3	-23.96	-15.87	-15.08
ligand 4	-15.34	-10.19	-12.24
ligand 5	-18.15	-12.04	-12.09
ligand 6	-19.65	-13.03	-12.77
ligand 7	-22.45	-14.88	-14.77
ligand 8	-13.05	-8.68	-8.80
ligand 9	-12.98	-8.63	-7.96
ligand 10	-12.81	-8.52	-8.65
ligand 11	-16.02	-10.64	-8.99
ligand 12	-14.70	-9.77	-8.72
	PMF (kcal/mol)	Scaled (kcal/mol)	
Overall MUE	5.96	0.85	
Overall RMSE	6.23	1.19	

Figure 9. Correlation between binding free energies calculated from PMF simulations and experiment.



Using the correlation equation, we scaled the PMF values of the benchmark systems and the scaled results are summarized in Tables 2 and 3. We found that after scaling, the MUE and RMSE are substantially reduced. For the OAs system, the MUE and RMSE was 2.85 kcal/mol and 2.96 kcal/mol for the PMF values, respectively. After scaling, they were reduced to 0.51 kcal/mol and 0.62 kcal/mol. And for the CB8 system, the MUE and RMSE goes from 5.96 kcal/mol to 0.85 kcal/mol and 6.23 kcal/mol to 1.19 kcal/mol, respectively. This improvement strongly supports the idea scaling to remove systematic errors^{111, 195}, in order to be able to reliably predict binding free energies.

With the benchmark results in hand we participated in the SAMPL6 challenge for the systems shown in Figure 6. In particular, we carried out the blinded PMF simulations and reported the raw data and the results derived by scaling the PMF predicted blind results using the equations derived above. The outcome of this process is summarized below.

2.3.2 Test Systems

OAs. The binding free energies obtained using PMF simulations on the OAs system are summarized in Table 4. There were 8 ligands for both OA and TEMOA. Most of the errors for the calculated binding free energies are greater than 2 kcal/mol. The MUE is 2.83 kcal/mol and 2.17 kcal/mol, and the RMSE is 3.01 kcal/mol and 2.76 kcal/mol for OA and TEMOA, respectively. The overall RMSE for the OAs system is 2.89 kcal/mol. These results suggest that the standard PMF studies on this system suffers from systematic errors that tend to give too favorable estimates for the binding free energies.

From the benchmark systems, we observed that by simply scaling the computed free energies we were able to drastically improve the quality of the results relative to experiment. We hypothesized that this is due to a large systematic error in the force field we built to study these systems. Using

equation 12 we scaled the raw free energies directly obtained from the PMF simulations (see Table 5). When compare to experiment, the scaled values realize errors less than 1 kcal/mol for most cases and less than 2 kcal/mol for all the cases. The MUE is reduced to 0.51 kcal/mol and 1.03 kcal/mol, and the RMSE is reduced to 0.60 kcal/mol and 1.20 kcal/mol for OA and TEMOA, respectively. The overall RMSE for the OA systems is now 0.95 kcal/mol, which is a dramatic improvement over the 2.89 kcal/mol obtained from the raw results. The scaled results for these systems were the best in the SAMPL6 competition. These results affirm that scaling raw free energies obtained using PMF simulations can yield remarkable agreement with experiment especially for the congeneric OA systems where the ligand characteristics are constant (aliphatic tail with charge head group). Overall, the PMF simulations were well converged: as you can see from Table 1, the RMSEs for the four independent runs for each ligand are within 1.1 kcal/mol, with most being within 0.7 kcal/mol.

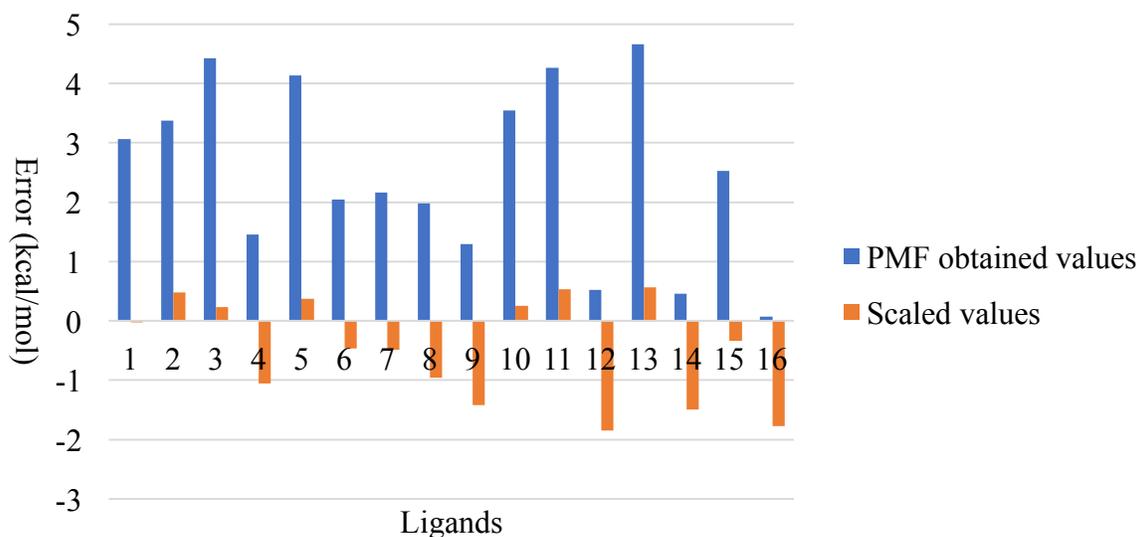
Table 4. Summary of the binding free energies for the test OA systems (see Figure 6).

Units: kcal/mol	OA			TEMOA		
	PMF	Experiment	Error	PMF	Experiment	Error
Ligand 1	-8.74	-5.68	3.06	-7.35	-6.06	1.29
Ligand 2	-8.02	-4.65	3.37	-9.52	-5.97	3.55
Ligand 3	-12.81	-8.38	4.43	-11.07	-6.81	4.26
Ligand 4	-6.63	-5.18	1.45	-6.12	-5.6	0.52
Ligand 5	-11.25	-7.11	4.14	-12.45	-7.79	4.66
Ligand 6	-6.63	-4.59	2.04	-4.62	-4.16	0.46
Ligand 7	-7.13	-4.97	2.16	-7.93	-5.4	2.53
Ligand 8	-8.2	-6.22	1.98	-4.2	-4.13	0.07
MUE	2.83			2.17		
RMSE	3.01			2.76		
Overall RMSE	2.89					

Table 5. Summary of the scaled (equation 12) binding free energies for the test OA systems (see Figure 6)

Units: kcal/mol	OA			TEMOA		
	Scaled	Experiment	Error	Scaled	Experiment	Error
Ligand 1	-5.66	-5.68	-0.02	-4.65	-6.06	-1.41
Ligand 2	-5.13	-4.65	0.48	-6.22	-5.97	0.25
Ligand 3	-8.61	-8.38	0.23	-7.35	-6.81	0.54
Ligand 4	-4.12	-5.18	-1.06	-3.75	-5.6	-1.85
Ligand 5	-7.48	-7.11	0.37	-8.35	-7.79	0.56
Ligand 6	-4.12	-4.59	-0.47	-2.66	-4.16	-1.50
Ligand 7	-4.49	-4.97	-0.48	-5.07	-5.4	-0.33
Ligand 8	-5.26	-6.22	-0.96	-2.36	-4.13	-1.77
MUE	0.51			1.03		
RMSE	0.60			1.20		
Overall RMSE	0.95					

Figure 10. Comparison of the errors of the PMF obtained values and scaled values for the test OAs. On the X axis: 1-8 represents ligands 1-8 binding to OA, while 8-16 represents ligands 1-8 binding to TEMOA.



CB8. The raw PMF binding free energies for the CB8 system are summarized in Table 6. There are 11 ligands in total and the structures can be found in Figure 6. The observed errors are large

for most cases, especially for ligand 3 and ligand 5. The MUE is 6.79 kcal/mol, and the RMSE is 8.04 kcal/mol. Using equation 13 we scaled the raw PMF free energy values and the results are summarized in Table 7. In comparison to experimental values, the errors are greatly reduced: the MUE goes from 6.79 kcal/mol to 2.44 kcal/mol. As you can see from Figure 11, for most of cases the error is ~ 2 kcal/mol. However, there're two major outliers that have error of 5.42 kcal/mol (ligand 3) and 8.87 kcal/mol (ligand 5), respectively. These large outliers combined with the remaining 9 systems yields a RMSE of 3.51 kcal/mol, which represents a large improvement, but shows that there are other issues at play for this set of ligands. Given the magnitude of these outliers we decided to examine them in more detail and these results are summarized in the next section. The scaled results ranked 7 out of 36 submissions. All the submissions for this system performed relatively poorly. The lowest RMSE reported was 1.92 kcal/mol and only 4 submissions had RMSE lower than 3 kcal/mol. For MD based methods, our scaled results ranked the third best. Overall, the PMF simulations were well converged: as you can see from Table 1, the RMSEs for the four independent runs for each ligand are within 1.3 kcal/mol, except for ligand 1, ligand 3 and ligand 5, which also have larger errors for our computed results vs experiment results.

Table 6. Summary of the binding free energies for the test CB8 system obtained from PMF simulations (see Figure 6).

	CB8		
	PMF (kcal/mol)	Experiment (kcal/mol)	Error (kcal/mol)
ligand 1	-15.74	-6.69	9.05
ligand 2	-12.25	-7.65	4.6
ligand 3	-19.72	-7.66	12.06
ligand 4	-11.67	-6.45	5.22
ligand 5	-25.17	-7.8	17.37
ligand 6	-14.27	-8.18	6.09

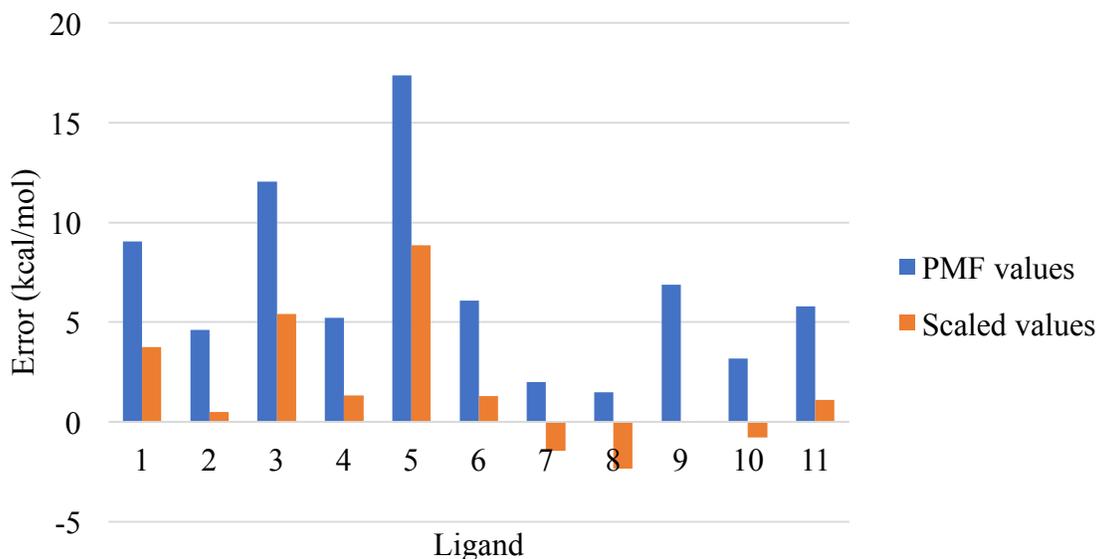
Table 6 (cont'd)

ligand 7	-10.34	-8.34	2
ligand 8	-11.49	-10	1.49
ligand 9	-20.37	-13.5	6.87
ligand 10	-11.87	-8.68	3.19
ligand 11	-14.02	-8.22	5.8
MUE (kcal/mol)	6.79		
RMSE (kcal/mol)	8.04		

Table 7. Summary of the scaled (equation 13) binding free energies for the test CB8 systems (see Figure 6)

	CB8		
	Scaled (kcal/mol)	Experiment (kcal/mol)	Error (kcal/mol)
ligand 1	-10.45	-6.69	3.76
ligand 2	-8.15	-7.65	0.5
ligand 3	-13.08	-7.66	5.42
ligand 4	-7.77	-6.45	1.32
ligand 5	-16.67	-7.8	8.87
ligand 6	-9.48	-8.18	1.3
ligand 7	-6.89	-8.34	-1.45
ligand 8	-7.65	-10	-2.35
ligand 9	-13.51	-13.5	0.01
ligand 10	-7.9	-8.68	-0.78
ligand 11	-9.32	-8.22	1.1
MUE (kcal/mol)	2.44		
RMSE (kcal/mol)	3.51		

Figure 11. Comparison of the errors of the PMF values and scaled values for the test CB8. X axis: 1-11 represents ligands 1-11 (see Figure 6).



2.3.3 Lesson Learned

2.3.1.1 The Two Faces of CB8

CB8 is a host that has two faces by which a ligand can enter. After submitting our initial results to the SAMPL6 challenge, we realized that we should explore both entrances/exits. Therefore we repeated the PMF studies pulling the ligand out of the pocket from the other direction for both the benchmark and test systems, except for symmetrical ligands. Comparing those two directions, the less negative binding free energy is identified as the final binding free energy. Although the overall results do not change much, we summarize the final values here.

Benchmark system. The correlation between the PMF values and experiment values is plotted in Figure 12, and the equation we use for scaling becomes:

$$\text{Experiment} = 0.5189 * \text{PMF} - 3.1898 \quad (14)$$

The PMF values and the scaled values using equation 14 are summarized in Table 8. Again, the scaling procedure reduces the MUE from 5.22 kcal/mol to 1.68 kcal/mol, and the RMSE from 5.53 kcal/mol to 2.13 kcal/mol.

Figure 12. Correlation between binding free energies calculated from PMF simulations and experiment.

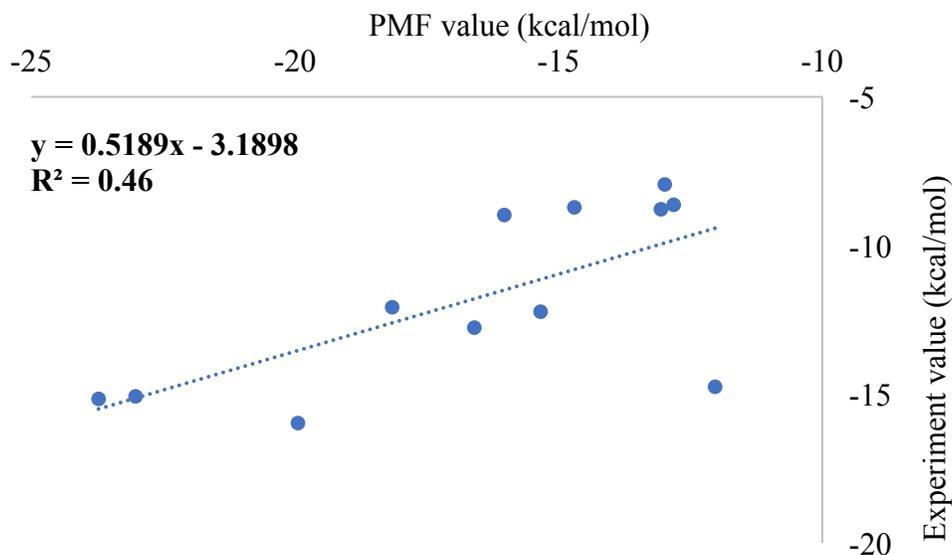


Table 8. Summary of the final binding free energies for the benchmark CB8 system.

	Forward* PMF	Reverse* PMF	Final PMF	Final Scaled	Experiment	Final PMF Error	Final Scaled Error
ligand 1	-19.94	-20.58	-19.94	-13.67	-15.97	3.97	-2.31
ligand 2	-23.72	-23.79	-23.72	-15.63	-15.16	8.56	0.47
ligand 3	-23.96	-23.02	-23.02	-15.26	-15.08	7.94	0.18
ligand 4	-15.34	-15.89	-15.34	-11.28	-12.24	3.10	-0.96
ligand 5	-18.15	-18.15	-18.15	-12.74	-12.09	6.06	0.65
ligand 6	-19.65	-16.59	-16.95	-11.93	-12.77	3.82	-0.84
ligand 7	-22.45	-12.03	-12.03	-9.56	-14.77	-2.74	-5.21

Table 8 (cont'd)

ligand 8	-13.05	-13.05	-13.05	-10.09	-8.80	4.26	1.30
ligand 9	-12.98	-12.98	-12.98	-10.05	-7.96	5.02	2.10
ligand 10	-12.81	-12.81	-12.81	-9.97	-8.65	4.16	1.31
ligand 11	-16.02	-16.02	-16.02	-11.63	-8.99	7.03	2.64
ligand 12	-14.70	-14.70	-14.70	-10.95	-8.72	5.98	2.22

Note: ‘Forward’ represents our initial pulling direction in section 3.2, same as our SAMPL6 submission; ‘Reverse’ represents the other pulling direction. The ones with light blue shading are identified as the final binding free energy values. Units are in kcal/mol.

Test system. The PMF values obtained from both directions and the final PMF value and scaled values obtained from equation 14 are summarized in Table 9. The less negative binding free energy obtained is then identified as the final binding free energy and is colored in blue in Table 9. Similar to the previous results, the scaling procedure reduces the errors significantly (see Figure 13). The MUE and RMSE are reduced from 5.65 kcal/mol to 2.35 kcal/mol and 6.88 kcal/mol to 3.19 kcal/mol. Most of the errors are ~2 kcal/mol after scaling, with one major outlier (ligand 5). We will analyze the issues with ligand 5 in further detail in the next section.

Table 9. Summary of the final binding free energies for the test CB8 systems (see Figure 6).

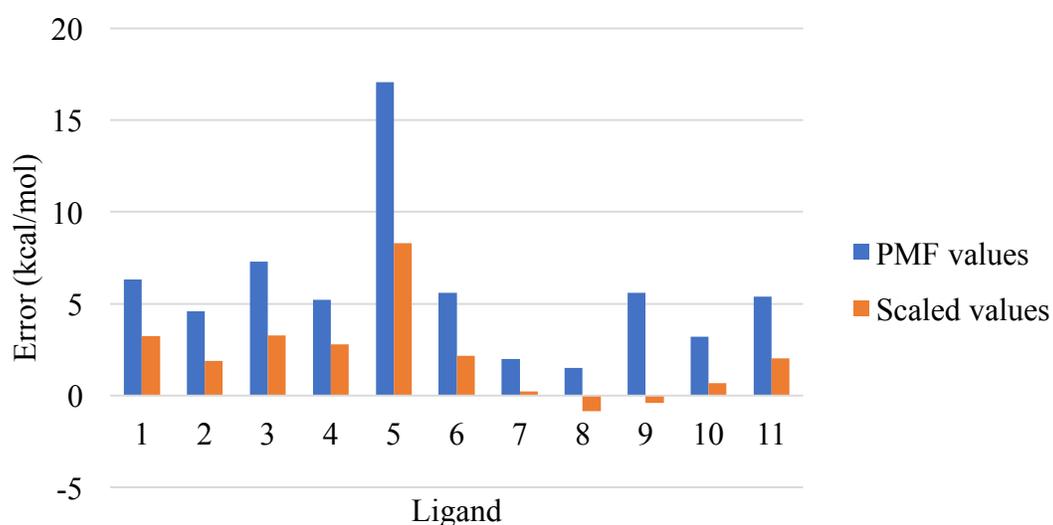
	Forward* PMF	Reverse* PMF	Final PMF	Final Scaled	Experiment	Final PMF Error	Final Scaled Error
ligand 1	-15.74	-13.00	-13.00	-9.94	-6.69	6.31	3.25
ligand 2	-12.25	-14.03	-12.25	-9.55	-7.65	4.60	1.90
ligand 3	-19.72	-14.95	-14.95	-10.95	-7.66	7.29	3.29
ligand 4	-11.67	-27.24	-11.67	-9.25	-6.45	5.22	2.80
ligand 5	-25.17	-24.86	-24.86	-16.09	-7.80	17.06	8.29
ligand 6	-14.27	-13.77	-13.77	-10.34	-8.18	5.59	2.16
ligand 7	-10.34	-10.59	-10.34	-8.56	-8.34	2.00	0.22
ligand 8	-11.49	-11.91	-11.49	-9.15	-10.00	1.49	-0.85

Table 9 (cont'd)

ligand 9	-20.37	-19.08	-19.08	-13.09	-13.50	5.58	-0.41
ligand 10	-11.87	-12.20	-11.87	-9.35	-8.68	3.19	0.67
ligand 11	-14.02	-13.60	-13.60	-10.25	-8.22	5.38	2.03

Note: ‘Forward’ represents our initial pulling direction in section 3.2, same as our SAMPL6 submission; ‘Reverse’ represents the other pulling direction. Units are in kcal/mol.

Figure 13. Comparison of the errors of the PMF values and scaled values for the test set for CB8. X axis: 1-11 represents ligands 1-11 (see Figure 6).

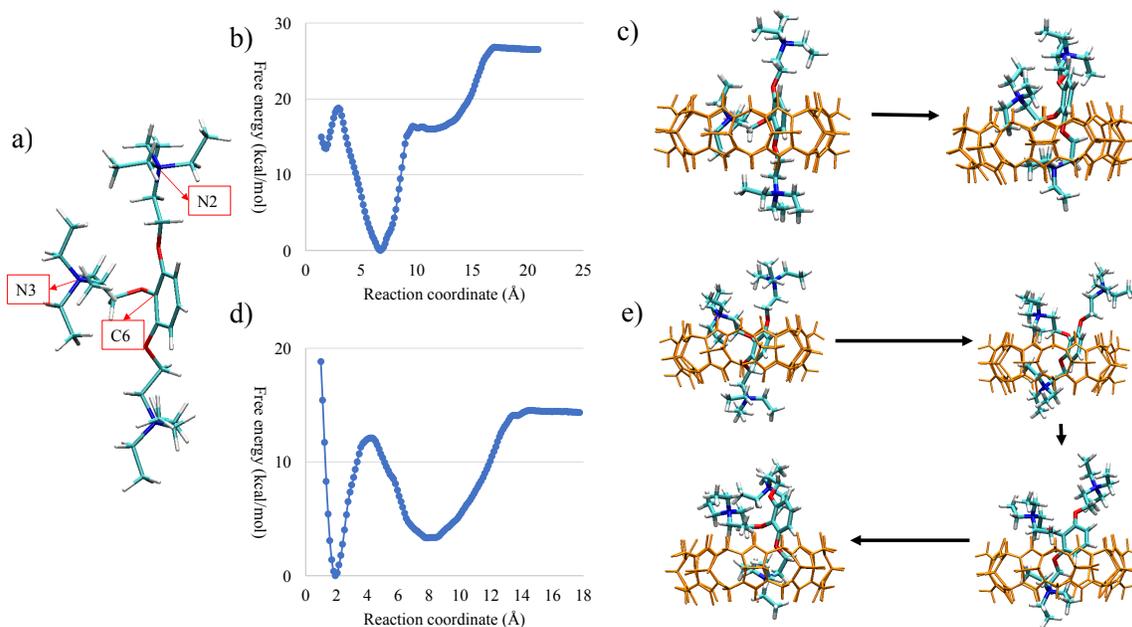


2.3.1.2 Lesson Learned From CB8-Ligand 5

To address the large errors found for ligand 5 binding to CB8, we repeated PMF simulations but chose a different atom of ligand 5 as the point from which we pulled the ligand out. Previously the reaction coordinate was the distance between the center of mass of CB8 and atom C6 (‘forward PMF’), or atom N2 (‘reverse PMF’), see Figure 14a, and the binding free energy obtained from these PMF’s was -25.17 kcal/mol and -24.86 kcal/mol, respectively. This time the reaction coordinate was the distance between the center of mass of CB8 and atom N3, and the binding free energy obtained from the PMF simulation was reduced to -13.83 kcal/mol. After scaling, the scaled binding free energy for ligand 5 is -10.37 kcal/mol, which has an error of only 2.57 kcal/mol

(versus the previous large error of 8.29 kcal/mol). The MUE and RMSE of the test system is reduced to 1.83 kcal/mol and 2.12 kcal/mol, respectively. After a close look at the trajectories, we found that pulling out using atom N3 resulted in a smoother transition of the bulky group out of the pocket, which reduced the sampling needed. Figure 14c shows the transition from the global minimum (~ 6.5 Å) to the local minimum (~ 11.5 Å) of Figure 14b, which is the free energy profile of the ‘reverse PMF’ of ligand 5. The aromatic ring and the middle chain moved out of the binding pocket simultaneously. However, when using the N3 atom, as you can see from Figure 14e, the flexible middle chain moved out first and then the aromatic ring, resulting in a much smoother transition and more sampling in this region. The sampling of the local minimum is also enhanced potentially due to the smoother transition, resulting in a much lower global minimum. Therefore, we conclude that choosing the reaction coordinate is very important. Multiple different trials using different reaction coordinates should be performed to get a better overall picture. Visualizing the overall reaction process and estimating the amount of sampling at the transition regions might help with the selection of the best reaction coordinate.

Figure 14. a) Structure of ligand 5 bound to CB8 system. Carbon, nitrogen, oxygen, hydrogen atoms are colored cyan, blue, red, white, respectively; b) free energy profile CB8-ligand 5 reverse PMF. the reaction coordinate is the distance between center of mass of CB8 and the N2 atom of ligand 5; c) the transition structures from the global minimum to the local minimum for b). CB8 is colored orange; d) free energy profile CB8-ligand 5 reverse PMF. the reaction coordinate is the distance between center of mass of CB8 and the N3 atom of ligand 5; e) the transition structures from the global minimum to the local minimum for d). CB8 is colored orange.



2.4 Conclusions

In the present work, we performed detailed PMF studies using US and WHAM on two host-guest systems, namely the OA and CB8 systems. We found that standard PMF studies on those systems using typical force field development protocols resulted in large overall RMSEs of 8.04 kcal/mol for CB8 and 2.89 kcal/mol for the OAs. Observing that for these systems our simulations tended to systematically overestimate the binding affinities, we developed a scaling procedure that improved our overall ability to accurately predict binding affinities to these systems (RMSE reduced to 3.51 kcal/mol for CB8 and 0.95 kcal/mol for the OAs). Moreover, some lessons were learned that further reduced simulation errors. Importantly, including considering the other possible entrance channels and the selection of multiple reaction coordinates should be carefully

considered.

As a final note, these PMF studies used the same force constant for all the umbrella sampling windows. To make sure that the distribution of each window centers around the targeted distance and neighboring window overlaps with each other, optimization of the restraint force constant for each window is required. This, along with the sampling and force field errors, could be the source of the large error for the PMF studies without scaling. All in all, through simple PMF studies on benchmark systems and test systems, binding free energies could be reliably predicted with an RMSE <2 kcal/mol.

CHAPTER 3: GPU TI STUDIES ON PROTEIN-LIGAND SYSTEMS

This chapter is drawn from the peer-reviewed publication with the title of “Using AMBER18 for Relative Free Energy Calculations” in the *Journal of chemical information and modeling* authored by Lin Frank Song, Taisung Lee, Chun Zhu, Darrin M. York, and Kenneth M. Merz. Taisung Lee and Darrin M. York implemented the GPU TI code into AMBER, and Chun Zhu helped to check the results.

3.1 Methods

3.1.1 System Preparation

All of the protein and ligand PDBs were obtained from the SI of the Wang *et al.* publication,⁸ see Appendix Figure 31 for the perturbation graph for each protein system. The atom names of the ligands were modified manually so that the common atoms of the ligands in each protein system have the same name and the unique atoms have different names. The protonation states of all the charged residues as well as Histidine residues were maintained as reported in Wang *et al.*. The AMBER FF14SB force field¹⁹⁶ was employed to describe the proteins and GAFF (version 1.8) was¹⁵² used for the ligands. Restrained electrostatic potential (RESP) charges for the ligands were calculated at the HF/6-31G(d) level of theory using the Gaussian 09 program¹⁹⁷ and AMBERTools16. The parmchk utility from AMBERTools16 was used to generate the missing parameters for the ligands. The systems were solvated using the SPC/E⁹⁶ water model using cubic simulation cells. 5 Å and 10 Å were used as the minimum distance between the edge of the cell and the solute atoms of the protein and ligand systems, respectively. The resulting solvated protein/ligand systems were then charge neutralized by adding Na⁺ or Cl⁻ ions¹⁹⁸. The particle mesh Ewald (PME) method¹⁹¹⁻¹⁹² was used to treat the long-range electrostatic interactions. All bonds involving hydrogen atoms were constrained using SHAKE⁷⁴. The AMBER16 package¹⁹⁹

was used to run the MD simulations. MD simulations for each protein-ligand system were performed to equilibrate the systems. Five steps of minimization were performed to remove close contacts. The first step minimizes the water molecules and counter ions, with the protein restrained. The second, third and fourth step restrains the heavy atoms, backbone heavy atoms, backbone carbon and oxygen atoms of the protein, while the last step minimizes the entire system. Each minimization step consisted of 20000 cycles of minimization using the steepest descent method. Afterwards the system was heated from 0 K to 300 K using the Langevin thermostat with a collision frequency of 2.5 ps^{-1} . The solute was restrained using a $5 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$ restraining potential. Finally, the system was equilibrated at 300 K for 5 ns employing the NPT ensemble using a Langevin thermostat with a collision frequency of 1 ps^{-1} . The Berendsen barostat was used for the pressure control with a pressure relaxation time of 10 ps. The time step was 2 fs and the nonbonded cutoff was 12 Å. The last snapshot was used to generate a pdb file. Using the generated pdb file, all the protein atoms were duplicated along with the common atoms of the second ligand. The unique atoms of the second ligand were added according to the mol2 file of the second ligand, the coordinates of which were obtained from the input files from Wang, *et al.* The “timerge” function of the parmed.py utility of the AMBER 14 package was used to generate the dual ligand topology.

3.1.2 TI Simulation: The One-step Protocol

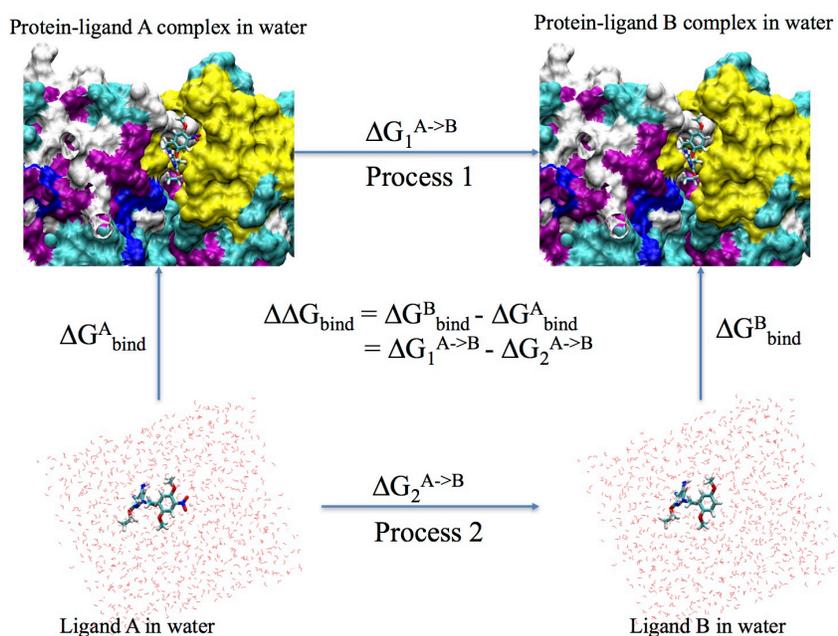
As shown in Figure 15, the relative binding free energy ($\Delta\Delta G$) can be calculated as the difference between the free energies (ΔG s) of changing one ligand to the other in the protein matrix and in solution. Therefore, TI simulations for both process 1 and 2 were performed. For each process, the one-step protocol was adopted, *i.e.* disappearing one ligand and appearing the other ligand simultaneously. The common atoms of the two ligands were linearly transformed and the unique

atoms were in the softcore region. Both the charge and vdW interactions between the disappearing (or appearing) unique atoms with the surrounding atoms were described by softcore potentials. Alternatively one can use the 3-step protocol, which consists of three steps: disappearing the charge interaction of one ligand, changing the vdW and bonded terms, and then appearing the charge of the second ligand. The one-step protocol not only takes less steps but also has the same charge for the initial and final state of the TI simulation. However, for the 3-step protocol, the charge of the system may change during the decharging/charging steps, which may affect the long-range electrostatic interactions via the use of a neutralizing background plasma in AMBER.

The detailed TI simulation protocol is as follows: First, using the dual ligand topology parameter file, 50000 steps of steepest descent minimization was performed. Then the system was heated from 0 to 300 K at the ps timescale, followed by 1 ns NVT equilibration at 300 K. Afterwards 1 ns pf NPT equilibration at 300 K and 1 bar was performed to equilibrate the density. These simulations were performed at $\lambda=0.5$ to equilibrate the system²⁰⁰⁻²⁰¹. No restraint was applied for these simulations and all structures were visually checked. For some perturbations, multiple runs had to be performed in order to obtain a stable starting structure. Afterwards the equilibrated structure was used as the starting structure for 12 λ windows (0.00922, 0.04794, 0.11505, 0.20634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, 0.99078). For each λ , 1 ns of NVT equilibration was performed with the initial velocities randomly generated to give a temperature of 300 K. Afterwards 5 ns of NVT simulation was performed to collect $\partial U/\partial \lambda$ data. A 12-point Gaussian quadrature was used for the numerical integration of $\partial U/\partial \lambda$ to obtain all necessary ΔG values. The non-bonded interaction cutoff was 9.0 Å and a softcore potential²⁰²⁻²⁰³ was used. The parameter α and β of softcore potential was 0.5 and 12 Å², respectively. The time step was 1 fs for all simulations and SHAKE was not used. All TI simulations used the Berendsen

thermostat with a coupling constant of 2 ps, except for the NPT equilibration step, which used the Langevin thermostat with a collision frequency of 2 ps^{-1} . We note that the Langevin thermostat is generally preferred over the Berendsen thermostat. The Berendsen barostat was used for NPT equilibration with a pressure relaxation time of 2 ps. NPT equilibration was performed using the CPU version of pmemd from the Amber14 package. The input files are available at GitHub: https://github.com/linfranksong/Input_TI

Figure 15. Thermodynamic cycle used for the calculation of the relative binding free energy between protein-ligand system A and protein-ligand system B.



3.2 Results and Discussion

3.2.1 Overall Results

The $\Delta\Delta G$ values directly obtained from the TI calculations can be found in the Appendix Table 20. The mean unsigned error (MUE) and root mean square deviation (RMSE) of these values compared to experiment are summarized in Table 10. After obtaining the $\Delta\Delta G$ values, we employed the cycle closure convergence strategy described previously²⁰⁴ and obtained our final $\Delta\Delta G$ values (summarized in the Appendix Table 20 as well). Thus, the following analysis is based

on the cycle-closure $\Delta\Delta G$ values. Table 11 summarizes the MUE and RMSE compared to experiment. The overall MUE obtained with GPU-TI of AMBER using the AMBER FF14SB/GAFF1.8 force field (AMBER for short) is 1.17 kcal/mol (0.27 kcal/mol larger than FEP+). Similarly, the RMSE is a bit higher for AMBER: 1.50 kcal/mol *versus* 1.14 kcal/mol for FEP+. Moreover, in our current work, we did not apply replica exchange, which could help enhance the overall sampling and improve the quality of the computed free energies. Future work will explore the role sampling (both in λ -space and phase space) plays on these systems versus the effect of force field errors.

Table 10. Summary of the MUE and RMSE of the eight systems based on $\Delta\Delta G$ values directly obtained from FEP or TI calculations.

System	# of ligands	# of perturbations	FEP+/OPLS 2.1 (kcal/mol)		AMBER GPU-TI/AMBER FF14SB + GAFF (1.8) (kcal/mol)		Difference* (kcal/mol)	
			MUE	RMSE	MUE	RMSE	MUE	RMSE
Thrombin	11	16	0.76	0.98	0.47	0.66	-0.29	-0.32
TYK2	16	24	0.74	0.95	1.07	1.29	0.33	0.34
JNK1	21	31	0.77	1.01	1.20	1.53	0.43	0.52
CDK2	16	25	0.95	1.14	0.95	1.14	0.00	0.00
PTP1B	23	49	0.93	1.27	1.08	1.49	0.15	0.22
BACE	36	58	0.87	1.05	1.33	1.79	0.46	0.74
MCL1	42	71	1.17	1.44	1.55	1.91	0.38	0.47
P38	34	56	0.86	1.06	1.41	1.82	0.55	0.76
Overall	199	330	0.92	1.17	1.25	1.64	0.33	0.47

* The difference is calculated as AMBER MUE or RMSE minus Schrödinger MUE or RMSE.

Table 11. Summary of the MUE and RMSE, R^2 and Kendall's tau coefficient (τ) of the eight systems based on cycle closure $\Delta\Delta G$ values.

System	# of ligands	# of perturbations	FEP+/OPLS 2.1 (kcal/mol)				AMBER GPU-TI/AMBER FF14SB + GAFF (1.8) (kcal/mol)				Difference* (kcal/mol)	
			MUE	RMSE	R^2	τ	MUE	RMSE	R^2	τ	MUE	RMSE
Thrombin	11	16	0.76	0.93	0.17	0.21	0.46	0.62	0.50	0.54	-0.30	-0.31
Tyk2	16	24	0.75	0.93	0.48	0.54	1.07	1.27	0.24	0.26	0.32	0.34
Jnk1	21	31	0.78	1.00	0.35	0.44	1.07	1.45	0.05	0.23	0.29	0.45
CDK2	16	25	0.91	1.11	0.15	0.30	0.97	1.13	0.35	0.46	0.06	0.02
PTP1B	23	49	0.89	1.22	0.43	0.55	1.06	1.40	0.35	0.51	0.17	0.18
BACE	36	58	0.84	1.03	0.37	0.36	1.20	1.47	0.27	0.31	0.36	0.44
MCL1	42	71	1.16	1.41	0.26	0.35	1.52	1.83	0.16	0.28	0.36	0.42
P38a	34	56	0.80	1.03	0.62	0.60	1.20	1.56	0.31	0.39	0.40	0.53
Overall	199	330	0.90	1.14	0.36	0.44	1.17	1.50	0.23	0.34	0.27	0.36

* The difference is calculated as AMBER MUE or RMSE minus Schrödinger MUE or RMSE.

With the cycle-closure $\Delta\Delta G$ values, we obtained the ΔG values following the procedure of Wang, *et al.*. In short, in this procedure all of the ligands' experimental values were used as a reference, and the sum of the predicted ΔG values was set to be equal to the sum of the experimental ΔG values. Though this way of calculating the offset can artificially improve the overall results, we adopted this procedure in order to better compare with Wang, *et al.* The predicted ΔG values were plotted against experimental ΔG values in Figure 16. We can see AMBER performs worse than FEP+. Out of the 199 ligands, 5 ligands (2.5%) for Schrödinger and 18 ligands (9.0%) for AMBER are more than 2kcal/mol off from experiment. The R^2 and Kendall's tau coefficient are listed in

Table 12. Figure 17 shows the individual plots of predicted *versus* experimental ΔG values for each of the 8 systems for both FEP+ and AMBER TI.

Figure 16. Correlation between predicted binding free energies and experimental data for the eight systems.

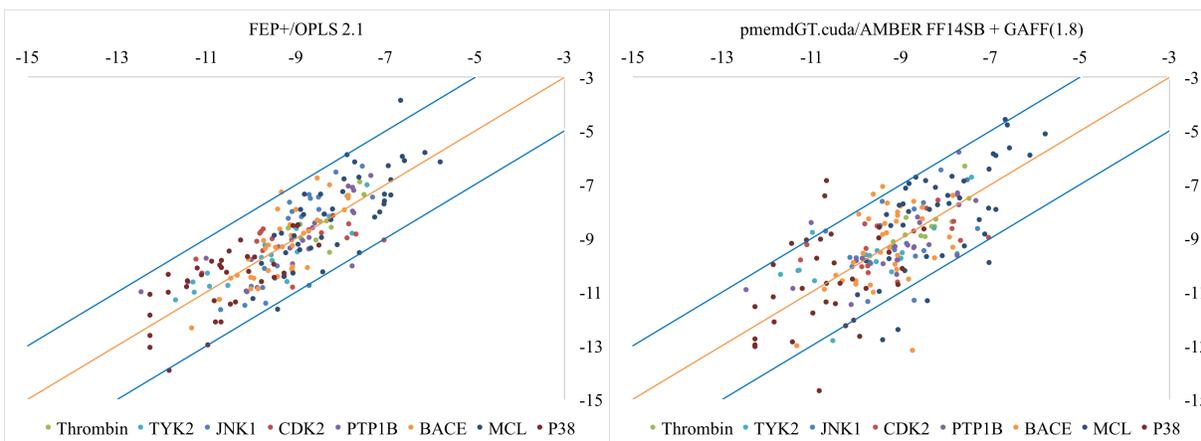
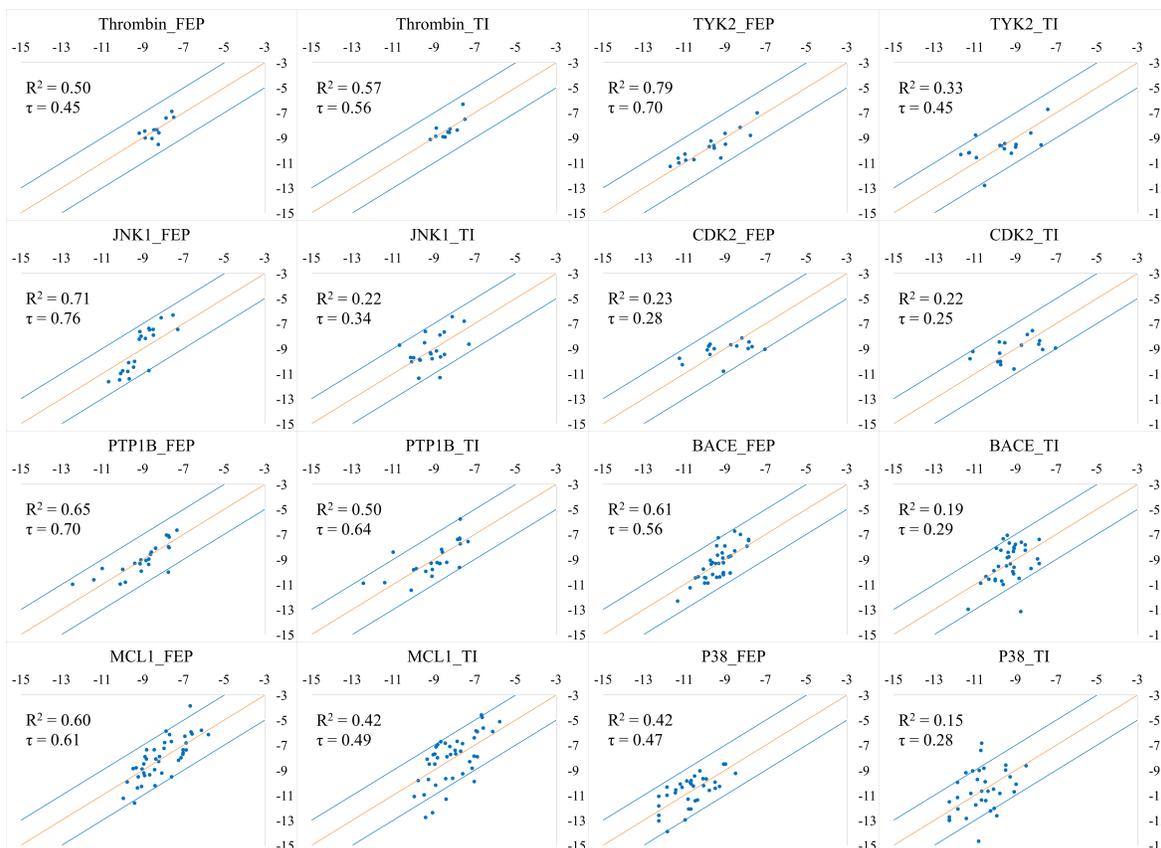


Table 12. R^2 and Kendall's tau coefficient for the correlation between predicted binding free energies and experimental data for the eight systems; τ represents the Kendall's tau coefficient.

System	# of ligands	# of perturbations	FEP+/OPLS 2.1 (kcal/mol)		AMBER GPU-TI/AMBER FF14SB + GAFF (1.8) (kcal/mol)	
			R^2	τ	R^2	τ
Thrombin	11	16	0.50	0.45	0.57	0.56
Tyk2	16	24	0.79	0.70	0.33	0.45
Jnk1	21	31	0.71	0.76	0.22	0.34
CDK2	16	25	0.23	0.28	0.22	0.25
PTP1B	23	49	0.65	0.70	0.50	0.64
BACE	36	58	0.61	0.56	0.19	0.29
MCL1	42	71	0.60	0.61	0.42	0.49
P38a	34	56	0.42	0.47	0.15	0.28
Overall	199	330	0.66	0.62	0.44	0.48

Figure 17. Correlation between the predicted binding free energies and experimental data for the eight systems studied herein. X axis: Experimental ΔG (kcal/mol); Y axis: Predicted ΔG (kcal/mol). τ is the Kendall's tau coefficient.



3.2.2 Uncertainty Estimate

To estimate the uncertainty in the calculations, we randomly selected 2 perturbations from each of the 8 systems and repeated the calculations described in section 3.1.2 twice. From Table 13, we can see most of the perturbations have standard deviations of less than 0.5 kcal/mol, except 4 of the perturbations. The overall standard deviation is 0.33 kcal/mol.

Table 13. Estimate of the uncertainty of the calculations.

System	Ligand 1	Ligand 2	Run_1 (kcal/mol)	Run_2 (kcal/mol)	Run_3 (kcal/mol)	Average (kcal/mol)	Standard Deviation (kcal/mol)
Thrombin	1d	1c	-0.20	-0.15	-0.15	-0.17	0.03
	6e	6b	0.60	0.75	0.75	0.70	0.09
TYK2	ejm_31	ejm_46	-0.75	-0.85	-0.65	-0.75	0.10
	jmc_28	jmc_30	-2.00	-2.00	-1.90	-1.97	0.06
JNK1	18626-1	18624-1	1.50	0.95	1.05	1.17	0.29
	18659-1	18634-1	-0.95	-1.10	-0.35	-0.80	0.40
CDK2	22	1h1r	-0.55	-0.90	-0.70	-0.72	0.18
	1oiy	1h1q	1.65	1.70	2.85	2.07	0.68
PTP1B	23466	23475	-1.50	-1.60	-2.65	-1.92	0.64
	20670(2qbs)	23330(2qbq)	-1.65	-1.40	-1.85	-1.63	0.23
BACE	CAT-13a	CAT-17g	1.95	1.10	1.65	1.57	0.43
	CAT-4p	CAT-13k	-1.45	-1.20	-0.85	-1.17	0.30
MCL1	26	57	-0.85	-1.05	-1.00	-0.97	0.10
	68	45	-0.75	-0.70	-0.85	-0.77	0.08
P38	p38a_2aa	p38a_2bb	-1.35	-0.20	0.65	-0.30	1.00
	p38a_3fly	p38a_3fmh	0.00	-0.35	0.85	0.17	0.62
Overall							0.33

3.2.3 The “Problematic Cases”

As alluded to in section 3.1.2, for some perturbations, multiple runs at $\lambda=0.5$ had to be run in order to obtain a stable starting structure; for example, the ligand significantly moved in the binding pocket or the conformation of the protein changed. In order to understand the origin of this problem better, we visually checked the initial structures and the structures after minimization, and found that there were a few cases that had close contacts in the initial structure, but after minimization, the structures had improved. No clashes between the ligand and the binding site of the protein were observed after minimization. We next hypothesized that our heating protocol was too fast, which

caused the observed structural issues. Hence, we repeated the “problematic cases” with a more rigorous minimization, heating and equilibration procedure. Five steps of minimization were performed to remove close contacts. The first step minimized the water molecules and the counter ions, with the protein restrained. The second, third and fourth step restrained the heavy atoms, backbone heavy atoms, backbone carbon and oxygen atoms of the protein, while the last step minimized the entire system. Each minimization step consisted of 20000 cycles of minimization using the steepest descent method. Afterwards the system was heated from 0 K to 300 K gradually over 1 ns with a coupling restraint of 5 kcal/(mol*Å²) on the solute, followed by equilibration at 300 K using the NPT ensemble for 200 ps with the same restraint. Then another 200 ps of NPT equilibration with a weaker restraint (2 kcal/(mol*Å²)) was performed. Finally the restraint was released and the system was equilibrated using NPT conditions for 600 ps. With these settings, the simulations successfully finished and the structures appeared fine after visual inspection. With the equilibrated structure, 12 λ windows were used for data collection with similar settings except: 1) the initial velocity was taken from the equilibrated structure as well as the coordinates; 2) the two end windows (0.00922 and 0.99078) used the velocity and coordinates from the equilibrated structure of the neighboring window (0.04794 and 0.95206). A few other differences between these new simulations and the former simulations include: 1) parmchk2 was used to generate the missing bond/angle/dihedral parameters for the ligands; 2) 22 and 12 Å was used as the minimum distance between the edge of the solvated cell and the ligand and protein/ligand systems respectively; 3) the protein/ligand system was thermalized more gradually and more steps of equilibration were used; 4) the Langevin thermostat was used with a collision frequency of 2 ps⁻¹ for all the TI simulations; 5) the CPU version of the AMBER 18 package was used instead of the AMBER 14 package for the TI simulations under NPT conditions. The overall MUE and RMSE

for these perturbations are about the same: 1.61 kcal/mol and 2.09 kcal/mol for the new protocol *versus* 1.52 kcal/mol and 1.93 kcal/mol for the former protocol. Even so, some of the individual changes were significant, but given the differing box sizes, thermalization protocols, thermostats, *etc.* this wasn't particularly surprising. Nonetheless, the average performance is relatively insensitive to the protocol employed. These data are summarized in the Appendix Table 21.

3.2.4 The Three-step Protocol

A recent publication highlighted differences between the one-step protocol and a 3-step protocol when employing AMBER TI calculations.²⁰⁵ In order to explore the impact of using one protocol over the other we performed 3-step calculations for one of the systems, *i.e.* the JNK1 system. As discussed above, the 3-step protocol consists of disappearing the charge, a vdW change and a charge reappearance step. For each step, the same minimization, heating and equilibration was performed at $\lambda=0.5$ as described in section 3.2.3. The equilibrated structure and velocities were used for the 12λ window TI calculation. The remaining settings for the TI calculations were the same as in section 3.1.2. We found that the MUE and RMSE is nearly the same as the one step protocol: 1.11 *versus* 1.07 kcal/mol, 1.43 *versus* 1.45 kcal/mol, respectively. This suggests that although there are differences between the two protocols that are worthy of in-depth exploration, the overall performance using either protocol is about the same, using the current code base and force fields. These data are summarized in the Appendix Table 22.

3.2.5 Discussion

In the Appendix Table 23 we summarize all of the 330 perturbations. Overall, we find that AMBER performs reasonably well for perturbations between halogens and H, CH₃ or CH₂CH₃: 44 of 49 perturbations have errors less than 2 kcal/mol, 34 of which have an error less than 1 kcal/mol. Perturbations involving large van der Waals radii changes, like Br to H or I to H, tend to have

larger errors. We further analyzed the perturbations based on the “size” of the perturbation; whether there is ring appearance/disappearance or whether there is a ring type change (for example, pyridine to benzene). We classified perturbations that involved 3 heavy atoms changing or more as "big change" perturbations, and the others as “small change” perturbations. AMBER performs well for "big change" as well as “small change” perturbations: 151 of the 194 "big change" perturbations (~80%) have errors less than 2 kcal/mol, 99 of which have errors less than 1 kcal/mol; 107 of the 136 “small change” perturbations have errors less than 2 kcal/mol, 99 of which have errors less than 1 kcal/mol. Compared to “big change” perturbations, a larger percentage of “small change” perturbations have errors less than 1 kcal/mol: 69% for “small change” vs 51% for “big change” perturbations. Moreover, ring disappearance/appearance and ring type changes are also often seen in perturbation studies and they’re present in this data set as well. From our analysis, we find that AMBER performs well for both: 54 of 68 ring disappearance/appearance perturbations have errors less than 2 kcal/mol, 35 of which have an error less than 1 kcal/mol; 52 of 78 ring type change perturbations have errors less than 2 kcal/mol, 29 of which have an error less than 1 kcal/mol. While it would have been helpful to find systematic issues within certain classes of perturbations when using the AMBER class of force fields, in order to help guide force field improvement efforts, we found this wasn’t the case in the present data set.

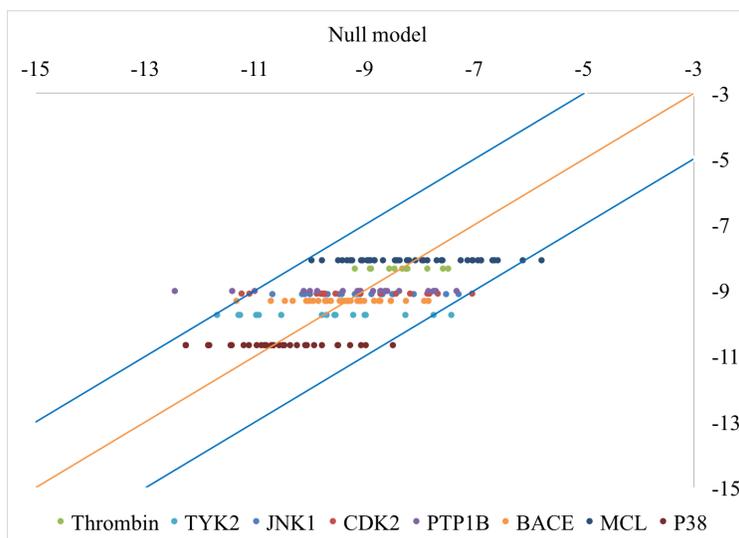
3.3 Conclusion

We repeated the relative binding free energy calculations on the data set described in previous work.⁸ Comparing to the Schrödinger FEP/OPLS 2.1 force field, GPU TI with AMBER FF14SB and the GAFF (1.8) force field performs reasonably well on this data set, with errors above those seen using the FEP/OPLS 2.1 force field. For the 330 perturbations, AMBER has MUE and RMSEs of 1.17 kcal/mol and 1.50 kcal/mol, which is a few tenths of kcal/mol larger than the

reported values (0.90 kcal/mol and 1.14 kcal/mol).⁸ For the 199 ligands, most of the binding free energy values are within 2 kcal/mol, except for 18 ligands (*versus* 5 reported previously⁸). Interestingly, a null model, which assumes all the $\Delta\Delta G$ values are 0 kcal/mol, gives similar results (Figure 18): 8 ligands are not within 2 kcal/mol. This is due to the small range of the experimental ΔG values: the widest range of ΔG values is 5.13 kcal/mol. To better demonstrate and test free energy approaches, data sets with larger experimental ΔG ranges should be explored. Future work will also explore the use of replica exchange and other features within AMBER to enhance the sampling (both in λ -space and orthogonal degrees of freedom). Along with technological advances, we will also explore the capabilities of the next generation GAFF2 and protein force fields. Finally, test procedures for creating benchmark quality results with meaningful error estimates that can be used as a baseline for other comparisons will be explored.

As a final note, Junmei and coworkers followed on our work and tested different protocols and computed the RBFEs of four of the systems using the same force field, where they showed slight improvements of MUE and RMSE.¹⁵¹ Interested readers are directed to the referred article.

Figure 18. Correlation between predicted binding free energies and experimental values for a null model, which has all the $\Delta\Delta G$ set to 0 kcal/mol. X axis: Experimental ΔG (kcal/mol); Y axis: Predicted ΔG (kcal/mol).



CHAPTER 4: THERMODYNAMICS OF TRANSITION METAL ION BINDING TO PROTEINS

This chapter is drawn from the peer-reviewed publication with the title of “Thermodynamics of Transition Metal Ion Binding to Proteins” in the *Journal of the American Chemical Society* authored by Lin Frank Song, Arkajyoti Sengupta, and Kenneth M. Merz. Arkajyoti Sengupta found the protein system that has experimental binding free energies for transition metal ions, performed 12-6-4 potential optimization, and contributed in manuscript composition.

4.1 Introduction

The coordination chemistry of TM ions has found wide-ranging applications in catalyst design²⁰⁶⁻²⁰⁷, energy conversion, biology,²⁰⁸⁻²⁰⁹ assembly of metal organic frameworks²¹⁰⁻²¹¹, *etc.*²¹²⁻²¹⁴ To study these processes computationally, the community has largely resorted to quantum chemistry. However, due to system size limitations only small TM containing cluster can be simulated; hence, the creation of effective simpler models has been an ongoing important research area.²¹⁵⁻²²⁰ An accurate representation of the structure and function and the thermodynamics of assembly of TM containing species is a highly challenging problem in computational chemistry, materials science and biology.²²¹⁻²²² Modeling the structural aspects of a TM binding complex is generally the easier task, with numerous approaches able to reproduce the experimentally observed structural details.²²⁰ However calculation of thermodynamics of TM ion/ligand association has proven to be far more challenging. In order to model the thermodynamics of TM ion binding to a host protein system, both the solvation free energy of the TM must be accurately modeled as well as the interactions of the TM ion with the coordinating groups.

Following the early works of the Kollman and McCammon groups on relative free energy based methods for ions,²²³⁻²²⁴ numerous computational studies have been performed to derive the relative

absolute metal binding free energy in host-guest systems including metalloproteins.²²⁵⁻²²⁸ Steep scaling and the crude approximate entropic contributions restrict ab initio studies to cluster models of metalloenzymes when determining the absolute metal binding energies.²²⁹ The calculated metal binding energies based on these models when consistently used across different metal bound proteins provide a basis for the cancellation of systematic errors and can derive accurate estimates for the relative metal binding energies. Similar ideas are applied on models based on quantum mechanical/molecular mechanical (QM/MM) and Poisson–Boltzmann approaches to systematically analyze metal binding affinity and selectivity. Rao *et. al.* found that their QM/MM model gave binding affinities and selectivity for the copper efflux regulator (CueR) toward different metal ions (Cu^+ , Ag^+ , Au^+ , Zn^{2+} , and Hg^{2+}) that were consistent with experiment.²³⁰ In a recent study, Alexandrova and co-workers apply mixed quantum-classical approach with QM and discrete molecular dynamics (DMD) method.²³¹ This method provides the advantages of fast sampling of protein conformations without the need to rely on parameterizations. Their calculations reproduced the experimentally observed trend in a metal-dependent HDAC8. The extensive sampling needed to describe the (un)binding of the host bound guest to the completely separated species restricts molecular dynamics (MD) techniques to simpler classical models rather than QM based models for the determination of absolute metal binding free energies.

Case and co-workers applied a combination of MD simulations, continuum electrostatics, and normal-mode analysis to derive absolute binding free energies for metal ion binding to RNA.²³²

Kollman and co-workers developed a novel thermodynamic cycle to derive absolute free energies for the binding of cations to a calixspherand.²²⁶ Despite these advances, the functional form of widely used 12–6 Lennard-Jones (LJ) nonbonded metal ion models often fail to simultaneously reproduce the structural and thermodynamics properties of metal ion solvation and its interactions

with metal ion binding groups, which restricts its applicability. Alternatively, polarizable force fields like SIBFA (sum of interactions between fragments ab initio computed),²³³ NEMO (nonempirical molecular orbital),²³⁴ and AMOEBA include polarization effects and charge transfer,²³⁵ to determine accurate metal binding energies. However, the lengthy parameterizations and higher computational cost of the methods restrict their applicability. Macchiagodena *et. al.* very recently developed and validated a novel force field in the context of the AMBER parameterization for simulation of zinc(II)-binding proteins.²³⁶ However, the 12-6 parameters utilized underestimate the Zn^{2+} hydration energy by 70 kcal/mol, which questions the broad applicability of their proposed parametrization. To reproduce the ion hydration free energy and the ion-water distance simultaneously, Li and Merz developed the 12-6-4 LJ-type nonbonded model that includes a $1/r^4$ term to incorporate charge-induced dipole interactions.^{56, 237} This modification has allowed us to reproduce multiple experimental properties of highly charged metal ions.⁵⁵ Recently, we have demonstrated that the properly optimized (m12-6-4) potentials can be effective in modeling a range of properties^{183, 238} including the chelate effect.⁵⁹ In light of these successes we wanted to explore how well the 12-6-4 model could tackle TM ion binding to a model protein system.

Metal ions play critical roles in the structure and function of numerous enzymes and proteins.²³⁹⁻²⁴⁰ The structure activity relationships of enzymes even within the same family may differ depending on the organism involved. For example, Glyoxalase I (GlxI) is the first of two enzymes in the two-component Glx system, and is responsible for the removal of cytotoxic α -ketoaldehydes. GlxI catalyzes the isomerization of the non-enzymically formed hemithioacetal of methylglyoxal and glutathione (GSH) to *S*-D-lactoylglutathione using several transition metal (TM) ions. In *Homo sapiens* and *Pseudomonas putida* GlxI, the essential metal was found to be Zn(II), while

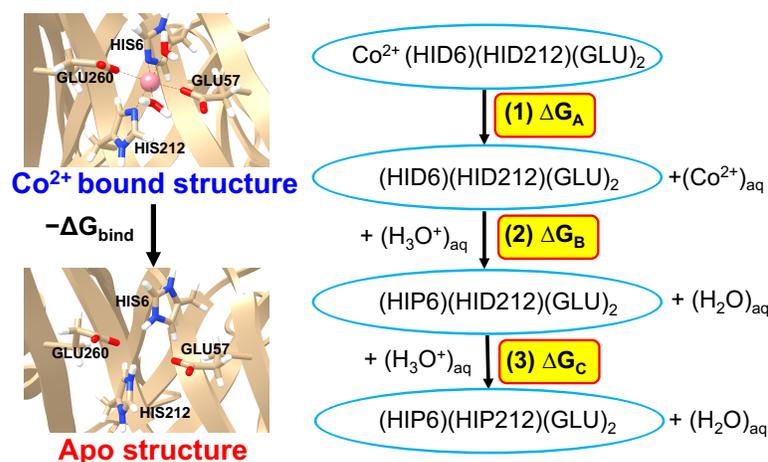
GlxI from *E. coli* is completely inactive in the presence of Zn(II), but was found to have maximal activity in the presence of Ni(II). Clearly, the specific selectivity towards an ion results from a delicate balance of a number of interactions. Hence, a comprehensive understanding of thermodynamics involved in the metal binding to the enzyme active site in aqueous solution is a prerequisite for the understanding, not only of GlxI, but of metal trafficking pathways, metal homeostasis and metal detoxification and for rational design of synthetic motifs²⁴¹⁻²⁴² with predictable properties.²⁴³

Despite the advances in macromolecular structure determination, correlation of structure with accurate thermodynamic data remains less common.²⁴⁴ Isothermal titration calorimetry (ITC) is one such technique that has been widely used to study the thermodynamics of metal ion-protein interactions. The technique is based on the quantitative measure of heat flow associated with a binding event conducted as a titration of one species into the other thereby yielding the binding free energy and by analogy the dissociation constant ($\Delta G = -RT \ln K_d$). While some experimental data are available more information on metal-ligand interactions (*e.g.*, M(II)-His or M(II)-Acetate) along with more data on TM-protein binding interactions would be most welcome in order to further push the present approach. One system that has available experimental data is GlxI where the free energy of binding a number of TM ions have been estimated.²⁴⁵

The GlxI protein is a homodimer and can be viewed as a large chelate complex: each monomer chelates to a M(II) with one histidine residue (HIS) and one glutamic acid residue (GLU) (see Figure 19 left panel). Prior experimental studies have reported the crystal structures and association constants for a range of TM ions to GlxI.²⁴⁵⁻²⁴⁶ We have studied Co^{2+} and Ni^{2+} in this work and will focus on Co^{2+} for our discussion. First the 12-6-4 potentials between the metal ion and the coordinating residues were optimized using available experimental data²⁴⁷⁻²⁴⁸ on imidazole

and acetate interacting with Co^{2+} and Ni^{2+} to mimic the HIS and GLU residues respectively. Then the optimized 12-6-4 potentials were used to simulate the GlxI protein. The simulated structures and the calculated thermodynamics were in excellent agreement with the experimental counterparts. We find that the protonation state change of the HIS residues is very intriguing and the incorporation of the associated free energy change is crucial for the metal ion binding free energy calculations.

Figure 19. Left: The binding site structure of GlxI in Co^{2+} bound (*holo*) and *apo* form. The Co^{2+} (pink) and its coordinating residues (two units each of HIS, GLU and water) are shown in a ball and stick representation. Right: scheme of calculating the binding free energy of Co^{2+} . HID: neutral form of HIS that is protonated at the δ nitrogen; HIP: +1 charged HIS that is protonated both at both the δ and ϵ nitrogens.



4.2 Methods

4.2.1 Optimization of the 12-6-4 Potentials

In the present work we utilized the 12-6-4 nonbonded model along with the AMBER force field:

$$U_{ij}(r_{ij}) = \frac{c_{12}^{ij}}{r_{ij}^{12}} - \frac{c_6^{ij}}{r_{ij}^6} - \frac{c_4^{ij}}{r_{ij}^4} + \frac{e^2 Q_i Q_j}{r_{ij}} \quad (15)$$

where e represents charge of the proton, Q_i and Q_j are partial charges of atoms i and j . The electrostatic interaction between atoms i and j is represented by the Coulomb pair potential, while the van der Waals interaction is represented by the classic Lennard-Jones (12-6) potential plus an

extra r^{-4} term. The C_4 terms between ions and water were parameterized in previous studies by Li et.al⁵⁵⁻⁵⁷. The C_4 terms between ions and other ligands are optimized based on the following equation:

$$C_4(\text{atom type}) = \frac{C_4(H_2O)}{\alpha_0(H_2O)} \times \alpha_0(\text{atom type}) \quad (16)$$

where α_0 is an atom type dependent polarizability. The metal binding site consists of two units of GLU, HIS and water each interacting with the metal ion. We used acetate and imidazole to mimic GLU and HIS amino acids respectively. Potential of mean force (PMF) calculations were used to optimize the pairwise parameters to reproduce the experimental free energies of metal binding with the individual ligands as shown in Figure 20 and Figure 21.

4.2.1.1 PMF calculations

Since the model systems are used to mimic the amino acid residues, we constrained the respective amino acid charges on the model systems. The RESP fitting was performed to derive the charges on the remaining atoms using antechamber. Parmchk2 was used to generate the frcmod files. GAFF2 was used as the force field for imidazole and acetate.

A free MD is first performed with 200 ps gradual heating under constant NVT condition, 1ns equilibration under constant NPT condition and 1ns sampling under constant NVT condition. The final geometry was then used to generate metal ligand complexes at various constrained distances using steered molecular dynamics. The constrained geometries were then used for umbrella sampling (US) studies to generate the potential of mean force (PMF). The US windows were spaced every 0.05 Å from around 2 Å to 5 Å and 0.1 Å from 5 Å to 11 Å. For each US window, 50000 steps of steepest descent minimization was performed followed by 50000 steps of conjugate gradient descent minimization. Afterwards the system was heated gradually from 0 to 300 K in 200 ps, followed by 2 ns NPT equilibration and 8 ns NVT production at 300 K. The reaction

coordinate distance was recorded every 100 steps with the step size of 2 fs. Weighted histogram method (WHAM) was used to generate the free energy profile with respect to the reaction coordinate. Berendsen barostat was used for pressure control and the pressure relaxation time was set to 5 ps. Langevin thermostat was used to maintain the constant temperature with a collision frequency of 2 ps^{-1} . The time step was 2 fs and the nonbonded cut off was 10 \AA . The restraint constant for each window was fine-tuned to ensure that the sampled distances are distributed around the targeted value and that neighboring windows overlap.

Figure 20. Comparison of potential of mean force (PMF) profiles for the default 12-6-4 and optimized m12-6-4 pairwise parameters for the Co^{2+} ion interacting with imidazole and acetate.

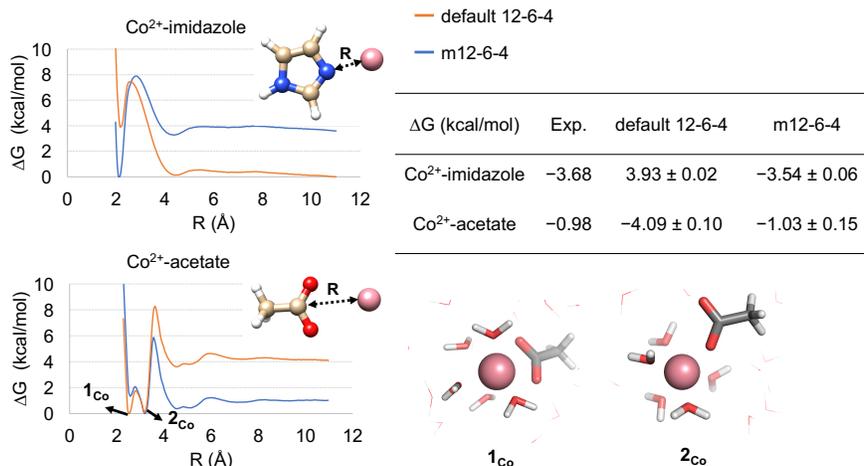
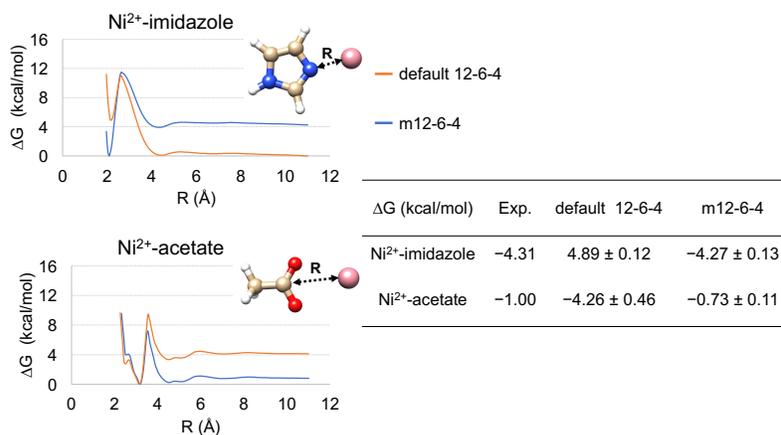


Figure 21. Free energy profiles calculated with the default and the optimized alpha values for Ni^{2+} acetate and Ni^{2+} imidazole complexes.



4.2.2 *Binding Free Energy Calculations*

4.2.2.1 *System Preparation*

The crystal structures of the Co²⁺ bound and Ni²⁺ bound GlxI of *Escherichia Coli* were downloaded from the Protein Data Bank (1FA6 and 1F9Z).²⁴⁶ The prepwizard utility of Schrödinger 2018-1 suite was used to add missing residues.²⁴⁹ Protonation states of the charged residues were determined by H++ server and were carefully visually examined.¹⁹⁸ The LEaP module of the assisted model building with energy refinement tools (AMBERTools 19, updated August 2019) was used to generate the topologies for MD simulations.²⁵⁰ The system was solvated by TIP3P⁹⁵ water molecules in a truncated octahedral simulation cell with a minimum of 10 Å from the solute to the cell boundary. The AMBER ff14sb force field was used to describe the protein.²⁵¹ Na⁺ ions described by the default 12-6-4 parameter were added to neutralize the system.⁵⁶

4.2.2.2 *Free MD simulation*

Five steps of minimization were performed to remove close contacts. The first step minimizes water molecules and counter ions, with the protein restrained. The second, third, fourth step restrains the heavy atoms, backbone heavy atoms, backbone carbon and oxygen atoms of the protein respectively, with the last step minimizing the whole system. Each minimization step consists of 10000 cycles of minimization using the steepest descent method. Afterwards the system was heated up to 300 K gradually during 1 ns NVT simulation with a weak coupling restraint (5 kcal/(mol*Å²)) on the protein. Then the density was equilibrated by six steps of NPT simulation at 300K with each step being 1 ns timescale. The restraint on the protein was gradually reduced from 5 kcal/(mol*Å²) to 0 kcal/(mol*Å²) during the NPT equilibrations. Finally 300 ns production run was performed under NPT condition at 300 K. Berendsen barostat was used for pressure

control and the pressure relaxation time was set to 5 ps. The Langevin thermostat was used to maintain a constant temperature with a collision frequency of 2 ps^{-1} . The time step was 2 fs. All simulations were performed using the CUDA version of PMEMD from the AMBER18 (updated August 2019) package.²⁵⁰ The routinely used particle mesh Ewald (PME) method was used to treat the long-range electrostatic interactions^{192, 252} and a 10 Å cutoff was used for the nonbonded interactions. All bonds with hydrogen atoms were constrained using SHAKE.²⁵³ The visual molecular dynamics (VMD) program was used to analyze the generated trajectories.²⁵⁴ The structure and velocity of the last snapshot at 100 ns, 200 ns and 300 ns were used for subsequent energy calculations: the results of the three sets of calculations are reported.

4.2.2.3 Energy Calculations

Figure 19 provides a schematic representation of the total free energy change involving (a) loss of metal ion and (b) subsequent protonation in the active site.

4.2.2.3.1 Loss of metal ion

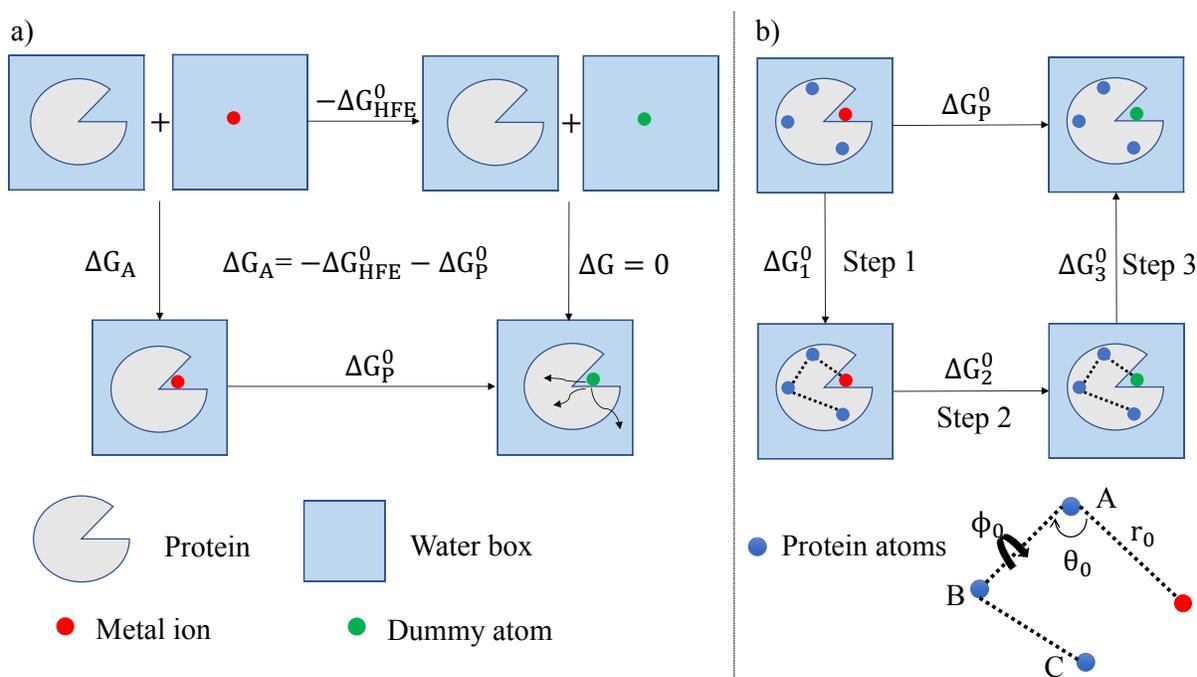
As mentioned in the above section, three sets of calculations were performed using the structure and velocity of the snapshot at 100 ns, 200 ns and 300 ns free MD simulation. Herein, we used the double decoupling method (DDM) strategy described by Boresch and Karplus to determine the free energy associated with the loss of the metal ion.²⁵⁵

4.2.2.3.1.1 The DDM theory and procedure

The DDM divides the calculation of binding free energy into two decoupling processes: one is the decoupling of the metal ion in water, which gives the negative of the hydration free energy (HFE) of the metal ion, *i.e.* $-\Delta G_{\text{HFE}}^0$, the other is decoupling of the metal ion in protein, which gives ΔG_{P}^0 (see Figure 22a). Both processes can be simulated by alchemical free energy methods. The latter process is more challenging since at the end state of this process, the dummy atom is free to

wander, hence the simulation is required to sample every possible position of the dummy atom, resulting in huge sampling issue. To circumvent this problem, a three-step-method is constructed (Figure 22b)). Step 1 represents the process of turning on some restraints on the metal ion. Step 2 is decoupling the metal ion in the protein with the restraints on the metal ion. Step 3 is turning off the restraints on the dummy atom. Step 1 and step 2 again can be simulated by alchemical free energy methods and step 3 can be described by an analytical equation. This protocol is based on the pioneering works on calculating absolute binding free energy of protein-ligand systems by Karplus, Roux, et al.^{61, 116} The details of each step are described below.

Figure 22. (a) Scheme of the DDM method. Dummy atom is an atom that has no interaction with the surroundings, so it can be viewed as the metal ion in gas phase. (b) Scheme of calculating the ΔG_P^0 . Dashed lines mean that the metal ion is restraint to the binding site by restraining to three of the protein atoms through distance, angle and dihedral restraints.



4.2.2.3.1.2 Hydration free energy (ΔG_{HFE}^0)

The ΔG_{HFE}^0 for Co^{2+} and Ni^{2+} ions are obtained from previous work by Pengfei Li, et al.⁵⁷

Simulation details can be found in the referenced publication.

4.2.2.3.1.3 Step1 of Figure 22(b) (ΔG_1^0)

ΔG_1^0 is calculated by free energy perturbation (FEP) based on the Zwanzig equation:

$$\Delta G_1^0 = -k_B T \ln \langle \exp(-\frac{E_1 - E_0}{k_B T}) \rangle_0 \quad (17)$$

where k_B is the Boltzmann constant, T is the simulation temperature, and E_1 and E_0 is the potential energy of the initial and the end state. The difference between the two potential energies is the restraint energy, which is calculated by:

$$E_{\text{restraint}} = \frac{1}{2}k_r(r - r_0)^2 + \frac{1}{2}k_\theta(\theta - \theta_0)^2 + \frac{1}{2}k_\phi(\phi - \phi_0)^2 \quad (18)$$

where r , θ , ϕ , k_r , k_θ , k_ϕ represents distance, angle, dihedral and the corresponding force constants. The equilibrium distance, angle and dihedral values (r_0 , θ_0 , ϕ_0) were calculated by taking the average values from the 100 ns, 200 ns, and 300 ns production run of the free MD simulation for the first, second, third set of DDM calculations, respectively. The structure and velocity at the 100 ns, 200 ns, and 300 ns production run were used to start the first, second, third set of calculations, respectively. For each set of calculations, multiple intermediate states were used to gradually turn on the restraints, with a linear coupling parameters k (with the value of 0, 0.0025, 0.005, 0.0075, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.4, 0.6, 0.8 and 1.0). For each intermediate state, 1 ns NPT simulation at 300 K was performed to equilibrate the system and 4 ns NVT production simulation at 300 K was performed to collect the distance, angle and dihedral values. The equilibrated structure and velocity from the 1 ns NPT equilibration were used for the 1 ns equilibration of the next intermediate state. Berendsen barostat was used for pressure control and the pressure relaxation time was set to 5 ps. Langevin thermostat was used to maintain the constant temperature with a collision frequency of 2 ps^{-1} . The time step was 2 fs and the nonbonded

cut off was 10 Å. The equilibrated structure and velocity from the 1 ns NPT equilibration of the last intermediate state ($k=1.0$) was used for the following step.

4.2.2.3.1.4 Step2 of Figure 22(b) (ΔG_2^0)

Two stages of TI simulations were performed using the AMBER18 GPU TI (updated August 2019).¹⁴⁷ The first stage decouples the electrostatic interactions between the metal ion and the surroundings, and the second stage decouples the vdW interactions. 14 λ windows of simulations (0, 0.00922, 0.04794, 0.11505, 0.20634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, 0.99078, 1.0) were performed and the Gaussian quadrature method was employed to calculate the free energy values. For each window, 1 ns NPT simulation at 300 K was performed to equilibrate the system and 4 ns NVT production simulation at 300 K was performed to collect the $\partial U/\partial \lambda$ data. The equilibrated structure and velocity from the 1 ns NPT equilibration were used for the 1 ns equilibration of the next window. Langevin thermostat was used to maintain the constant temperature with a collision frequency of 2 ps⁻¹. The nonbonded cut off was 10 Å. For the electrostatic decoupling stage, the linear mixing potential was used and the time step was 2 fs. For the vdW decoupling stage, the softcore potential was used, the SHAKE was not used and the time step was 1 fs. The parameter α and β of softcore potential was 0.5 and 12 Å², respectively. With the collected $\partial U/\partial \lambda$ from each window and the corresponding weights, ΔG_2^0 was calculated.

4.2.2.3.1.5 Step3 of Figure 22(b) (ΔG_3^0)

Here is the final equation:

$$\Delta G_3^0 = -RT \ln \left(\sqrt{k_r k_\theta k_\phi} * V_0 / (\sqrt{(2\pi k_B T)^3} * r_0^2 * \sin \theta_0) \right) \quad (19)$$

where V_0 is the standard volume (1661 Å³) for a one molar standard state, R is the gas constant, and other terms were defined in above sections. ΔG_3^0 was derived based on the same idea from the work by Karplus, et al.¹¹⁶

Consider the process of step3: $(P\text{-----}D)_{\text{aq}} = (P)_{\text{aq}} + (D)_{\text{aq}}$,

where $(P\text{-----}D)_{\text{aq}}$ represents the complex in solution and the dummy atom is restrained on the protein, $(P)_{\text{aq}}$ and $(D)_{\text{aq}}$ represents the protein and dummy atom in solution, respectively. The free energy change of this process can be calculated as (refer to equation 31 of ref 1):

$$\Delta G_3^0 = -RT \ln (V_0 * Z_P * Z_D)/(V * Z_{P-D}) \quad (20)$$

V is the simulation volume, and Z_P , Z_D and Z_{P-D} is the configurational partition function of the protein, the dummy atom, and the complex in solution.

Using the internal coordinate, $Z_P * Z_D$ can be written as:

$$Z_P * Z_D = Z_P * V \quad (21)$$

and Z_{P-D} can be written as:

$$\begin{aligned} Z_{P-D} &= \int d\mathbf{R} \exp\left(-\frac{U(\mathbf{R})}{k_B T}\right) \\ &= Z_P * \int dr d\theta d\phi r^2 \sin \theta \exp\left(-\frac{U(r) + U(\theta) + U(\phi)}{k_B T}\right) \quad (22) \end{aligned}$$

where \mathbf{R} represents all the degrees of freedom, $U(r)$, $U(\theta)$, and $U(\phi)$ is the restraint potential:

$$U(r) = \frac{1}{2} k_r (r - r_0)^2 \quad (23)$$

$$U(\theta) = \frac{1}{2} k_\theta (\theta - \theta_0)^2 \quad (24)$$

$$U(\phi) = \frac{1}{2} k_\phi (\phi - \phi_0)^2 \quad (25)$$

Combining equation (22) to (25), we get:

$$Z_{P-D} = Z_P * \left(r_0^2 * \sin \theta * \sqrt{\frac{(2\pi k_B T)^3}{k_r k_\theta k_\phi}} \right) \quad (26)$$

Plug equation (21) and (26) into equation (20), we can obtain the final equation for ΔG_3^0 (equation 19), in which all the terms are known constants or pre-set restraint values, so that ΔG_3^0 can be calculated numerically.

4.2.2.3.1.6 Restraint set-up

The information of the restraints is summarized in Table 14. Again, the equilibrium distance, angle and dihedral values were calculated by taking the average values from the 100 ns, 200 ns, and 300 ns production run of the free MD simulation for the first, second, third set of DDM calculations, respectively.

Table 14. The distance, angle and dihedral restraints for the three sets of DDM calculations on the GlxI-Ni²⁺ system. See Figure 22 for the info of the restraints and protein atoms A, B, and C. Nomenclature :7(THR)@C means backbone carbonyl carbon from number 7 residue, which is a THR amino acid (Threonine).

GlxI-Ni ²⁺ A :7(THR)@C ; B :8(MET)@C; C :212(HID)@C		r ₀	θ ₀	φ ₀
Set 1	Default 12-6-4	6.47 Å	87.80°	68.74°
	Optimized 12-6-4	6.52 Å	86.27°	67.59°
Set 2	Default 12-6-4	6.50 Å	87.56°	68.28°
	Optimized 12-6-4	6.51 Å	86.51°	67.65°
Set 3	Default 12-6-4	6.51 Å	87.47°	68.19°
	Optimized 12-6-4	6.50 Å	86.95°	67.71°

4.2.2.3.2 Subsequent protonation

In the calculations involving DDM we assume the protonation state of the binding site residues do not change upon release of the metal ion. However, in the *apo* crystal structure, the two HIS and two GLU residues have similar structural features as in the *holo* (Co²⁺ bound structure) form (Fig 1), hence we hypothesized that upon release of the +2 charged metal ion, the two HIS residues

become protonated thereby conserving the total charge of the system. In this case incorporating the free energy change associated with the protonation of the two HIS residues is crucial. Similar considerations were shown to be important for the zinc transporter, ZnT₂, where “two protons are exchanged for each zinc ion transported”.²⁵⁶ Moreover, ZnT₂ also has two HIS and two carboxylate groups (ASP in this case) in the binding site, and the authors concluded that it was the two HIS that were protonated, which is what we observed here as well. Herein, we performed pKa calculations using TI to obtain the free energy change of protonating the two HIS residues.

The free energy changes of perturbing HIP to the HID in both water and protein were calculated by TI, from which the pKa of the HIS residue can be computed, see Figure 27 for the HIS6 example.

4.2.2.3.2.1 System preparation

For the perturbations in water, N-terminal and C-terminal residues, i.e. NHCH₃ and CHCH₃, were used to cap the HID (or HIP) residue. The systems were solvated by TIP3P water molecules in a truncated octahedral cell with a minimum of 20 Å and 10 Å from the solute to the cell boundary for perturbations in water and in protein, respectively. Na⁺ ions described by the default 12-6-4 parameter were added to neutralize the system. The AMBER ff14sb force field was used to describe the amino acid residues. The “timerge” function of the parmed utility of the AMBER 18 package (updated August 2019) was used to generate the topology for TI simulations.

4.2.2.3.2.2 TI simulations

First, 100 ns free MD was performed to equilibrate the system. With the equilibrated structure and velocity, the one-step protocol is used to disappear HIP and appear HID simultaneously. SHAKE was not used. Both the charge and vdW interactions between the disappearing (or appearing) unique atoms with the surrounding atoms were described by the softcore potentials. The other

simulation settings were the same as section 4.2.2.3.1.4. Nine independent runs were performed for each perturbation and the averaged values are reported.

4.3 Results and Discussion

4.3.1 Optimization of the 12-6-4 Potentials

As shown in Figure 20 and Figure 21, the default 12-6-4 potential was found to either underestimate (for imidazole) or overestimate (for acetate) the binding free energy of the TM ion. Moreover, simulations using the default 12-6-4 potential for acetate yielded a too stable bidentate complex (the $1C_6$ in Figure 20) relative to experiment. However, on optimizing the 12-6-4 potential to reproduce the experimental binding free energy we also were able to observe the experimental monodentate binding mode (the $2C_6$ in Figure 20). These findings align with our experiences from our previous study on the chelate effect.⁵⁹ The optimized α_0 and C_4 values are listed in Table 15. The alpha values were optimized to obtain an average binding energy of three independent runs within 0.3 kcal/mol of experimental binding energy.

Table 15. The optimized α_0 and C_4 terms. α_0 is the polarizability for atom type of “nd” and “o” for imidazole and acetate, respectively.

	Default 12-6-4			Optimized 12-6-4			ΔG_{bind} (Expt.) (kcal/mol)
	α_0	C_4^{ij}	ΔG_{bind} (Calc.) (kcal/mol)	α_0	C_4^{ij}	ΔG_{bind} (Calc.) (kcal/mol)	
Co ²⁺ - imidazole	1.090	158.292	3.93 ± 0.02	2.230	323.844	-3.54 ± 0.06	-3.68 ²⁴⁷
Co ²⁺ -acetate	0.569	82.631	-4.09 ± 0.10	0.120	17.427	-1.03 ± 0.15	-0.98 ²⁴⁸
Ni ²⁺ - imidazole	1.090	160.632	4.89 ± 0.12	2.310	340.421	-4.27 ± 0.13	-4.31 ²⁴⁷
Ni ²⁺ -acetate	0.569	83.853	-4.26 ± 0.46	0.145	21.368	-0.73 ± 0.11	-1.00 ²⁴⁸

4.3.2 Geometries of Metalloproteins

The MD simulations on the GlxI-M(II) complexes with the optimized 12-6-4 (m12-6-4) parameters resulted in structural features that maintained the metal coordination environment. Figure 23 represents the geometries of Co^{2+} bound protein obtained after 300 ns of free MD simulations with default 12-6-4 and m12-6-4 potential. The octahedral coordination is well preserved; over the 300 ns free MD simulation, the average root mean square deviation (RMSD) for the binding site comparing to crystal structure is 0.65 and 0.68 Å for the default 12-6-4 and m12-6-4 potential, respectively. The bond distances between the metal ion and coordinating residues are summarized in Table 16.

Figure 23. Geometries of Co^{2+} bound protein obtained after 300 ns MD simulations with (a) default 12-6-4 and (b) m12-6-4 potential aligned with crystal structure (light orange). The RMSD measurements are based on the side chain of the two HIS, two GLU, two water molecules along with the metal ion.

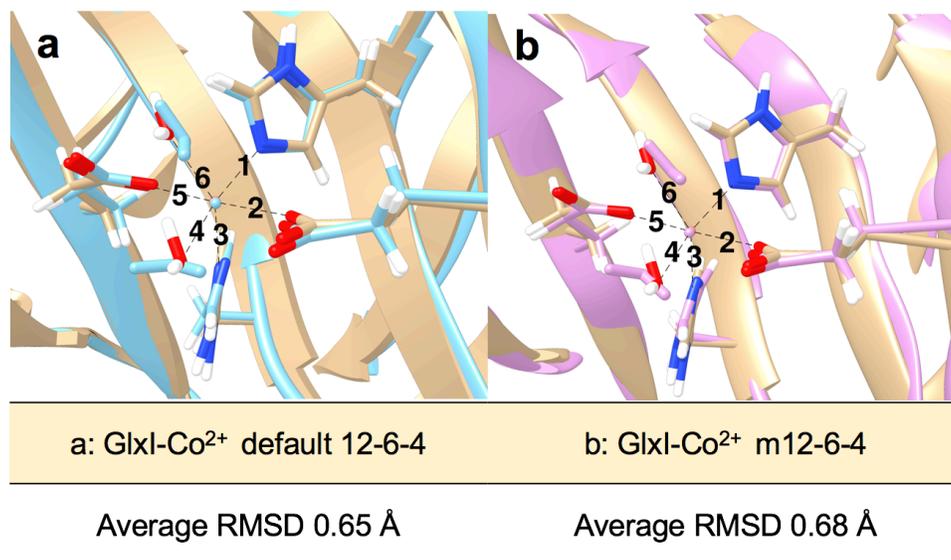


Table 16. The coordinating bond distances. The simulated bond distance is the average bond distance from the 300 ns free MD simulations.

Coordinating bond distances (Å)	(a) GlxI-Co ²⁺ default 12-6-4	(b) GlxI-Co ²⁺ optimized 12-6-4	Crystal structure (1FA6)	(c) GlxI-Ni ²⁺ default 12-6-4	(d) GlxI-Ni ²⁺ optimized 12-6-4	Crystal structure (1FA9)
Label 1 (Co,HID6@NE2)	2.21 ± 0.06	2.14 ± 0.05	2.24	2.16 ± 0.06	2.09 ± 0.04	2.03
Label 2 (Co,GLU57@OE1)	2.04 ± 0.04	2.06 ± 0.04	2.12	2.00 ± 0.04	2.03 ± 0.04	2.16
Label 3 (Co,HID212@NE2)	2.23 ± 0.06	2.15 ± 0.05	2.34	2.18 ± 0.06	2.11 ± 0.05	2.15
Label 4 (Co,WAT1)	2.12 ± 0.05	2.12 ± 0.05	2.22	2.08 ± 0.05	2.08 ± 0.05	2.10
Label 5 (Co,GLU260@OE1)	2.04 ± 0.04	2.06 ± 0.04	2.12	2.00 ± 0.04	2.03 ± 0.04	2.09
Label 6 (Co,WAT2)	2.12 ± 0.05	2.13 ± 0.05	2.41	2.08 ± 0.05	2.09 ± 0.05	2.29

4.3.3 Binding Free Energy Calculations

As shown in the Figure 19 the total metal binding free energy includes the free energy change corresponding to the loss of metal ion (ΔG_A) and the subsequent protonation of the binding site residues (ΔG_B and ΔG_C). ΔG_A was calculated using the DDM and includes three steps as discussed in the section 4.2.2. To ensure the convergence of the results, three sets of DDM calculations were carried out, which started with the structure and velocity from the last snap-shot of the 100 ns, 200 ns, 300 ns free MD simulations, respectively. The free MD simulations on the metal bound proteins provided a good estimate of the equilibrium distance, angle, and dihedral values that the restraints should use to hold the metal ion, see Figure 24. The equilibrium distance, angle and dihedral values

are all about the same for the three sets of calculations, and for each set, the distributions of the distance, angle and dihedral values are close to Gaussian shape and are centered around the average values with small fluctuation ranges (see Figure 25 and Figure 26).

Figure 24. The distance, angle and dihedral restraints for the three sets of DDM calculations on the GlxI-Co²⁺ system. r_0 , θ_0 , ϕ_0 is the equilibrium distance, angle and dihedral values. ‘:7(THR)@C’: backbone carbonyl carbon of number 7 residue, which is a THR (Threonine) amino acid. Set 1, Set 2 and Set 3: the three sets of DDM calculations starting with the structure and velocity from the last snapshot of the 100 ns, 200 ns, 300 ns free MD simulations, respectively.

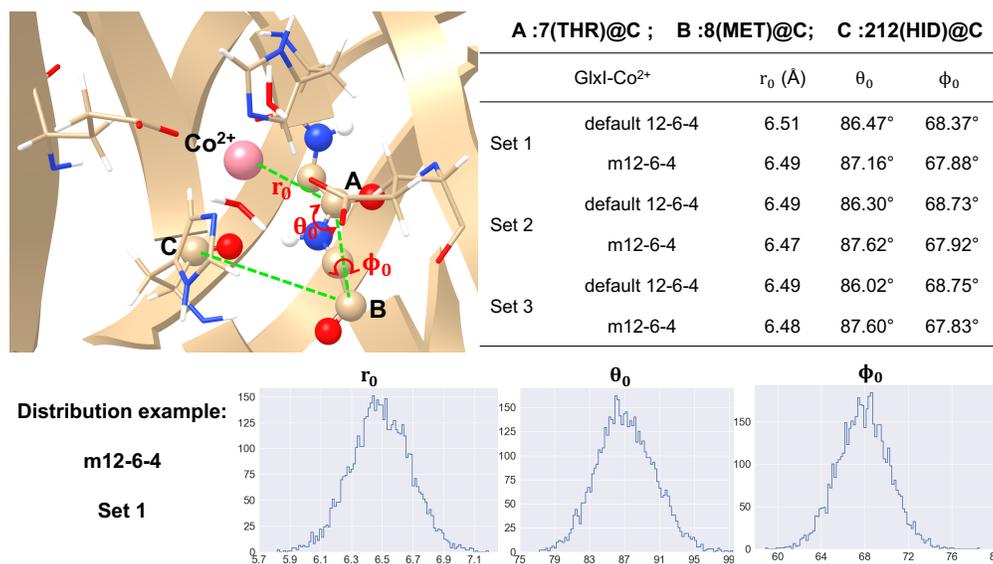


Figure 25. Distributions of the selected distance, angle and dihedral in the free MD simulations for Glx-Co²⁺. The averaged values were used as the equilibrium distance, angle and dihedral values for the three set of DDM calculations for the default and optimized 12-6-4 potential.

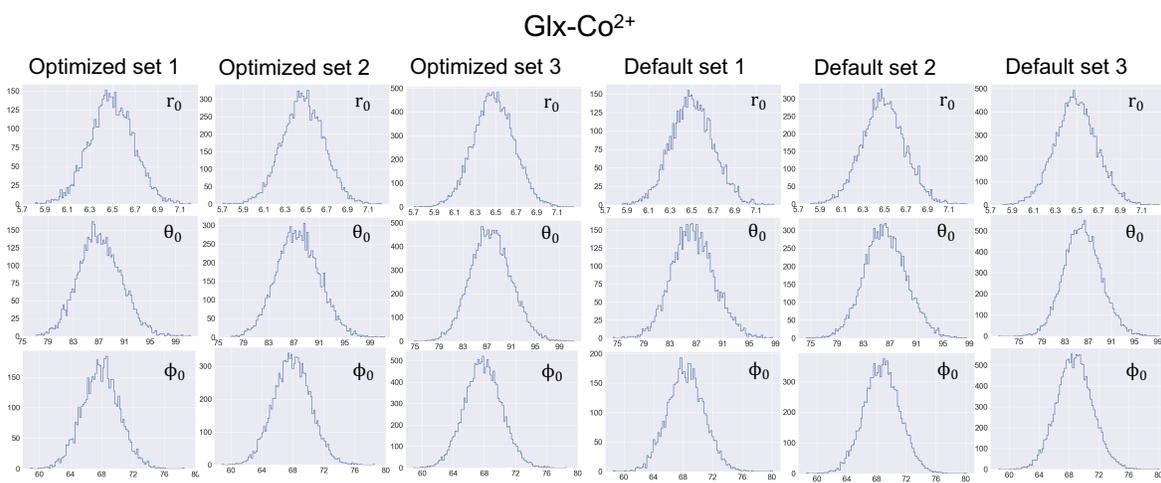
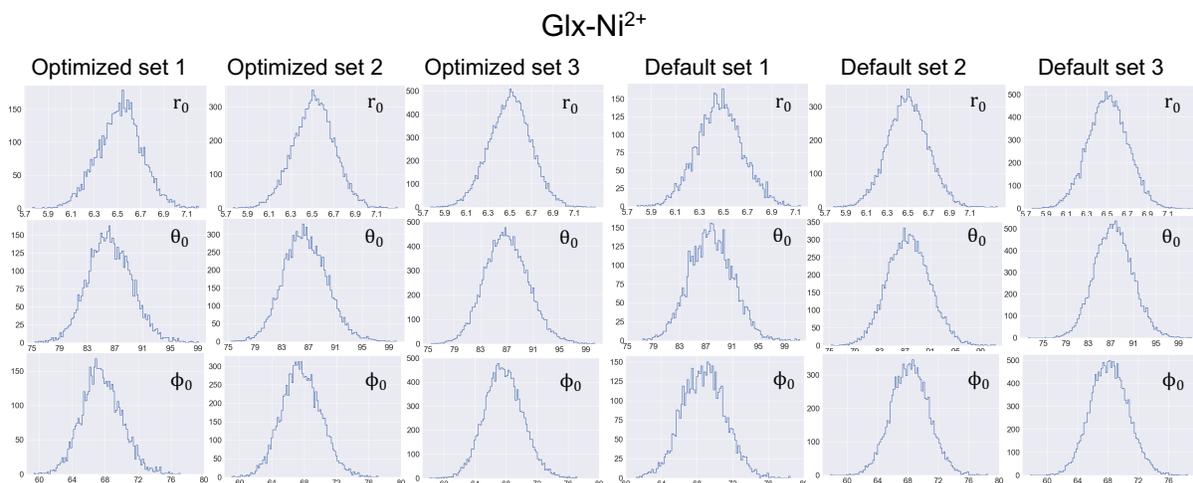


Figure 26. Distributions of the selected distance, angle and dihedral in the free MD simulations for Glx-Ni²⁺. The averaged values were used as the equilibrium distance, angle and dihedral values for the three set of DDM calculations for the default and optimized 12-6-4 potential.



For the DDM calculations, within each set, nine calculations were performed with different restraint strength. The calculated ΔG_A values for the GlxI-Co²⁺ system are listed in Table 17. The standard deviation for each set of calculations is within 3.0 kcal/mol. The overall ΔG_A averaging the three sets of calculations is 31.5 ± 3.6 kcal/mol and 39.0 ± 2.6 kcal/mol for the default and optimized 12-6-4, respectively. The standard deviations are relatively small compared to the free energy change of disappearing the metal ion in protein we computed (~ 500 kcal/mol).

Table 17. Summary of ΔG_A values of the GlxI-Co²⁺ system by the double decoupling method (DDM). For each system, nine runs were performed using different restraint strength. Set 1, Set 2 and Set 3 are the three sets of DDM calculations starting with the structure and velocity from the last snap-shot of the 100 ns, 200 ns, 300 ns free MD simulations, respectively.

Restraint force constants \mathbf{k}_r : kcal/(mol*Å ²), \mathbf{k}_θ : kcal/(mol*rad ²) , \mathbf{k}_ϕ : kcal/(mol*rad ²)	ΔG_A (kcal/mol), GlxI-Co ²⁺					
	Set 1		Set 2		Set 3	
	Default 12-6-4	m 12-6-4	Default 12-6-4	m 12-6-4	Default 12-6-4	m 12-6-4
400,400,400	31.7	43.8	30.8	40.4	32.5	36.8
500,500,500	30.7	36.8	36.3	41.2	30.0	40.6

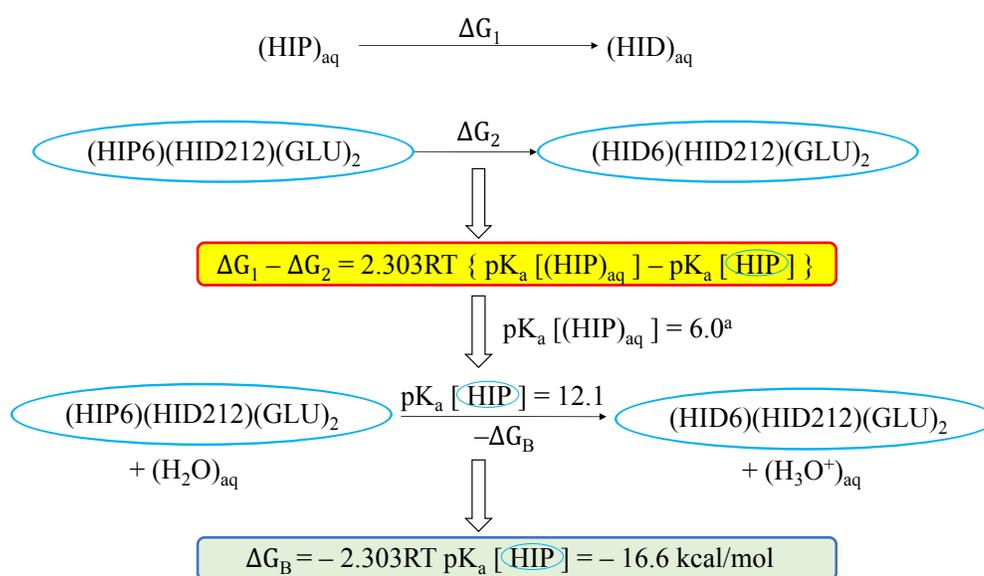
Table 17 (cont'd)

600,600,600	31.3	39.9	34.9	39.1	27.3	37.4
700,700,700	30.2	40.0	37.4	36.9	28.2	38.2
800,800,800	30.5	42.1	28.4	37.1	30.5	38.1
900,900,900	32.4	39.6	32.6	38.0	28.1	38.1
1000,1000,1000	34.6	44.5	37.8	33.7	28.3	38.7
1100,1100,1100	31.4	42.0	33.1	37.0	23.6	36.4
1200,1200,1200	39.7	41.0	31.7	40.6	27.3	34.6
Average	32.5	41.1	33.6	38.2	28.4	37.7
Standard Deviation	3.0	2.3	3.2	2.4	2.5	1.7
Overall	Default 12-6-4	m12-6-4				
Average	31.5	39.0				
Standard Deviation	3.6	2.6				

Figure 27 represents the scheme to obtain the free energy change corresponding to the protonation of the HIS6, *i.e.* ΔG_B . Thermodynamic integration (TI) calculations were performed to determine the free energy changes ΔG_1 and ΔG_2 for the deprotonation of a free HIP unit in water and the HIP unit in the protein respectively. The difference between the alchemical free energy changes, *i.e.* $\Delta G_1 - \Delta G_2$, is used to determine a pK_a value of 12.1 ± 1.2 for the specific HIP unit.²⁵⁷ The calculated pK_a was subsequently used to calculate ΔG_B . Similar calculations for the HIS212 results in the determination of ΔG_C . The calculated ΔG_B and ΔG_C were found to be -16.6 and -16.4 kcal/mol. Both protonation state changes correspond changing from HID form to HIP form, with the hydrogen added on the ϵ nitrogens which were coordinating with the TM ion. The calculated pK_a values are large for both HIS residues, indicating a highly negatively charged sphere after the loss of the +2 charged TM ion. The corresponding ΔG_B and ΔG_C are crucial contributions of the computed TM ion binding free energy. Herein we considered both HIS residues being HIP form

after the loss of the TM ion, however, more discussion on other possible protonation states will be discussed in the next section.

Figure 27. Scheme for computing the free energy change (ΔG_B) associated with the protonation of the HIS6 based on a pK_a shift calculation. \circ represents the protein. ^a approximate value, see reference.²⁵⁸



The calculated ΔG values of Co^{2+} and Ni^{2+} binding to GlxI using both default and optimized 12-6-4 parameters are listed in Table 18. The calculated ΔG_{bind} by the default 12-6-4 model underestimates the binding free energy with respect to the experimental value by more than 10 kcal/mol for both ions, respectively. The calculations with m12-6-4 predicts a ΔG_{bind} of -6.0 ± 3.5 kcal/mol and -7.9 ± 4.7 kcal/mol for Co^{2+} and Ni^{2+} ion respectively, against the experimental value of > -9.6 kcal/mol. The error bars are around or less than 4.5 kcal/mol and are considerably small considering the large free energy change of disappearing the metal ion in protein we computed (~ 500 kcal/mol). Moreover, we find against the default 12-6-4 potential, the optimized m12-6-4 potential model corrects the Co^{2+} binding free energy to imidazole by around -7.5 kcal/mol and the binding free energy of Co^{2+} to acetate by around $+3.1$ kcal/mol (Figure 20). With two HIS and

two GLU in the binding site of the protein, the optimized m12-6-4 potential should give a ΔG_{bind} more negative than the default 12-6-4 potential by $2*(-7.5+3.1) = -8.8$ kcal/mol, assuming the terms are additive. This agrees with the difference between ΔG_{bind} obtained by m12-6-4 and by the default 12-6-4 potential, i.e. -7.5 kcal/mol. Similar agreement is also found for the Ni^{2+} ion.

Table 18. Summary of the overall averaged ΔG (in kcal/mol) results.

	GlxI-Co^{2+}		GlxI-Ni^{2+}	
	12-6-4	m12-6-4	12-6-4	m12-6-4
ΔG_A	31.5 ± 3.6	39.0 ± 2.6	29.5 ± 4.0	40.9 ± 4.1
ΔG_B	-16.6 ± 1.7			
ΔG_C	-16.4 ± 1.6			
$-\Delta G_{\text{bind}}$	-1.5 ± 4.3	6.0 ± 3.5	-3.5 ± 4.6	7.9 ± 4.7
$-\Delta G_{\text{bind}}^{\text{exp}}$	> 9.6		> 9.6	

4.3.4 Apo State Discussion

To explore the other possibilities of the *apo* state, we first performed free MD simulations with combinations of different protonation states of the two HIS residues numbered 6 and 212 in the present system: HIP6_HID212, HIP6_HIE212, and HIP6_HIP212. Figure 28 shows the RMSD of the 100-ns free MD trajectories with respect to the *apo* crystal structure (PDB ID: 1fa6); the measurements of RMSD are based on the heavy atoms of the two HIS and the two GLU in the binding site. All the three combinations have rather low averaged RMSD and the *apo* crystal structure was well reproduced. Herein we only considered HIS6 being protonated form (HIP6), because the computed pKa of HIP6 at both nitrogen positions are greater than 7.0 (10.6 for the δ nitrogen and 12.1 for the ϵ nitrogen), as shown in Figure 29. The H++ server confirms our result: the estimated pKa of HIP6 for the δ nitrogen is 9.2.

As mentioned in above sections and as is shown in Figure 29, the HID212 is also very likely to be protonated since the corresponding pKa for the ϵ nitrogen is calculated to be 11.9. Moreover, with

a careful observation on the 100 ns free MD simulations with HIP6_HID212, we observe a Na⁺ ion in the binding site to interact with one of the GLUs (see Figure 30), suggesting that the binding site is too negative charged. This confirms with our pKa calculation that the HIS212 should be protonated at the ϵ nitrogen position. Hence, the only two remaining possibilities are HIP6_HIE212 and HIP6_HIP212. Although our calculations on the pKa of HIP212 at the δ nitrogen position gives a pKa of 9.0, the H++ server estimates a lower pKa (6.2). The HIS212 δ nitrogen is more solvent accessible than the other nitrogens of the HIS6 and HIS212 (see Figure 30), which explains its lower pKa.

Figure 28. Upper panel: geometries of apo protein obtained after 100 ns MD simulations with different HIS protonation states aligned with the *apo* crystal structure (light orange); lower panel: RMSD of the heavy atoms of the two HIS and the two GLU in the binding site comparing to the apo crystal structure (PDB ID: 1fa6).

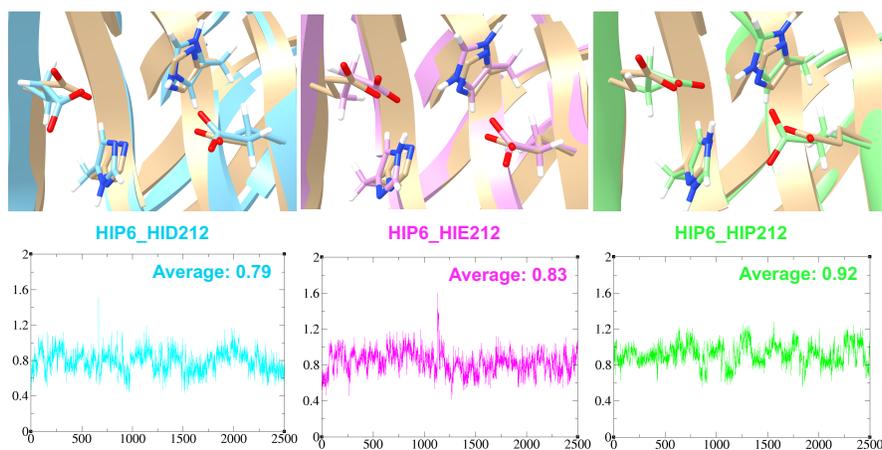


Figure 29. The calculated pKa of the HIS6 and HIS212 at both the ϵ and the δ positions. pKa^{C} is the value calculated by TI using method similar to Figure 27; pKa^{H} is the value estimated by H++ server.

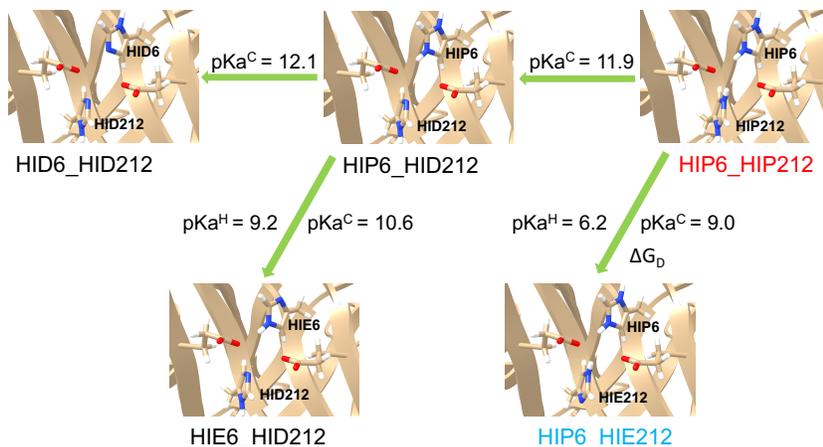
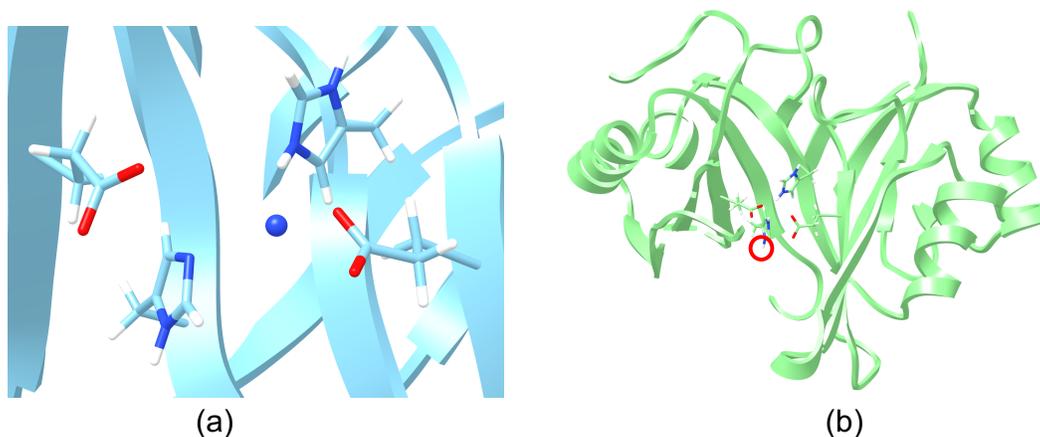


Figure 30. (a) The snapshot after 100 ns MD simulation with the two HIS residues being HIP6_HID212, the blue ball represents the Na^+ ion that came into the binding site; (b) the bottom part is more open to solvent, and the red circled hydrogen is the hydrogen on the δ nitrogen position of HIS212.



Therefore, with the two possible *apo* states, the ΔG_{bind} could be in a range. As shown in Table19, the ΔG_{bind} for m12-6-4 for GlxI- Co^{2+} and GlxI- Ni^{2+} is $-6.0 \pm 3.5 \sim -18.3 \pm 3.6$ kcal/mol and $-7.9 \pm 4.7 \sim -20.2 \pm 4.8$ kcal/mol, respectively. As the pKa of HIP212 at the δ nitrogen position estimated by H++ server is not far from 7.0, and the calculated ΔG_{bind} for HIP6_HIE212 is too

negative for a typical divalent metal ion binding free energy, we argue that the *apo* protein should be HIP6_HIP212 dominated.

Table 19. Summary of the overall averaged ΔG (in kcal/mol) results considering two possible *apo* states.

	GlxI-Co ²⁺		GlxI-Ni ²⁺	
	12-6-4	m12-6-4	12-6-4	m12-6-4
ΔG_A	31.5 ± 3.6	39.0 ± 2.6	29.5 ± 4.0	40.9 ± 4.1
ΔG_B	-16.6 ± 1.7			
ΔG_C	-16.4 ± 1.6			
$-\Delta G_{\text{bind}}$ (Apo state: HIP6_HIP212)	-1.5 ± 4.3	6.0 ± 3.5	-3.5 ± 4.6	7.9 ± 4.7
ΔG_D (See Figure 29)	12.3 ± 0.9			
$-\Delta G_{\text{bind}}$ (Apo state: HIP6_HIE212)	10.8 ± 4.4	18.3 ± 3.6	8.8 ± 4.7	20.2 ± 4.8
$-\Delta G_{\text{bind}}^{\text{exp}}$	> 9.6		> 9.6	

4.4 Conclusions

In this work, we have shown that the 12-6-4 nonbonded model could be extended from modeling metal ions in water and with small ligands to modeling TMs in proteins. We have presented a computational route to determine TMs binding affinities in metalloprotein (Co²⁺ and Ni²⁺ in the GlxI enzyme) by MD based free energy simulations: the double-decoupling method (DDM) with distance/angle/dihedral restraints. Optimization of the 12-6-4 potential between the TMs and the binding site residues is critical to derive accurate TMs binding free energies in the protein. Furthermore, we find that the consideration of protonation state changes of the binding site residues associated with (un)binding is crucial, and the corresponding free energy changes are important contributions to the computed binding free energies.

This work shows, for the first time, that it is possible to create a thermodynamically balanced model that can then be used to estimate absolute binding free energies of coordination transition metal ions. The present model provides an accurate approach representing the true solvation free

energies of the ions along with accurate representation of the interactions between the ion and the host. With this model in hand we now have a framework that will allow us to tackle a range of problems associated with the coordination chemistry of transition metal ions as well as even more highly charge ions.

APPENDICES

APPENDIX: Tables

Table 20. The $\Delta\Delta G$ values directly obtained from the TI calculations as well as the cycle-closure $\Delta\Delta G$ values as mentioned in section 3.2.1.

System	Ligand 1	Ligand 2	$\Delta\Delta G$ (kcal/mol)						
			Experiment	Forward	Reverse	Average	Error	Cycle-closure $\Delta\Delta G$	Error
Thrombin	1d	1c	-0.31	-0.3	-0.1	-0.2	-0.11	-0.34	0.03
	3a	1b	-0.14	-0.1	-0.7	-0.4	0.26	-0.39	0.25
	3a	1d	0.07	-0.1	0.1	0	0.07	0.01	0.06
	1b	1c	-0.10	0	-0.2	-0.1	0	0.04	-0.14
	1d	1a	0.77	1.1	0.7	0.9	-0.13	1.05	-0.28
	1d	7a	0.03	0.7	-0.4	0.15	-0.12	0.29	-0.26
	1b	1a	0.98	0.8	1.2	1	-0.02	1.43	-0.45
	1b	7a	0.24	2.6	-1	0.8	-0.56	0.66	-0.42
	1a	5	-0.10	1.1	0.9	1	-1.1	1.2	-1.3
	1a	3b	-0.38	-0.4	-2.1	-1.25	0.87	-0.87	0.49
	1b	3b	0.60	2.1	-0.2	0.95	-0.35	0.57	0.03
	1d	6e	-0.66	-0.3	-0.4	-0.35	-0.31	-0.31	-0.35
	1d	5	0.67	2.3	2.6	2.45	-1.78	2.25	-1.58
	6a	1b	0.72	0.1	0.2	0.15	0.57	0.19	0.53
	6a	6b	0.29	0.8	1.1	0.95	-0.66	0.91	-0.62
6e	6b	0.02	1.1	0.1	0.6	-0.58	0.64	-0.62	
TYK2	ejm_31	ejm_46	-1.77	-0.5	-1	-0.75	-1.02	-0.74	-1.03
	ejm_31	ejm_43	1.28	0.8	0.9	0.85	0.43	0.84	0.44
	ejm_31	ejm_45	-0.02	-0.5	-0.3	-0.4	0.38	-0.41	0.39
	ejm_31	jmc_28	-1.44	0.5	0.6	0.55	-1.99	0.69	-2.13

Table 20 (cont'd)

TYK2	ejm_31	ejm_48	0.54	-0.2	-0.3	-0.25	0.79	-0.28	0.82
	ejm_42	ejm_48	0.78	-0.3	0	-0.15	0.93	-0.12	0.9
	ejm_42	ejm_55	0.57	-0.9	-0.8	-0.85	1.42	-0.61	1.18
	ejm_42	ejm_54	-0.75	-2.5	-2.6	-2.55	1.8	-3.18	2.43
	ejm_43	ejm_55	-0.95	-0.9	-2.3	-1.6	0.65	-1.61	0.66
	ejm_44	ejm_55	-1.79	-4.2	-3.8	-4	2.21	-3.51	1.72
	ejm_44	ejm_42	-2.36	-2.4	-2.4	-2.4	0.04	-2.89	0.53
	ejm_45	ejm_42	-0.22	0.1	0.4	0.25	-0.47	0.24	-0.46
	ejm_47	ejm_31	0.16	0	-0.1	-0.05	0.21	0.19	-0.03
	ejm_47	ejm_55	0.49	0.7	-1.4	-0.35	0.84	-0.59	1.08
	ejm_49	ejm_31	-1.79	0.4	0.1	0.25	-2.04	0.11	-1.9
	ejm_49	ejm_50	-1.23	0.3	-0.5	-0.1	-1.13	0.04	-1.27
	ejm_50	ejm_42	-0.80	-0.2	-0.3	-0.25	-0.55	-0.11	-0.69
	ejm_55	ejm_54	-1.32	-3.6	-2.8	-3.2	1.88	-2.57	1.25
	jmc_23	ejm_55	2.49	-0.3	0.2	-0.05	2.54	0.1	2.39
	jmc_23	ejm_46	0.39	0.1	0.2	0.15	0.24	0.14	0.25
	jmc_23	jmc_27	0.42	-0.2	0.4	0.1	0.32	0.16	0.26
	jmc_23	jmc_30	0.76	0.1	-0.2	-0.05	0.81	-0.24	1
	jmc_28	jmc_27	-0.30	-1.4	-1.3	-1.35	1.05	-1.41	1.11
	jmc_28	jmc_30	0.04	-2.3	-1.7	-2	2.04	-1.81	1.85
JNK1	17124-1	18634-1	-0.32	-0.1	1	0.45	-0.77	0.2	-0.52
	17124-1	18631-1	0.26	2	2.1	2.05	-1.79	2.3	-2.04
	18626-1	18624-1	0.38	1.1	1.9	1.5	-1.12	1.51	-1.13
	18626-1	18658-1	-0.83	0.3	-0.5	-0.1	-0.73	-0.69	-0.14
	18626-1	18625-1	0.77	4.1	3.3	3.7	-2.93	2.74	-1.97
	18626-1	18632-1	-0.21	-0.1	-1.3	-0.7	0.49	-0.62	0.41
	18626-1	18630-1	-0.27	-0.5	-0.2	-0.35	0.08	-0.23	-0.04
	18626-1	18627-1	0.39	-0.7	0.4	-0.15	0.54	-0.27	0.66
	18626-1	18634-1	-1.12	-1.8	-2.5	-2.15	1.03	-0.54	-0.58
	18626-1	18628-1	0.17	1.6	1.2	1.4	-1.23	1.28	-1.11
	18626-1	18660-1	0.17	-0.8	-3.9	-2.35	2.52	-2.12	2.29
	18626-1	18659-1	-0.59	2.3	-0.5	0.9	-1.49	0.66	-1.25
	18627-1	18630-1	-0.66	0	0.3	0.15	-0.81	0.03	-0.69
	18628-1	18624-1	0.21	0.8	-0.1	0.35	-0.14	0.23	-0.02
18629-1	18627-1	0.19	0.1	0.3	0.2	-0.01	0.2	-0.01	

Table 20 (cont'd)

JNK1	18631-1	18660-1	0.71	-2.9	-4	-3.45	4.16	-3.68	4.39
	18631-1	18624-1	0.92	-0.5	-1.6	-1.05	1.97	-0.05	0.97
	18631-1	18652-1	-1.27	-1.6	-0.6	-1.1	-0.17	-1.1	-0.17
	18632-1	18624-1	0.59	2.2	1.9	2.05	-1.46	2.13	-1.54
	18633-1	18624-1	0.68	1.9	1.4	1.65	-0.97	1.65	-0.97
	18634-1	18637-1	-0.15	-0.8	-0.2	-0.5	0.35	0.02	-0.17
	18635-1	18625-1	-0.82	1.7	1.3	1.5	-2.32	2.19	-3.01
	18635-1	18624-1	-1.21	1.9	1.4	1.65	-2.86	0.96	-2.17
	18636-1	18625-1	-0.59	0.2	0	0.1	-0.69	0.37	-0.96
	18636-1	18624-1	-0.98	-0.4	-0.8	-0.6	-0.38	-0.87	-0.11
	18637-1	18631-1	0.73	1.6	1.5	1.55	-0.82	2.07	-1.34
	18638-1	18658-1	0.39	-0.5	-0.3	-0.4	0.79	0.17	0.22
	18638-1	18634-1	0.1	1.3	0.5	0.9	-0.8	0.33	-0.23
	18639-1	18658-1	0.04	2.1	0.8	1.45	-1.41	1.47	-1.43
	18639-1	18634-1	-0.25	1.4	1.9	1.65	-1.9	1.63	-1.88
18659-1	18634-1	-0.53	0	-1.9	-0.95	0.42	-1.19	0.66	
CDK2	22	1h1r	0.19	-0.4	-0.7	-0.55	0.74	-0.43	0.62
	17	1h1q	-1.14	1.7	1.4	1.55	-2.69	1.39	-2.53
	17	21	-0.79	0.7	0.4	0.55	-1.34	0.59	-1.38
	17	22	-0.82	-0.1	0.5	0.2	-1.02	0.32	-1.14
	20	1h1q	0.54	1	0.9	0.95	-0.41	1.15	-0.61
	26	1h1q	0.25	0.4	0.5	0.45	-0.20	0.31	-0.06
	26	1oi9	-1.31	-3	-2	-2.5	1.19	-2.4	1.09
	28	26	2.68	0.7	1.5	1.1	1.58	1.35	1.33
	28	31	1.57	1.4	0.5	0.95	0.62	0.7	0.87
29	26	1.45	3.1	1.1	2.1	-0.65	2.17	-0.72	

Table 20 (cont'd)

CDK2	30	26	1.38	0.9	0.2	0.55	0.83	0.58	0.80
	30	31	0.27	-0.2	0.1	-0.05	0.32	-0.08	0.35
	31	32	-0.21	-0.2	-2.4	-1.3	1.09	-1.49	1.28
	1h1r	1oi9	-2.07	-0.9	-1.8	-1.35	-	-1.19	-0.88
	1h1r	21	-0.16	1.3	0.2	0.75	-	0.71	-0.87
	1h1s	1oiy	1.46	0.4	-0.4	0	1.46	0.43	1.03
	1h1s	26	2.82	2.3	2.4	2.35	0.47	1.92	0.90
	1oi9	20	1.02	2.4	0.3	1.35	-	1.55	-0.53
	1oiu	26	0.65	2	3.4	2.7	-	2.75	-2.10
	1oiu	1h1q	0.90	3.3	2.9	3.1	-	3.05	-2.15
	1oiy	1oi9	0.04	-0.9	-0.8	-0.85	0.89	-0.91	0.95
	1oiy	32	0.04	0.2	-1.9	-0.85	0.89	-0.66	0.70
	1oiy	29	-0.10	-1.4	-0.1	-0.75	0.65	-0.68	0.58
PTP1B	23466	23475	-0.87	-0.6	-2.4	-1.5	0.63	-1.92	1.05
	23467	23466	-0.51	0.5	-0.1	0.2	-	0.18	-0.69
	23467	23468	-0.41	-0.1	0.7	0.3	-	0.28	-0.69
	23467	23469	-0.38	-0.4	3	1.3	-	1.80	-2.18
	23467	23470	-0.38	-0.1	0	-0.05	-	-0.17	-0.21
	23467	23473	-1.05	-1.1	-1.8	-1.45	0.40	-1.64	0.59
	23467	23474	-1.77	-3.1	-2.6	-2.85	1.08	-2.76	0.99
	23467	23475	-1.38	-1.8	-2.5	-2.15	0.77	-1.73	0.35
	23467	23476	-2.07	-2.4	-2.1	-2.25	0.18	-2.36	0.29
	23469	23472	-0.92	-2.7	-2.9	-2.8	1.88	-2.42	1.5
	23469	20669(2qbr)	-0.88	-3.7	-1.8	-2.75	1.87	-2.63	1.75
	23471	23466	-0.1	0.1	0.1	0.1	-0.2	-0.03	-0.07
	23471	23468	0	0	0.1	0.05	-	0.07	-0.07
	23471	23470	0.03	-0.4	-0.6	-0.5	0.53	-0.38	0.41
23473	20669(2qbr)	-0.22	1	1	1	-	0.81	-1.03	

Table 20 (cont'd)

PTP1B	23474	23466	1.26	2.9	2.8	2.85	- 1.59	2.94	-1.68
	23476	23466	1.57	2.1	3.2	2.65	- 1.08	2.54	-0.97
	23477	23466	1.01	2.4	1.6	2	- 0.99	1.84	-0.83
	23477	23467	1.51	1.1	1.1	1.1	0.41	1.66	-0.15
	23477	23479	-0.29	-0.5	0.5	0	- 0.29	-0.05	-0.24
	23477	23482	-1.15	0	-1.1	-0.55	- 0.60	-0.58	-0.57
	23477	23483	-1.02	0.6	-1.3	-0.35	- 0.67	-0.50	-0.52
	23477	23330(2qbp)	-1.27	-1.8	-2.6	-2.2	0.93	-2.20	0.93
	23480	23479	-0.42	0.3	0	0.15	- 0.57	0.04	-0.46
	23480	23482	-1.29	-0.8	-0.4	-0.6	- 0.69	-0.49	-0.8
	23482	23479	0.86	1.2	-0.2	0.5	0.36	0.53	0.33
	23482	23485	-1.01	0.3	1.9	1.1	- 2.11	1.41	-2.42
	23482	23486	-2.46	-0.9	-1.6	-1.25	- 1.21	-1.08	-1.38
	23483	23479	0.73	1.1	-0.5	0.3	0.43	0.45	0.28
	23484	23479	-1.39	5.5	-1.4	2.05	- 3.44	0.34	-1.73
	23484	23482	-2.25	-1.2	-1.6	-1.4	- 0.85	-0.18	-2.07
	23484	23485	-3.26	1	0.9	0.95	- 4.21	1.23	-4.49
	23484	23486	-4.72	-0.8	-1.1	-0.95	- 3.77	-1.27	-3.45
	23485	23479	1.87	-3.3	-2.3	-2.8	4.67	-0.88	2.75
	23486	23479	3.33	2.6	1.5	2.05	1.28	1.61	1.72
	23486	23485	1.46	2.1	2.1	2.1	- 0.64	2.49	-1.03
	20667(2qbp)	23479	2.28	0.7	2.6	1.65	0.63	1.56	0.72
	20667(2qbp)	23482	1.42	1.4	1.5	1.45	- 0.03	1.03	0.39
	20667(2qbp)	23484	3.67	3.6	-0.1	1.75	1.92	1.22	2.45
	20667(2qbp)	23485	0.41	1.2	1.8	1.5	- 1.09	2.44	-2.03

Table 20 (cont'd)

PTP1B	20667(2qbp)	23486	-1.05	0	-0.3	-0.15	-0.9	-0.05	-1
	20669(2qbr)	23466	0.76	0.6	0.8	0.7	0.06	1.01	-0.25
	20669(2qbr)	23472	-0.04	-0.7	1.9	0.6	-0.64	0.22	-0.26
	20670(2qbs)	23466	1.24	2.8	2.8	2.8	-1.56	2.4	-1.16
	20670(2qbs)	23477	0.23	0.2	0.6	0.4	-0.17	0.56	-0.33
	20670(2qbs)	23479	-0.06	0.1	0.3	0.2	-0.26	0.51	-0.57
	20670(2qbs)	23482	-0.92	0.5	0.2	0.35	-1.27	-0.02	-0.9
	20670(2qbs)	23483	-0.79	0	-0.5	-0.25	-0.54	0.05	-0.84
	20670(2qbs)	23330(2qbq)	-1.04	-1.4	-1.9	-1.65	0.61	-1.65	0.61
BACE	CAT-13a	CAT-17g	-0.9	2.1	1.8	1.95	-2.85	2.25	-3.15
	CAT-13a	CAT-17i	-0.63	1.6	1	1.3	-1.93	1	-1.63
	CAT-13a	CAT-13m	0.08	-0.7	-0.9	-0.8	0.88	-2.65	2.73
	CAT-13b	CAT-17g	-0.62	-0.3	-2.4	-1.35	0.73	0.1	-0.72
	CAT-13c	CAT-17i	-0.15	0.2	-2.6	-1.2	1.05	-1.83	1.68
	CAT-13d	CAT-13h	0.84	3.8	3.8	3.8	-2.96	3.02	-2.18
	CAT-13d	CAT-17h	0.14	0.8	0.4	0.6	-0.46	-0.22	0.36
	CAT-13d	CAT-17d	1.05	4.3	4.1	4.2	-3.15	3.3	-2.25
	CAT-13d	CAT-13b	1.35	2.2	-1.1	0.55	0.8	2	-0.65
	CAT-13d	CAT-13f	1.38	2.8	-1	0.9	0.48	2.22	-0.84
	CAT-13d	CAT-17a	-0.26	-0.3	-0.7	-0.5	0.24	-0.53	0.27
	CAT-13d	CAT-13i	1.2	0	2.3	1.15	0.05	1.02	0.18
	CAT-13e	CAT-17g	0.22	-0.7	-1.7	-1.2	1.42	-0.17	0.39
	CAT-13e	CAT-17i	0.49	0.4	-1.2	-0.4	0.89	-1.43	1.92
	CAT-13g	CAT-17g	-0.65	1.9	1.4	1.65	-2.3	-0.25	-0.4
	CAT-13g	CAT-17i	-0.38	-4.4	-2.4	-3.4	3.02	-1.5	1.12
	CAT-13h	CAT-17i	0.16	-4.6	1.8	-1.4	1.56	-2.18	2.34
CAT-13j	CAT-4o	-0.65	-1.7	0.8	-0.45	-0.2	-0.81	0.16	
CAT-13k	CAT-4d	0.59	0.8	-0.1	0.35	0.24	0.57	0.02	

Table 20 (cont'd)

BACE	CAT-13k	CAT-4b	0.07	0	0	0	0.07	-0.28	0.35
	CAT-13n	CAT-13k	-1.16	-1.8	-1.1	-1.45	0.29	-2.46	1.3
	CAT-13n	CAT-13a	-0.3	-0.8	0	-0.4	0.1	-2.25	1.95
	CAT-13n	CAT-4i	0.28	-2.9	-5.7	-4.3	4.58	-1.45	1.73
	CAT-13o	CAT-17i	-0.93	-1	-0.7	-0.85	-0.08	-1.67	0.74
	CAT-13o	CAT-17h	-1.79	-3.1	-4	-3.55	1.76	-2.73	0.94
	CAT-17b	CAT-13d	-0.45	0.2	0	0.1	-0.55	0.21	-0.66
	CAT-17b	CAT-17e	0	-0.1	0.1	0	0	-0.11	0.11
	CAT-17c	CAT-17e	-0.16	-1.2	-1.5	-1.35	1.19	-1.14	0.98
	CAT-17f	CAT-17e	-0.6	-2.4	-1.8	-2.1	1.5	-1.84	1.24
	CAT-17g	CAT-17c	-0.12	-1.8	-1.2	-1.5	1.38	-1.29	1.17
	CAT-17g	CAT-17f	0.32	-1.2	-0.5	-0.85	1.17	-0.59	0.91
	CAT-17g	CAT-13i	0.47	-0.8	-1.6	-1.2	1.67	-1.07	1.54
	CAT-17g	CAT-13c	0.42	2.8	-0.4	1.2	-0.78	0.57	-0.15
	CAT-17g	CAT-17d	0.32	0.4	0.2	0.3	0.02	1.2	-0.88
	CAT-17i	CAT-13f	0.38	2.3	3.1	2.7	-2.32	1.38	-1
	CAT-17i	CAT-17a	-1.26	-1.6	-1.2	-1.4	0.14	-1.37	0.11
	CAT-24	CAT-17e	1.33	3.4	1.9	2.65	-1.32	2.29	-0.96
	CAT-24	CAT-17i	1.88	3.9	2.3	3.1	-1.22	3.46	-1.58
	CAT-4a	CAT-4o	-1.45	-0.4	0.6	0.1	-1.55	0.52	-1.97
	CAT-4a	CAT-13k	-1.77	-1.5	-1.2	-1.35	-0.42	-1.77	0
	CAT-4c	CAT-4o	-1.53	0.9	1.4	1.15	-2.68	0.89	-2.42
	CAT-4i	CAT-13m	-0.5	-5.9	-6.7	-6.3	5.8	-3.45	2.95
	CAT-4j	CAT-4o	-0.36	-0.6	-0.5	-0.55	0.19	-0.64	0.28
	CAT-4k	CAT-4o	-1.53	-1.8	-1	-1.4	-0.13	-1.04	-0.49
	CAT-4l	CAT-13k	-0.36	0.4	-2	-0.8	0.44	-1.93	1.57
CAT-4m	CAT-4c	1.3	0	1.8	0.9	0.4	0.64	0.66	
CAT-4m	CAT-13j	0.42	2.1	3.3	2.7	-2.28	2.34	-1.92	

Table 20 (cont'd)

BACE	CAT-4m	CAT-4j	0.13	2.1	2.4	2.25	-2.12	2.16	-2.03
	CAT-4m	CAT-4n	0.06	1.1	0.6	0.85	-0.79	0.34	-0.28
	CAT-4m	CAT-13k	-0.55	-3.1	-3.2	-3.15	2.6	-0.77	0.22
	CAT-4m	CAT-13m	0.39	-2.5	-1.9	-2.2	2.59	-3.21	3.6
	CAT-4m	CAT-4l	-0.19	3.5	1.1	2.3	-2.49	1.17	-1.36
	CAT-4m	CAT-4k	1.3	2.7	1.7	2.2	-0.9	2.56	-1.26
	CAT-4m	CAT-4p	-0.93	-0.7	-0.4	-0.55	-0.38	0.07	-1
	CAT-4n	CAT-13k	-0.61	-0.8	-0.4	-0.6	-0.01	-1.11	0.5
	CAT-4o	CAT-4b	-0.25	-2.9	-2.8	-2.85	2.6	-2.57	2.32
	CAT-4o	CAT-4d	0.27	-2.1	-0.9	-1.5	1.77	-1.72	1.99
	CAT-4p	CAT-13k	0.38	-1.2	-1.7	-1.45	1.83	-0.83	1.21
MCL1	26	44	-0.44	1	0.3	0.65	-1.09	0.37	-0.81
	26	57	-0.8	-1	-0.7	-0.85	0.05	-0.65	-0.15
	26	64	-1.26	-3.9	-3.8	-3.85	2.59	-3.87	2.61
	27	23	-2.71	-0.9	-1.9	-1.4	-1.31	-2.1	-0.61
	27	45	-2.84	-4	-2.6	-3.3	0.46	-2.6	-0.24
	27	46	-1.48	-1.1	-1.2	-1.15	-0.33	-0.96	-0.52
	28	27	0.51	-3	0	-1.5	2.01	-1.13	1.64
	28	35	-2.19	-0.4	-3.4	-1.9	-0.29	-2.18	-0.01
	28	47	0.85	1.2	-1.7	-0.25	1.1	-0.34	1.19
	29	27	0.82	-1.2	4.3	1.55	-0.73	-0.05	0.87
	29	35	-1.87	-1.8	-0.3	-1.05	-0.82	-1.1	-0.77
	29	40	-0.31	-5.1	-3.1	-4.1	3.79	-2.45	2.14
	30	27	1.74	0.5	0	0.25	1.49	0.79	0.95
	30	35	-0.96	-0.9	-0.8	-0.85	-0.11	-0.26	-0.7
	30	40	0.6	1.2	-1.1	0.05	0.55	-1.6	2.2
	30	48	1.19	1.1	2.1	1.6	-0.41	2.12	-0.93
	31	35	-0.89	-1.1	0.4	-0.35	-0.54	0.74	-1.63
	32	34	-0.29	0.2	-0.3	-0.05	-0.24	-0.29	0
	32	46	-1.02	-0.7	-1.4	-1.05	0.03	-1.24	0.22
33	27	0.76	2.1	2	2.05	-1.29	2.02	-1.26	
35	33	1.94	-1.2	-0.7	-0.95	2.89	-0.98	2.92	
35	34	1.94	1.2	0.4	0.8	1.14	1.04	0.9	

Table 20 (cont'd)

MCL1	35	36	0.63	-1.6	-1.2	-1.4	2.03	-0.79	1.42
	35	37	-0.14	-0.1	-1.8	-0.95	0.81	-0.76	0.62
	35	39	1.79	-0.5	1.1	0.3	1.49	0.54	1.25
	35	53	-1.15	-1.3	-6.2	-3.75	2.6	-4.14	2.99
	35	60	-0.1	-3.7	-3	-3.35	3.25	-3.24	3.14
	38	35	-1.79	4.6	0.5	2.55	-4.34	2.94	-4.73
	38	60	-1.9	-0.5	0.7	0.1	-2	-0.29	-1.61
	39	32	0.44	0.8	0.3	0.55	-0.11	0.79	-0.35
	41	32	0.55	3.1	2.9	3	-2.45	3.22	-2.67
	41	35	-1.68	2.1	2.1	2.1	-3.78	1.88	-3.56
	42	51	0.45	0.5	1.1	0.8	-0.35	0.29	0.16
	42	64	-0.6	-3.8	-3.9	-3.85	3.25	-3.83	3.23
	43	27	0.92	2.9	1.2	2.05	-1.13	1.96	-1.04
	43	47	1.26	3.1	2.2	2.65	-1.39	2.74	-1.48
	44	23	-0.16	-1.1	-0.9	-1	0.84	-1.28	1.12
	48	27	0.55	-1.5	-2.2	-1.85	2.4	-1.33	1.88
	49	35	-0.45	1.4	1.1	1.25	-1.7	0.96	-1.41
	49	67	0.78	-2.3	-1	-1.65	2.43	-1.36	2.14
	50	60	0.41	-1.7	-2	-1.85	2.26	-2.08	2.49
	51	45	-0.51	-0.9	-1.4	-1.15	0.64	-1.66	1.15
	52	60	0.31	-1.4	-1.4	-1.4	1.71	-1.68	1.99
	54	23	0.95	1.5	1.7	1.6	-0.65	1.82	-0.87
	54	42	0.88	3	2.8	2.9	-2.02	2.68	-1.8
	56	35	0.45	2	3.5	2.75	-2.3	2.79	-2.34
	56	60	0.34	-0.4	-0.4	-0.4	0.74	-0.44	0.78
	57	23	0.21	-0.6	-0.3	-0.45	0.66	-0.25	0.46
	58	60	0.49	2.3	2.9	2.6	-2.11	2.57	-2.08
	60	36	0.74	3.6	2.5	3.05	-2.31	2.44	-1.7
	61	60	-0.84	-1.6	-1	-1.3	0.46	-0.55	-0.29
	62	26	-0.28	0.3	0.5	0.4	-0.68	0.3	-0.58
62	45	-1	-0.8	-1.6	-1.2	0.2	-1.1	0.1	
63	60	0.15	2	3	2.5	-2.35	2.2	-2.05	
65	60	-0.51	1.3	1.3	1.3	-1.81	1.12	-1.63	
65	67	0.83	1.9	1.8	1.85	-1.02	2.03	-1.2	

Table 20 (cont'd)

MCL1	66	23	-0.4	1.4	1.4	1.4	-1.8	1.67	-2.07
	66	42	-0.47	2.8	2.8	2.8	-3.27	2.53	-3
	67	27	1.46	4.2	1.6	2.9	-1.44	3.38	-1.92
	67	31	-0.34	-2.1	3.1	0.5	-0.84	1.59	-1.93
	67	32	1	4.5	4.6	4.55	-3.55	3.66	-2.66
	67	35	-1.23	2.6	2.7	2.65	-3.88	2.33	-3.56
	67	37	-1.37	1.9	1.6	1.75	-3.12	1.56	-2.93
	67	50	-1.75	1.8	1	1.4	-3.15	1.17	-2.92
	67	52	-1.64	0.8	1.3	1.05	-2.69	0.77	-2.41
	67	53	-2.38	-2.2	-2.2	-2.2	-0.18	-1.81	-0.57
	67	58	-1.83	-3.4	-3.5	-3.45	1.62	-3.48	1.65
	67	61	-0.5	0.8	-3	-1.1	0.6	-0.35	-0.15
	67	63	-1.48	-2.9	-2.7	-2.8	1.32	-3.1	1.62
	68	23	-1.14	-0.8	-0.8	-0.8	-0.34	-0.52	-0.62
	68	45	-1.26	-0.1	-1.4	-0.75	-0.51	-1.03	-0.23
P38	p38a_2aa	p38a_2bb	0.2	-2.7	0	-1.35	1.55	-1.2	1.4
	p38a_2aa	p38a_3flw	-1.41	-3.3	-0.2	-1.75	0.34	-1.85	0.44
	p38a_2e	p38a_2j	0.62	-4.3	-3.4	-3.85	4.47	-3.09	3.71
	p38a_2ee	p38a_2j	2.18	6.2	-0.6	2.8	-0.62	2.51	-0.33
	p38a_2ee	p38a_3fln	1.38	-1.1	7.1	3	-1.62	3.29	-1.91
	p38a_2g	p38a_2c	0.2	-1.3	-0.6	-0.95	1.15	-0.75	0.95
	p38a_2g	p38a_2f	2.18	1.6	1.6	1.6	0.58	2.03	0.15
	p38a_2g	p38a_2h	1.18	4.7	3.1	3.9	-2.72	2.08	-0.9
	p38a_2g	p38a_2i	0.61	-2	-1.6	-1.8	2.41	-1.37	1.98
	p38a_2gg	p38a_2j	0.58	-4.4	-3.2	-3.8	4.38	-3.67	4.25
	p38a_2j	p38a_2f	1.6	2	2.6	2.3	-0.7	1.87	-0.27
	p38a_2j	p38a_2ff	-1.36	0.3	2.2	1.25	-2.61	1.32	-2.68
	p38a_2j	p38a_2h	0.6	-0.6	0.8	0.1	0.5	1.92	-1.32
	p38a_2j	p38a_2q	-2.18	-7.8	0	-3.9	1.72	-2.44	0.26
	p38a_2j	p38a_2r	-0.71	2.2	2.6	2.4	-3.11	1.49	-2.2
	p38a_2l	p38a_2j	2.18	0	-0.5	-0.25	2.43	1.02	1.16
	p38a_2l	p38a_2n	0.41	0.2	0.3	0.25	0.16	0.35	0.06
p38a_2l	p38a_2o	1.77	3.5	1.1	2.3	-0.53	1.63	0.14	

Table 20 (cont'd)

P38	p38a_2l	p38a_2p	1.06	1.1	1.2	1.15	- 0.09	0.68	0.38
	p38a_2m	p38a_2j	0.88	0.7	0.7	0.7	0.18	1.26	-0.38
	p38a_2m	p38a_2k	0.41	3	4	3.5	- 3.09	2.94	-2.53
	p38a_2s	p38a_2l	-1.15	-2.8	-2.6	-2.7	1.55	-2.48	1.33
	p38a_2t	p38a_2j	1.77	-1.4	-2.2	-1.8	3.57	-0.54	2.31
	p38a_2u	p38a_2k	1.71	1.6	3.3	2.45	- 0.74	3.91	-2.2
	p38a_2u	p38a_2q	0	1.1	1.4	1.25	- 1.25	-0.21	0.21
	p38a_2v	p38a_2bb	-0.09	-1	0.1	-0.45	0.36	-0.6	0.51
	p38a_2v	p38a_2j	-1.11	3.6	-1	1.3	- 2.41	-0.39	-0.72
	p38a_2v	p38a_2x	-1.26	-0.4	-1.4	-0.9	- 0.36	-2.12	0.86
	p38a_2v	p38a_2y	-0.8	-2.1	-1.4	-1.75	0.95	-0.73	-0.07
	p38a_2v	p38a_3fln	-1.91	-1.7	-0.9	-1.3	- 0.61	0.39	-2.3
	p38a_2v	p38a_3fly	-2.45	-3.7	-2.2	-2.95	0.5	-2.73	0.28
	p38a_2v	p38a_3fmh	-1.85	-7.2	-5.5	-6.35	4.5	-4.54	2.69
	p38a_2v	p38a_3fmk	-2.86	-1.9	1.3	-0.3	- 2.56	-1.98	-0.88
	p38a_2z	p38a_2aa	1.09	1.4	0.8	1.1	- 0.01	1.15	-0.06
	p38a_2z	p38a_2y	0.58	1	0.7	0.85	- 0.27	-0.17	0.75
	p38a_2z	p38a_3flq	0.43	-3.4	1.1	-1.15	1.58	-1.96	2.39
	p38a_2z	p38a_3flw	-0.32	-1.9	0.3	-0.8	0.48	-0.7	0.38
	p38a_2z	p38a_3fmk	-1.47	-4.2	-2	-3.1	1.63	-1.42	-0.05
	p38a_3fln	p38a_2e	0.18	1.5	1.6	1.55	- 1.37	2.31	-2.13
	p38a_3fln	p38a_2ff	-0.56	1.5	-0.3	0.6	- 1.16	0.53	-1.09
	p38a_3fln	p38a_2g	0.22	-0.9	-0.1	-0.5	0.72	-0.95	1.17
	p38a_3fln	p38a_2gg	0.22	2.6	2.9	2.75	- 2.53	2.88	-2.66
p38a_3fln	p38a_2i	0.83	-1.4	-2.4	-1.9	2.73	-2.33	3.16	
p38a_3fln	p38a_2k	0.33	1.5	2.1	1.8	- 1.47	0.9	-0.57	

Table 20 (cont'd)

P38	p38a_3fln	p38a_2n	-0.97	-1.6	-1.1	-1.35	0.38	-1.45	0.48
	p38a_3fln	p38a_2o	0.39	-1.8	0.1	-0.85	1.24	-0.18	0.57
	p38a_3fln	p38a_2p	-0.32	-3	-0.2	-1.6	1.28	-1.13	0.81
	p38a_3fln	p38a_2r	0.09	0	-0.4	-0.2	0.29	0.71	-0.62
	p38a_3fln	p38a_2s	-0.23	1.3	-0.4	0.45	0.68	0.67	-0.9
	p38a_3fln	p38a_2t	-0.97	-1.4	-1.6	-1.5	0.53	-0.24	-0.73
	p38a_3fln	p38a_3flz	1.39	0.9	1.6	1.25	0.14	0.75	0.64
	p38a_3fly	p38a_2x	1.19	-0.8	-0.4	-0.6	1.79	0.62	0.57
	p38a_3fly	p38a_3flq	1.49	-0.1	-1.1	-0.6	2.09	0.21	1.28
	p38a_3fly	p38a_3fmh	0.6	-0.1	0.1	0	0.6	-1.81	2.41
	p38a_3flz	p38a_2c	-0.97	-1.1	-3.4	-2.25	1.28	-2.45	1.48
	p38a_3flz	p38a_2g	-1.17	-1.6	-1.2	-1.4	0.23	-1.7	0.53

Table 21. $\Delta\Delta G$ results for the “problematic cases” obtained from TI calculations using the protocol mentioned in section 3.2.3.

System	Ligand 1	Ligand 2	$\Delta\Delta G$ (kcal/mol)	Error (kcal/mol)
MCL1	27	23	-2.30	-0.41
	67	27	4.70	-3.24
	30	48	3.30	-2.11
	32	46	-0.05	-0.97
	32	34	0.10	-0.39
	28	47	1.50	-0.65
	43	47	2.50	-1.24
	49	67	-1.45	2.23
	50	60	-2.20	2.61
	67	53	-2.65	0.27
	61	60	-0.65	-0.19
	67	31	2.85	-3.19
	67	37	1.70	-3.07
	35	34	-0.20	2.14
	35	37	-2.25	2.11
38	35	2.55	-4.34	

Table 21 (cont'd)

MCL1	35	39	0.85	0.64
	41	35	1.35	-3.03
	49	35	0.85	-1.30
	56	35	4.60	-4.15
	67	35	2.65	-3.88
	57	23	-0.75	0.96
	28	35	-1.50	-0.69
	31	35	-0.50	-0.39
	39	32	0.85	-0.41
	27	46	-0.75	-0.73
	60	36	2.65	-1.91
	35	33	-0.75	2.69
	35	53	-6.90	5.75
	30	35	-0.45	-0.51
	35	60	-3.35	3.25
	35	36	-0.50	1.13
	29	35	-2.70	0.83
TYK2	49	31	0.00	-1.79
	28	30	-1.90	1.94
JNK1	31	24	2.40	-1.48
	35	24	2.20	-3.41
	36	24	-0.25	-0.73
	26	32	-0.20	-0.01
	26	24	1.45	-1.07
	26	27	0.20	0.19
	26	30	0.60	-0.87
BACE	3G	7I	-0.10	-0.28
	7G	7F	-0.75	1.07
	7G	3C	2.55	-2.13
P38	3FN	2GG	0.45	-0.23
	2V	2X	-2.15	0.89
	2M	2K	3.25	-2.84
	2G	2F	0.95	1.23
	2G	2H	3.65	-2.47

Table 21 (cont'd)

P38	3FN	2G	-0.05	0.27
	2Z	3FK	-2.40	0.93
	2EE	3FN	-0.55	1.93
	3FN	2K	1.15	-0.82
	2GG	2J	-3.80	4.38
	2L	2J	-0.95	3.13
	2T	2J	-1.10	2.87
	3FN	2R	0.45	-0.36
	2V	3FY	-2.60	0.15
	2J	2Q	-1.85	-0.33
	2J	2R	2.00	-2.71
	3FY	2X	0.20	0.99
	2Z	2AA	2.15	-1.06
	3FZ	2C	-2.40	1.43
	3FN	3FZ	0.85	0.54
	L2E	2J	-4.25	4.87
	2Z	2Y	0.65	-0.07
	2V	2Y	-0.65	-0.15
	2V	3FN	-4.35	2.44
	2V	3FK	-6.75	3.89
	3FN	2FF	-0.20	-0.36
	3FN	2O	0.75	-0.36
	3FN	2T	-1.35	0.38
	3FN	2I	-2.40	3.23
	2EE	2J	0.25	1.93
	2M	2J	-0.10	0.98
2S	2L	-1.75	0.60	
3FN	2S	-0.90	0.67	

Table 22. $\Delta\Delta G$ results for the “JNK1” system using the three-step protocol mentioned in section 3.2.4

JNK1	Ligand1	Ligand2	$\Delta\Delta G$ (kcal/mol)					
			Forward	Reverse	Average	Error	Cycle-closure $\Delta\Delta G$	Error
1	17124-1	18634-1	0.9	0.8	0.85	-1.17	0.58	-0.9
2	17124-1	18631-1	2.1	1.8	1.95	-1.69	2.22	-1.96
3	18626-1	18624-1	2.4	2	2.2	-1.82	2.16	-1.78
4	18626-1	18658-1	-1.6	-2.4	-2	1.17	-1.92	1.09
5	18626-1	18625-1	2.7	2	2.35	-1.58	2.52	-1.75
6	18626-1	18632-1	0.9	0.9	0.9	-1.11	0.55	-0.76
7	18626-1	18630-1	0.4	0.6	0.5	-0.77	0.55	-0.82
8	18626-1	18627-1	0.3	0.8	0.55	-0.16	0.5	-0.11
9	18626-1	18634-1	-1.2	-2.6	-1.9	0.78	-1.29	0.17
10	18626-1	18628-1	2	2.1	2.05	-1.88	2.08	-1.91
11	18626-1	18660-1	-1.7	-3.5	-2.6	2.77	-2.82	2.99
12	18626-1	18659-1	0.1	-1.8	-0.85	0.26	-1.12	0.53
13	18627-1	18630-1	0.1	0.1	0.1	-0.76	0.05	-0.71
14	18628-1	18624-1	0.3	-0.2	0.05	0.16	0.08	0.13
15	18629-1	18627-1	0.2	0	0.1	0.09	0.1	0.09
16	18631-1	18660-1	-4.4	-2.4	-3.4	4.11	-3.18	3.89
17	18631-1	18624-1	1.8	1.4	1.6	-0.68	1.8	-0.88
18	18631-1	18652-1	-1.9	-2.6	-2.25	0.98	-2.25	0.98
19	18632-1	18624-1	2.1	1.8	1.95	-1.36	1.6	-1.01
20	18633-1	18624-1	1.7	1.3	1.5	-0.82	1.5	-0.82
21	18634-1	18637-1	-0.5	0.4	-0.05	-0.1	0.1	-0.25
22	18635-1	18625-1	1.4	1.3	1.35	-2.17	1.45	-2.27
23	18635-1	18624-1	1.4	1	1.2	-2.41	1.1	-2.31
24	18636-1	18625-1	0.2	0.1	0.15	-0.74	-0.12	-0.47
25	18636-1	18624-1	-0.5	-1	-0.75	-0.23	-0.48	-0.5
26	18637-1	18631-1	1.1	1.7	1.4	-0.67	1.55	-0.82
27	18638-1	18658-1	0	-0.2	-0.1	0.49	0.01	0.38
28	18638-1	18634-1	1	0.5	0.75	-0.65	0.64	-0.54
29	18639-1	18658-1	1.1	1.8	1.45	-1.41	1.26	-1.22

Table 22 (cont'd)

30	18639-1	18634-1	1.5	1.9	1.7	-1.95	1.89	-2.14
31	18659-1	18634-1	-0.1	0.3	0.1	-0.63	-0.17	-0.36

Table 23. Summary of the 330 perturbations based on size, ring changes, etc.

Ligand1	Ligand2	$\Delta\Delta G$ error (kcal/mol)	Mutation		Size	Ring disappear ?	Ring type change
1b	1c	0	CL	BR	small change		
CAT-17b	CAT-17e	0	m- OCH3- Benzene	m- OCH3- Pyridine	big change		One Nring- Cring change
CAT-4n	CAT-13k	0.01	Benzene + CN- Pyridine	Pyridine + CL- Benzene	big change		Two Nring- Cring change
18629-1	18627-1	0.01	CH3 + H	H + CL	small change		
p38a_2z	p38a_2aa	0.01	CH3 + H	H + OH	small change		
CAT-17g	CAT-17d	0.02	F	H	small change		
1b	1a	0.02	CL	F	small change		
32	46	0.03	CH3- Benzene	Benzene- Cyclopen tane	big change	ring disappear	
20667(2q bp)	23482	0.03	Benzene	H	big change	Ring disappear	
ejm_44	ejm_42	0.04	C(CH3)3	CH2CH3	small change		
1oiy	1h1q	0.04	CONH2	H	big change		
26	57	0.05	O	NCH3	small change		Oring --> Nring
23471	23468	0.05	OCH3	CH2CH3	small change		
CAT-13d	CAT-13i	0.05	3Membe rRing	CH2- 3Membe rRing	small change		

Table 23 (cont'd)

CAT-13d	CAT-13i	0.05	3Member Ring	CH2-3Member Ring	small change		
20669(2q br)	23466	0.06	CH2-Benzene	H	big change	Ring disappear	
3a	1d	0.07	CH3	I	small change		
CAT-13k	CAT-4b	0.07	Pyridine + CL	Benzene + OCH3	big change		One Nring-Cring change
18626-1	18630-1	0.08	CL + H	H + CH3	small change		
CAT-13o	CAT-17i	0.08	SNring + CL-Benzene	3Member Ring + CH3-Pyridine	big change		Two Nring-Cring change
p38a_2l	p38a_2p	0.09	H	CH2SO2 CH3	big change		
CAT-13n	CAT-13a	0.1	2Nring	CH3	big change	ring disappear	
30	35	0.11	H	CL	small change		
1d	1c	0.11	I	BR	small change		
39	32	0.11	Benzene	CH3	big change	ring disappear	
1d	7a	0.12	I + H	CH3 + CH3	small change		
1d	1a	0.13	I	F	small change		
CAT-4k	CAT-4o	0.13	Pyridine	F-Pyridine	big change		
CAT-17i	CAT-17a	0.14	OCH3-Pyridine	CL-Benzene	big change		One Nring-Cring change
p38a_3fl n	p38a_3fl z	0.14	F + F	H + H	small change		
18628-1	18624-1	0.14	CH3	H	small change		
p38a_2l	p38a_2n	0.16	H	CH2OH	small change		

Table 23 (cont'd)

18631-1	18652-1	0.17	H + H + CH3	OCH3 + SO2CH3 + CH(CH3) 2	big change		
20670(2q bs)	23477	0.17	CH2	O	small change		
67	53	0.18	H + O	CL + NH	small change		Oring --> Nring
p38a_2m	p38a_2j	0.18	3Member Ring	CH3	big change	ring disappear	
23467	23476	0.18	OCH3	NH- 7Member Ring	big change	Ring disappear	
CAT-4j	CAT-4o	0.19	H	F	small change		
26	1h1q	0.20	OCH3	H	small change		
23471	23466	0.2	COOCH 3	H	big change		
CAT-13j	CAT-4o	0.2	CL- Benzene + Pyridine	F- Pyridine + Benzene	big change		Two Nring- Cring change
62	45	0.2	S + CL + Naphthal ene	NH + H + Benzene- Cyclohex ane	big change		Nring --> Sring
ejm_47	ejm_31	0.21	4Member Ring	CH3	big change	Ring disappear	
p38a_3fl z	p38a_2g	0.23	H	F	small change		
CAT-13k	CAT-4d	0.24	CL + Pyridine	OCH2C H3 + Benzene	big change		One Nring- Cring change
jmc_23	ejm_46	0.24	F	H	small change		
CAT-13d	CAT-17a	0.24	H	CL	small change		
32	34	0.24	CH3	CF3	big change		

Table 23 (cont'd)

3a	1b	0.26	CH3	CL	small change		
20670(2qbs)	23479	0.26	CH2	N-COCH3	big change		
p38a_2z	p38a_2y	0.27	C(CH3)2 CH2OH	CH2C(C H3)2OH	big change		
23477	23479	0.29	O	N-COCH3	big change		
28	35	0.29	H + H + CH3	CH3 + CL + H	big change		
p38a_3fl n	p38a_2r	0.29	H	C(CH3)2 OH	big change		
CAT-13n	CAT-13k	0.29	2Nring	Pyridine	big change		one ring change
1d	6e	0.31	I + H	CL + F	small change		
jmc_23	jmc_27	0.32	CL	F	small change		
30	31	0.32	SO2CH3	SOCH3	small change		
1oi9	20	0.33	OH + H	H + CH2OH	big change		
23467	23470	0.33	OCH3	NHCON HCH(CH 3)2	big change		
27	46	0.33	Benzene	Benzene- Cyclopentane	big change	ring disappear	
68	23	0.34	O + CL	S + H	small change		Oring --- > Sring
p38a_2aa	p38a_3fl w	0.34	OH + OH	CH2OH + CH2OH	small change		
18634-1	18637-1	0.35	H	NHCOC H3	big change		
42	51	0.35	H	CL	small change		
1b	3b	0.35	CL	CH2CH3	small change		
23482	23479	0.36	SO2CH3	COCH3	big change		
p38a_2v	p38a_2b b	0.36	H	CH2SO2 CH3	big change		

Table 23 (cont'd)

p38a_2v	p38a_2x	0.36	CH3	Oring	big change	ring disappear	
18636-1	18624-1	0.38	BR + BR	H + H	small change		
ejm_31	ejm_45	0.38	H	3Member Ring	big change	Ring disappear	
CAT-4m	CAT-4p	0.38	CH3	CL	small change		
p38a_3fl n	p38a_2n	0.38	H	CH2CH2 OH	big change		
CAT-4m	CAT-4c	0.4	CH3- Pyridine	OCH3- Benzene	big change		One Nring- Cring change
23467	23473	0.4	H	Benzene	big change	Ring disappear	
23477	23467	0.41	NH- 6Member Ring	OCH3	big change	Ring disappear	
30	48	0.41	CH3- Benzene	1H- Indole	big change	ring disappear	
20	1h1q	0.41	CH2OH	H	small change		
18659-1	18634-1	0.42	OCH3	H	small change		
CAT-4a	CAT-13k	0.42	Benzene + H	Pyridine + CL	big change		One Nring- Cring change
ejm_31	ejm_43	0.43	CH3	CH(CH3) 2	small change		
23483	23479	0.43	SO2- Benzene	COCH3	big change	Ring disappear	
CAT-4l	CAT-13k	0.44	2Nring + Benzene	CL- Benzene + Pyridine	big change		Two Nring- Cring change
27	45	0.46	Benzene	Benzene- Cyclohex ane	big change	ring disappear	
CAT-13d	CAT-17h	0.46	CL- Benzene	CL- Pyridine	big change		One Nring- Cring change

Table 23 (cont'd)

61	60	0.46	CL	H	small change		
1h1s	26	0.47	SO2NH2	OCH3	big change		
ejm_45	ejm_42	0.47	3MemberRing	CH3	big change	Ring disappear	
p38a_2z	p38a_3flw	0.48	CH3 + H + OH	H + CH2OH + CH2OH	big change		
CAT-13d	CAT-13f	0.48	3MemberRing	5MemberRing	big change		
18626-1	18632-1	0.49	CL	OCH3	small change		
p38a_2j	p38a_2h	0.5	CH2-Benzene-2F	S-Benzene-F	big change		
p38a_2v	p38a_3fl_y	0.5	CH3	CH(CH3) ₂	small change		
1oiy	31	0.50	CONH2	SOCH3	small change		
68	45	0.51	Benzene + CL + O	Cyclohexane + H + NH	big change		Oring --> Nring
p38a_2l	p38a_2o	0.53	H	SO2CH3	big change		
23471	23470	0.53	OCH3	NH-CH(CH3) ₂	big change		
p38a_3fl_n	p38a_2t	0.53	H	CN	small change	triple bond	
18626-1	18627-1	0.54	CL + H	H + CL	small change		
20670(2qbs)	23483	0.54	CH2	N-SO2-Benzene	big change	Ring disappear	
31	35	0.54	CF3 + H	CH3 + CL	big change		
ejm_50	ejm_42	0.55	CH2OH	CH2CH3	small change		
CAT-17b	CAT-13d	0.55	OCH3	CL	small change		
30	40	0.55	CH3 + H	H + O-Benzene	big change	ring disappear	

Table 23 (cont'd)

1b	7a	0.56	CL + H	CH ₃ + CH ₃	small change		
6a	1b	0.57	CL	H	small change		
23480	23479	0.57	OCH ₃	CH ₃	small change		
6e	6b	0.58	F	CH ₃	small change		
p38a_2g	p38a_2f	0.58	F + H	H + F	small change		
23477	23482	0.6	O	N-SO ₂ CH ₃	big change		
p38a_3fly	p38a_3fmh	0.6	H	tetrazole	big change	ring disappear	
67	61	0.6	O + H	S + CL	small change		Oring --> Sring
20670(2qbs)	23330(2qbq)	0.61	H+H+H+H	CH ₃ +CH ₃ +CH ₃ +CH ₃	big change		
p38a_2v	p38a_3fln	0.61	CH ₃	Oring	big change	ring disappear	
p38a_2ee	p38a_2j	0.62	O + N-COCH ₃	CH ₂ + O	big change		Nring --> Oring
28	31	0.62	SO ₂ NHC H ₃	SOCH ₃	big change		
20667(2qbp)	23479	0.63	SO ₂ -CH ₂ -Benzene	COCH ₃	big change	Ring disappear	
23466	23475	0.63	NH ₂	5Member Ring	big change	Ring disappear	
51	45	0.64	Benzene + CL + N	Cyclohexane + H + NH	big change		Sring --> Nring
20669(2qbr)	23472	0.64	CH ₂ -Benzene	SO ₂ -Benz-CL	big change		
23486	23485	0.64	CL	CH ₃	small change		
29	26	0.65	SO ₂ N(C H ₃) ₂	OCH ₃	big change		
ejm_43	ejm_55	0.65	CH(CH ₃) ₂	OCH ₃	small change		
54	23	0.65	CL + NH	H + S	small change		Nring --> Sring

Table 23 (cont'd)

loi9	29	0.65	CONH2	SO2N(C H3)2	big change		
57	23	0.66	NCH3	S	small change		Nring --> Sring
6a	6b	0.66	CL	CH3	small change		
23477	23483	0.67	O	NHSO2- Benzene	big change	Ring disappear	
62	26	0.68	CL + S	H + O	small change		Sring --> Oring
p38a_3fl n	p38a_2s	0.68	H	CH(OH) CH2OH	big change		
18636-1	18625-1	0.69	BR + BR	CL + H	small change		
23480	23482	0.69	COOCH 3	SO2CH3	big change		
p38a_2j	p38a_2f	0.7	F + H + F	H + F + H	big change		
23467	23466	0.71	OCH3	NH2	small change		
23467	23468	0.71	OCH3	NHCOC H2CH3	big change		
p38a_3fl n	p38a_2g	0.72	F	H	small change		
1h1r	loi9	0.72	CL + H	H + OH	small change		
CAT-13b	CAT-17g	0.73	CH2CH3 + CL- Benzene	3Member Ring + F- Pyridine	big change	ring disappear	Cring --> Nring
29	27	0.73	CF3	H	big change		
18626-1	18658-1	0.73	CL + H + H	OCH3 + OCH3 + OH	big change		
56	60	0.74	NCH3	S	small change		Nring --> Sring
p38a_2u	p38a_2k	0.74	CH3SO2 -Nring	H	big change	ring disappear	
22	1h1r	0.74	SCH3	CL	small change		
17124-1	18634-1	0.77	BR	H	small change		

Table 23 (cont'd)

23467	23475	0.77	OCH3	NH-5Member Ring	big change	Ring disappear	
CAT-17g	CAT-13c	0.78	3Member Ring + F-Pyridine	CH(CH3)2 + CL-Benzene	big change	ring disappear	Nring --> Cring
ejm_31	ejm_48	0.79	CH3	5Member Ring	big change	Ring disappear	
18638-1	18658-1	0.79	SO2CH3 + H	H + OH	big change		
CAT-4m	CAT-4n	0.79	CH3	CN	small change		
18638-1	18634-1	0.8	SO2CH3	H	big change		
CAT-13d	CAT-13b	0.8	CH2CH3	3Member Ring	big change	ring disappear	
jmc_23	jmc_30	0.81	F	CN	small change		
18627-1	18630-1	0.81	CL	CH3	small change		
35	37	0.81	H	CH3	small change		
18637-1	18631-1	0.82	OCH3 + NHCOC H3	H + H	big change		
29	35	0.82	CF3 + H	H + CH3	big change		
30	26	0.83	SO2CH3	OCH3	big change		
ejm_47	ejm_55	0.84	4Member Ring	OCH3	big change	Ring disappear	
44	23	0.84	NH + CL	S + H	small change		Nring --> Sring
67	31	0.84	O + CH3 + CL + CH3	NH + CF3 + H + H	big change		Oring --> Nring
23484	23482	0.85	Benzene + H	H + BR	big change	Ring disappear	
1a	3b	0.87	F	CH2CH3	small change		
CAT-13a	CAT-13m	0.88	CH3	2Nring	big change	ring disappear	
loiy	32	0.89	CONH2	SO2NH2	big change		

Table 23 (cont'd)

CAT-13e	CAT-17i	0.89	4Member Ring + CL-Benzene	3Member Ring + CH3-Pyridine	big change		Cring --> Nring
loi9	loi9	0.89	CONH2	OH	big change		
20667(2q bp)	23486	0.9	BR	CL	small change		
CAT-4m	CAT-4k	0.9	CH3-Pyridine	Pyridine	big change		
1h1r	21	0.91	CL	OCH3	small change		
ejm_42	ejm_48	0.93	CH2CH3	5Member Ring	big change	Ring disappear	
23477	23330(2q bq)	0.93	H+H+O+H+H	CH3+CH3+CH2+CH3+CH3	big change		
p38a_2v	p38a_2y	0.95	H	C(CH3)2 OH	big change		
18633-1	18624-1	0.97	OCH3	H	small change		
23477	23466	0.99	NH-6Member Ring	H	big change	Ring disappear	
ejm_31	ejm_46	1.02	CH3	3Member Ring	big change	Ring disappear	
65	67	1.02	S + CL	O + H	small change		Sring --> Oring
17	22	1.02	BR	SCH3	small change		
18626-1	18634-1	1.03	CL + H	OCH3 + OCH3	big change		
jmc_28	jmc_27	1.05	CH3	CL	small change		
CAT-13c	CAT-17i	1.05	CH(CH3)2 + CL-Benzene	3Member Ring + CH3-Pyridine	big change	ring disappear	Cring --> Nring
23467	23474	1.08	OCH3	NHCH2-6Member Ring	big change	Ring disappear	
23476	23466	1.08	6Member Ring	H	big change	Ring disappear	

Table 23 (cont'd)

31	32	1.09	CH3 + H	NH2 + OCH3	big change		
20667(2q bp)	23485	1.09	BR	CH3	small change		
26	44	1.09	O	NH	small change		Oring --> Nring
1a	5	1.1	F	H	small change		
28	47	1.1	CH3-Benzene	1H-Indole	big change	ring disappear	
18626-1	18624-1	1.12	CL	H	small change		
ejm_49	ejm_50	1.13	Benz	CH2OH	big change		
43	27	1.13	Naphthalene	Benzene	big change	ring disappear	
35	34	1.14	CH3 + CL	H + CF3	big change		
p38a_2g	p38a_2c	1.15	O-Benzene-F	Benzene-CL	big change		
p38a_3fln	p38a_2ff	1.16	O	CHOH	small change		Oring --> Cring
CAT-17g	CAT-17f	1.17	F	CN	small change		
26	loi9	1.19	OCH3	OH	small change		
CAT-17c	CAT-17e	1.19	CN-Benzene	OCH3-Pyridine	big change		One Nring-Cring change
23482	23486	1.21	H	Benzene	big change	Ring disappear	
CAT-24	CAT-17i	1.22	CCCH3	CH3	small change		
23473	20669(2q br)	1.22	O	NH	small change		
18626-1	18628-1	1.23	CL + H	H + CH3	small change		
p38a_3fln	p38a_2o	1.24	H	CH2SO2 CH3	big change		
p38a_2u	p38a_2q	1.25	O-SO2-CH3	SO2	big change		Nring --> Sring
20670(2q bs)	23482	1.27	CH2	N-SO2CH3	big change		

Table 23 (cont'd)

p38a_3fl n	p38a_2p	1.28	H	CH2CH2 SO2CCH 3	big change		
p38a_3fl z	p38a_2c	1.28	O- Benzene	CL- Benzene	big change		
23486	23479	1.28	SO2- CH2- Benzene + CL	COCH3 + BR	big change	Ring disappear	
33	27	1.29	CL	H	small change		
27	23	1.31	NH + Benzene	S + Naphthal ene	big change	ring disappear	Nring --> Sring
CAT-24	CAT-17e	1.32	CCCH3	OCH3	small change		
67	63	1.32	O + H	S + CL	small change		Oring --> Sring
17	21	1.34	BR	OCH3	small change		
p38a_3fl n	p38a_2e	1.37	F	H	small change		
CAT-17g	CAT-17c	1.38	F- Pyridine	CN- Benzene	big change		One Nring- Cring change
43	47	1.39	Naphthal ene	1H- Indole	big change		Cring --> Nring
18639-1	18658-1	1.41	NO2	H	big change		
ejm_42	ejm_55	1.42	CH2CH3	OCH3	small change		
CAT-13e	CAT-17g	1.42	4Member Ring + CL- Benzene	3Member Ring + Pyridine	big change		Cring --> Nring
67	27	1.44	O + CH3 + CL + CH3	NH + H + H + H	big change		Nring --> Oring
18632-1	18624-1	1.46	OCH3	H	small change		
1h1s	1oiy	1.46	SO2NH2	CONH2	small change		
p38a_3fl n	p38a_2k	1.47	CH3	H	small change		

Table 23 (cont'd)

18626-1	18659-1	1.49	CL + H + H	OCH3 + OCH3 + OCH3	big change		
30	27	1.49	CH3	H	small change		
35	39	1.49	CL + CH3	Benzene + H	big change	ring disappear	
CAT-17f	CAT-17e	1.5	CN	OCH3	small change		
CAT-4a	CAT-4o	1.55	Benzene	F- Pyridine	big change		One Nring- Cring change
p38a_2aa	p38a_2b b	1.55	C(CH2O H)2	CH2CH2 SO2CH3	big change		
p38a_2s	p38a_2l	1.55	CH(CH2 OH)OH	CH3	big change		
20670(2q bs)	23466	1.56	6Member Ring	H	big change	Ring disappear	
CAT-13h	CAT-17i	1.56	CH3 + m-CL- Benzene	H + CH3- Pyridine	big change		One Nring- Cring change
p38a_2z	p38a_3fl q	1.58	CH3 + CH2OH	H + CH2SO2 CH3	big change		
28	26	1.58	SONHC H3	OCH3	big change		
23474	23466	1.59	CH2- 6Member Ring	H	big change	Ring disappear	
67	58	1.62	O + H	NCH3 + CL	big change		Oring --> Nring
p38a_2ee	p38a_3fl n	1.62	NCOCH 3	O	big change		Nring --> Oring
p38a_2z	p38a_3f mk	1.63	CH3 + CH2OH	H + CH2- tetrazole	big change	ring disappear	
CAT-17g	CAT-13i	1.67	3Member Ring + F- Pyridine	CH2- 3Member Ring + CL- Benzene	big change		Nring --> Cring
23467	23469	1.68	OCH3	NHCO- Benzene	big change	Ring disappear	

Table 23 (cont'd)

49	35	1.7	CL	H	small change		
52	60	1.71	NH + CL + H	S + H + CH ₃	big change		Nring --> Sring
p38a_2j	p38a_2q	1.72	CH ₂ + CH ₃	O + Sring	big change	ring disappear	
CAT-13o	CAT-17h	1.76	SNring + CL-Benzene	3Member Ring + CL-Pyridine	big change		Two Nring-Cring change
CAT-4o	CAT-4d	1.77	Pyridine	CH ₃ CH ₂ O-Benzene	big change		One Nring-Cring change
1d	5	1.78	I	H	small change		
17124-1	18631-1	1.79	BR + OCH ₃	H + H	big change		
p38a_3fl_y	p38a_2x	1.79	CH(CH ₃) ₂	Oring	big change	ring disappear	
ejm_42	ejm_54	1.8	CH ₂ CH ₃	NHCH ₂ CH ₃	small change		
66	23	1.8	CL	H	small change		
65	60	1.81	CL	H	small change		
CAT-4p	CAT-13k	1.83	Benzene + CL-Pyridine	Nring + CL-Benzene	big change		Two Nring-Cring change
23469	20669(2q br)	1.87	CO-Benzene	CH ₂ -Benzene	small change		
23469	23472	1.88	CO-Benzene	SO ₂ -Benz-CL	big change		
ejm_55	ejm_54	1.88	OCH ₃	NHCH ₂ CH ₃	small change		
18639-1	18634-1	1.9	NO ₂	H	big change		
20667(2q bp)	23484	1.92	BR	H	small change		
CAT-13a	CAT-17i	1.93	CH ₃ + m-CL-Benzene	3Member Ring + CH ₃ -Pyridine	big change	ring disappear	Cring --> Nring

Table 23 (cont'd)

18631-1	18624-1	1.97	OCH3	H	small change		
ejm_31	jmc_28	1.99	CH3	3Member Ring-CH3	big change	Ring disappear	
38	60	2	NH + H + H + Benzene	S + CH3 + CL + CH3	big change	ring disappear	Nring --> Sring
28	27	2.01	CH3	H	small change		
54	42	2.02	CL	H	small change		
35	36	2.03	H	CH3	small change		
ejm_49	ejm_31	2.04	Benz	CH3	big change		
jmc_28	jmc_30	2.04	CH3	CN	small change		
1oiu	26	2.05	SO2NH2 + H	H + OCH3	big change		
p38a_3fly	p38a_3flq	2.09	H	CH2SO2 CH3	big change		
58	60	2.11	NCH3 + CL	S + H	big change		Nring --> Sring
23482	23485	2.11	BR + H	CH3 + CH2-Benzene	big change	Ring disappear	
CAT-4m	CAT-4j	2.12	H	CH3	small change		
1oiu	1h1q	2.20	SO2NH2	H	big change		
ejm_44	ejm_55	2.21	C(CH3)3	OCH3	big change		
50	60	2.26	NH + CL	S + H	small change		Nring --> Sring
CAT-4m	CAT-13j	2.28	Benzene + CH3-Pyridine	Pyridine + CL-Benzene	big change		Two Nring-Cring change
CAT-13g	CAT-17g	2.3	5Member Ring + CL-Benzene	3Member Ring + Pyridine	big change		Cring --> Nring
56	35	2.3	CH3 + CH3	H + H	small change		

Table 23 (cont'd)

60	36	2.31	S + CH3 + H	NH + H + CH3	big change		Sring --> Nring
18635-1	18625-1	2.32	CH3 + CH3	CL + H	small change		
CAT-17i	CAT-13f	2.32	3Member Ring + CH3-Pyridine	5Member Ring + CL-Benzene	big change		Nring --> Cring
63	60	2.35	CL	H	small change		
48	27	2.4	Benzene	1H-Indole	big change	ring disappear	
p38a_2g	p38a_2i	2.41	O	CH2	small change		
p38a_2v	p38a_2j	2.41	CH3 + O	Oring + CH2	big change	ring disappear	
49	67	2.43	NH + CL + H	O + H + CH3	big change		
p38a_2l	p38a_2j	2.43	CH3 + O	H + CH2	small change		
41	32	2.45	O-Benzene	CH3	big change	ring disappear	
CAT-4m	CAT-4l	2.49	CH3-Pyridine	2Nring	big change		one ring change
18626-1	18660-1	2.52	CL + H + H	OCH3 + OCH3 + SO2CH3	big change		
p38a_3fln	p38a_2g	2.53	O	SO2	big change		Oring --> Sring
jmc_23	ejm_55	2.54	F-3Member Ring	OCH3	big change	Ring disappear	
p38a_2v	p38a_3fmk	2.56	CH3	CH(CH3)(CH2-tetrazole)	big change	ring disappear	
26	64	2.59	O + H	S + CL	small change		Oring --> Sring
CAT-4m	CAT-13m	2.59	Benzene +CH3-Pyridine	2Nring + CL-Benzene	big change		Two Nring-Cring change
CAT-4o	CAT-4b	2.6	Pyridine	OCH3-Benzene	big change		one Nring-Cring change

Table 23 (cont'd)

35	53	2.6	H + H	CL + CH ₃	small change		
CAT-4m	CAT-13k	2.6	Benzene + CH ₃ -Pyridine	Nring + CL-Benzene	big change		Two Nring-Cring change
p38a_2j	p38a_2ff	2.61	CH ₂ + O	O + CH(OH)	big change		Oring --> Cring
CAT-4c	CAT-4o	2.68	OCH ₃ -Benzene	Pyridine	big change		one Nring-Cring change
17	1h1q	2.69	BR	H	big change		
67	52	2.69	O + H + CH ₃	NH + CL + H	big change		Oring --> Nring
p38a_2g	p38a_2h	2.72	O	S	small change		
p38a_3fl n	p38a_2i	2.73	O + F	CH ₂ + H	small change		
CAT-13a	CAT-17g	2.85	CH ₃ + m-CL-Benzene	3Member Ring + Pyridine	big change	ring disappear	Cring --> Nring
18635-1	18624-1	2.86	CH ₃ + CH ₃	H + H	small change		
35	33	2.89	CH ₃	H	small change		
18626-1	18625-1	2.93	H + CL	CL + H	small change		
CAT-13d	CAT-13h	2.96	H	CH ₃	small change		
CAT-13g	CAT-17i	3.02	6Member Ring + m-CL-Benzene	3Member Ring + Pyridine	big change		Cring --> Nring
p38a_2m	p38a_2k	3.09	3Member Ring	H	big change	ring disappear	
p38a_2j	p38a_2r	3.11	CH ₂ + H	O + C(OH)(C H ₃) ₂	big change		
67	37	3.12	O	NH	small change		Oring --> Nring

Table 23 (cont'd)

CAT-13d	CAT-17d	3.15	m-CL-Benzene	Pyridine	big change		One Nring-Cring change
67	50	3.15	O + H	NH + CL	small change		Oring --> Nring
42	64	3.25	NH + H	S + CL	small change		Nring --> Sring
35	60	3.25	NH + H	S + CH3	small change		Nring --> Sring
66	42	3.27	S + CL	NH + H	small change		Sring --> Nring
23484	23479	3.44	SO2-CH2-Benzene + H	COCH3 + BR	big change	Ring disappear	
67	32	3.55	CH3 + CL + CH3 + O	H + CH3 + H + NH	big change		Oring --> Nring
p38a_2t	p38a_2j	3.57	H	CN	small change	triple bond	
23484	23486	3.77	H	CL	small change		
41	35	3.78	O-Benzene + H	CL + CH3	big change	ring disappear	
29	40	3.79	CF3 + H	H + O-Benzene	big change	ring disappear	
67	35	3.88	CH3 + O	H + NH	small change		Oring --> Nring
18631-1	18660-1	4.16	H + H	OCH3 + SO2CH3	big change		
23484	23485	4.21	H	CH3	small change		
38	35	4.34	H + H + Benzene	H + CH3 + CL	big change	ring disappear	
p38a_2g_g	p38a_2j	4.38	O + SO2	CH2 + O	big change		Sring --> Oring
p38a_2e	p38a_2j	4.47	O + H	CH2 + F	small change		
p38a_2v	p38a_3f_mh	4.5	CH3	CH(CH3)(CH2-tetrazole)	big change	ring disappear	

Table 23 (cont'd)

CAT-13n	CAT-4i	4.58	m-CL-Benzene + 2Nring	Pyridine + Benzene	big change		Two Nring-Cring change
23485	23479	4.67	SO2-CH2-Benzene + CH3	COCH3 + BR	big change	Ring disappear	
CAT-4i	CAT-13m	5.8	Pyridine + Benzene	m-CL-Benzene + 2Nring	big change		Two Nring-Cring change

Table 24. The ΔG values obtained in pKa calculations mentioned in section 4.2.2.3.2

ΔG (kcal/mol)					
HIS6 (HIP--->HID)		HIS212 (HIP--->HID)		HIS_water (HIP--->HID)	
run1	36.4	run1	35.9	run1	29.3
run2	34.7	run2	35.4	run2	29.2
run3	36.9	run3	36.3	run3	29.3
run4	35.4	run4	35.9	run4	29.2
run5	36.5	run5	35.8	run5	28.4
run6	36.2	run6	36.1	run6	27.5
run7	36.0	run7	35.6	run7	25.7
run8	36.4	run8	36.3	run8	26.6
run9	37.1	run9	36.6	run9	26.2
Average	36.2	Average	36.0	Average	27.9
Standard deviation	0.7	Standard deviation	0.4	Standard deviation	1.5
HIS6 (HIP--->HIE)		HIS212 (HIP--->HIE)		HIS_water (HIP--->HIE)	
run1	35.2	run1	33.0	run1	30.2
run2	39.3	run2	34.1	run2	30.2
run3	34.8	run3	33.9	run3	29.8
run4	35.0	run4	33.3	run4	30.7
run5	34.3	run5	33.8	run5	29.7
run6	36.3	run6	34.3	run6	29.0
run7	37.8	run7	33.4	run7	30.5
run8	35.0	run8	33.7	run8	29.0

Table 24 (cont'd)

run9	36.9	run9	34.9	run9	29.0
Average	36.1	Average	33.8	Average	29.8
Standard deviation	1.7	Standard deviation	0.6	Standard deviation	0.7

Table 25. The binding free energies for Ni²⁺ and Co²⁺ to imidazole and acetate via PMF calculations mentioned in section 4.2.1.1

Optimized m12-6-4						
Imidazole						
	Alpha	Run1	Run2	Run3	Average (kcal/mol)	Standard Deviation (kcal/mol)
Co ²⁺	2.230	-3.60	-3.48	-3.55	-3.54	0.06
Ni ²⁺	2.310	-4.27	-4.40	-4.14	-4.27	0.13
Acetate						
Co ²⁺	0.120	-1.19	-1.02	-0.89	-1.03	0.15
Ni ²⁺	0.145	-0.82	-0.61	-0.75	-0.73	0.11
Default 12-6-4						
Imidazole						
Co ²⁺	1.090	3.92	3.96	3.92	3.93	0.02
Ni ²⁺	1.090	4.92	4.76	4.99	4.89	0.12
Acetate						
Co ²⁺	0.569	-4.11	-3.98	-4.18	-4.09	0.10
Ni ²⁺	0.569	-4.13	-3.88	-4.77	-4.26	0.46

Table 26. Summary of ΔG_A values of the GlxI-Ni²⁺ system by the double decoupling method (DDM). For each system, nine runs were performed using different restraint strength. Set 1, Set 2 and Set 3 are the three sets of DDM calculations starting with the structure and velocity from the last snap-shot of the 100 ns, 200 ns, 300 ns free MD simulations, respectively.

Restraint force constants k _r : kcal/(mol*Å ²), k _θ : kcal/(mol*rad ²), k _φ : kcal/(mol*rad ²)	ΔG_A (kcal/mol), GlxI-Ni ²⁺					
	Set 1		Set 2		Set 3	
	Default 12-6-4	m12-6-4	Default 12-6-4	m12-6-4	Default 12-6-4	m12-6-4

Table 26 (cont'd)

400,400,400	37.2	39.3	30.9	45.8	30.0	35.8
500,500,500	31.3	41.4	26.0	40.6	26.4	39.0
600,600,600	30.0	43.7	30.7	42.8	25.2	42.6
700,700,700	35.7	41.5	26.2	49.4	24.0	41.1
800,800,800	37.4	45.5	28.4	44.4	29.7	36.2
900,900,900	32.7	39.1	27.8	40.0	28.5	43.2
1000,1000,1000	27.3	31.3	28.3	38.0	31.3	38.6
1100,1100,1100	26.5	31.6	28.4	43.2	28.9	44.9
1200,1200,1200	37.6	44.0	28.0	38.0	21.5	42.2
Average	32.9	39.7	28.3	42.5	27.3	40.4
Standard Deviation	4.4	5.1	1.7	3.8	3.2	3.2
Overall	Default 12-6-4	m12-6-4				
Average	29.5	40.9				
Standard Deviation	4.0	4.1				

APPENDIX: Figures

Figure 31. The perturbation graph plotted based on the work of Wang et.al⁸ for the GPU TI study in Chapter 2.

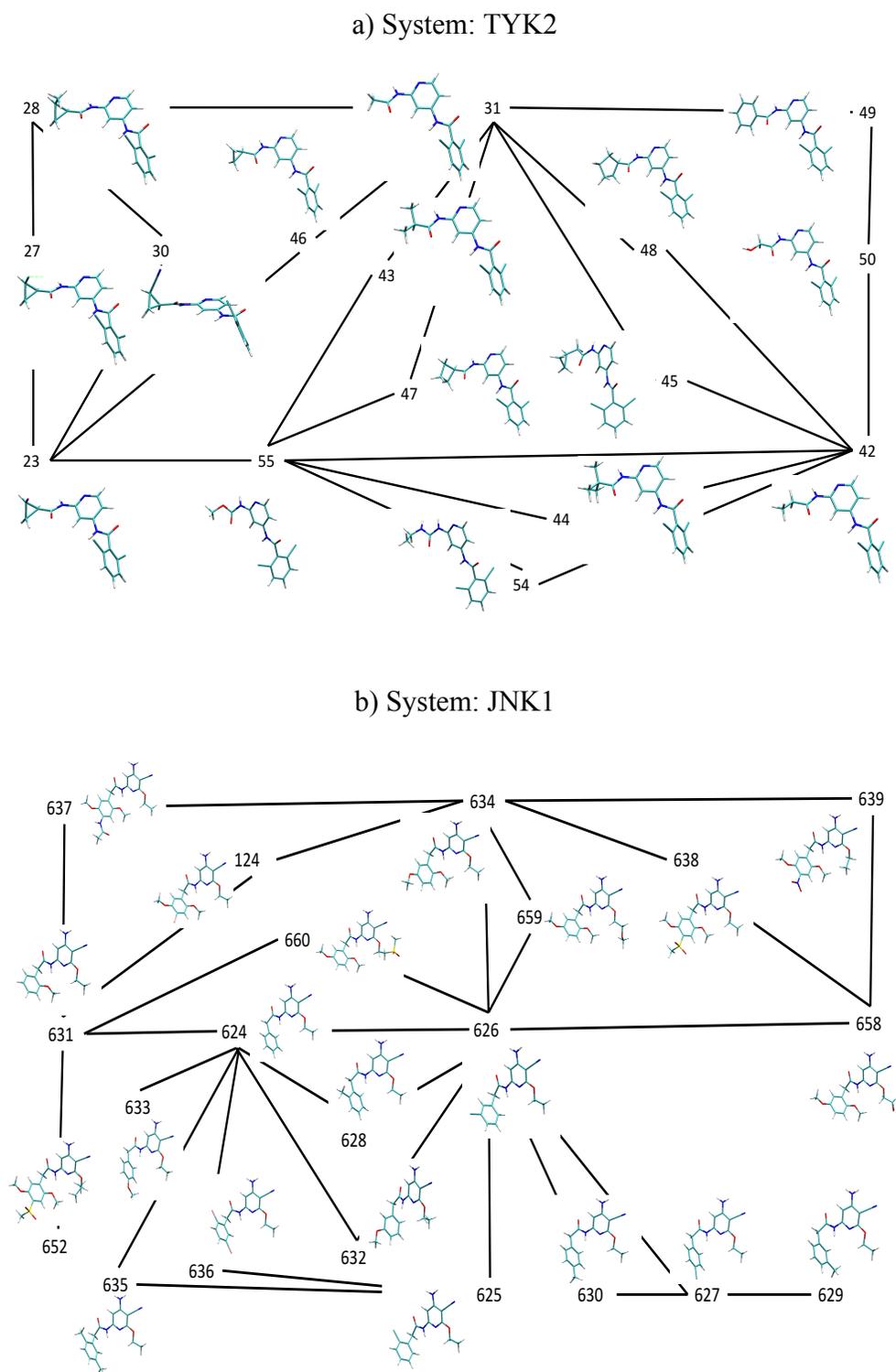
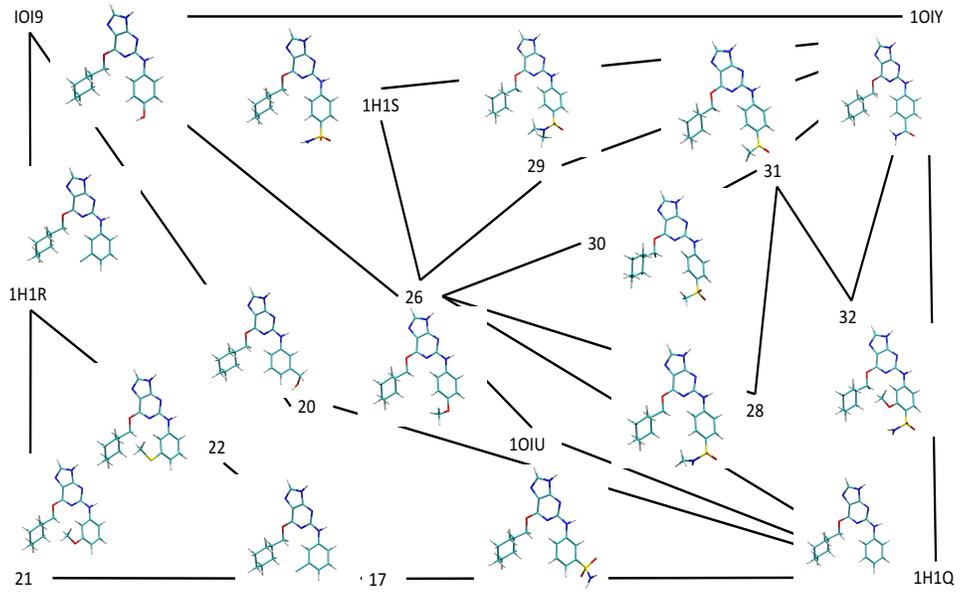


Figure 31 (cont'd)

c) System: CDK2



d) System: PTP1B

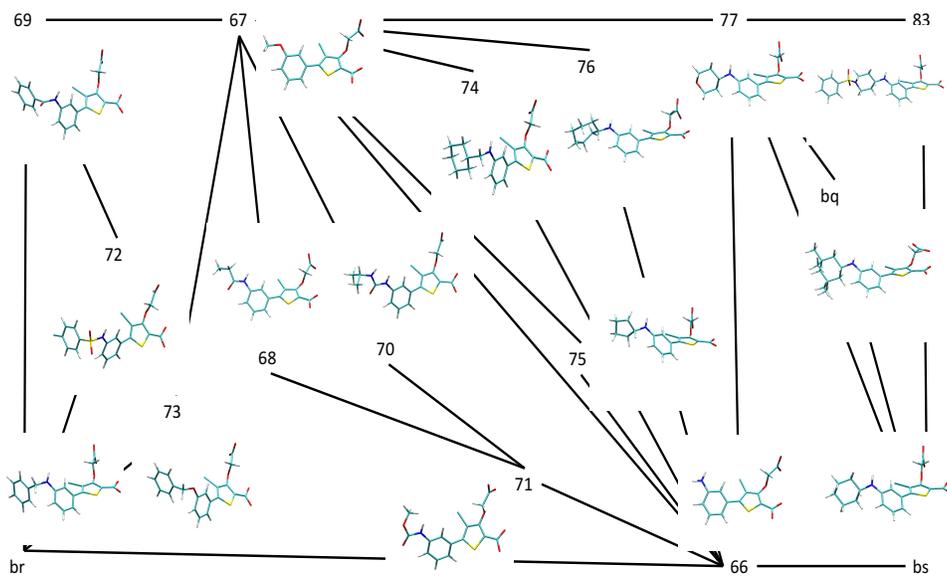
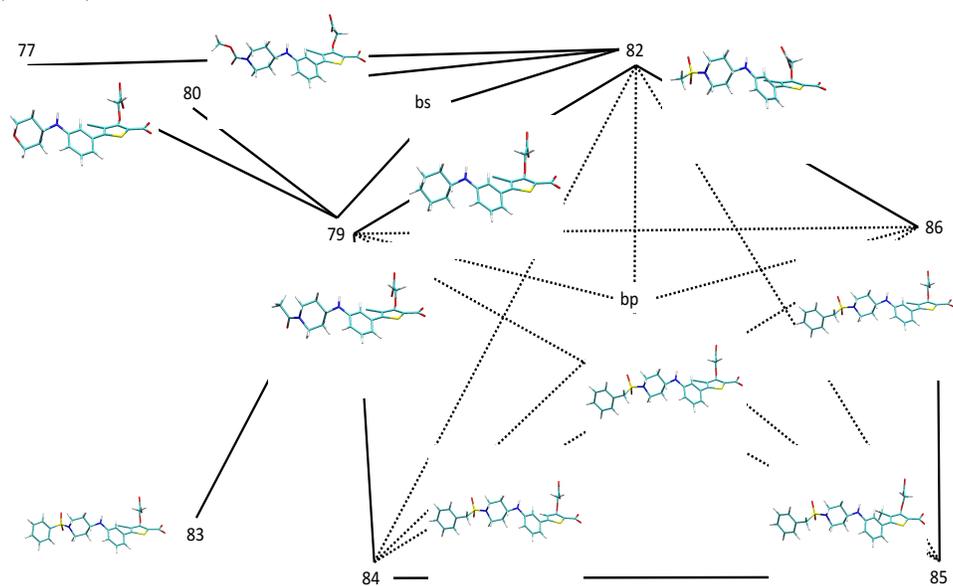


Figure 31 (cont'd)



e) System: BACE

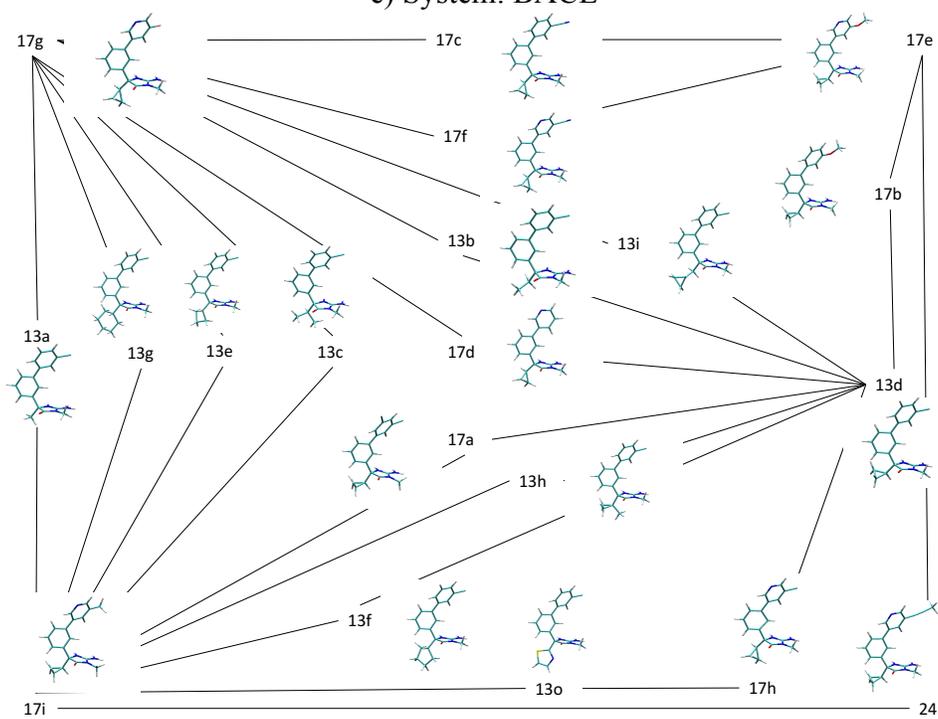
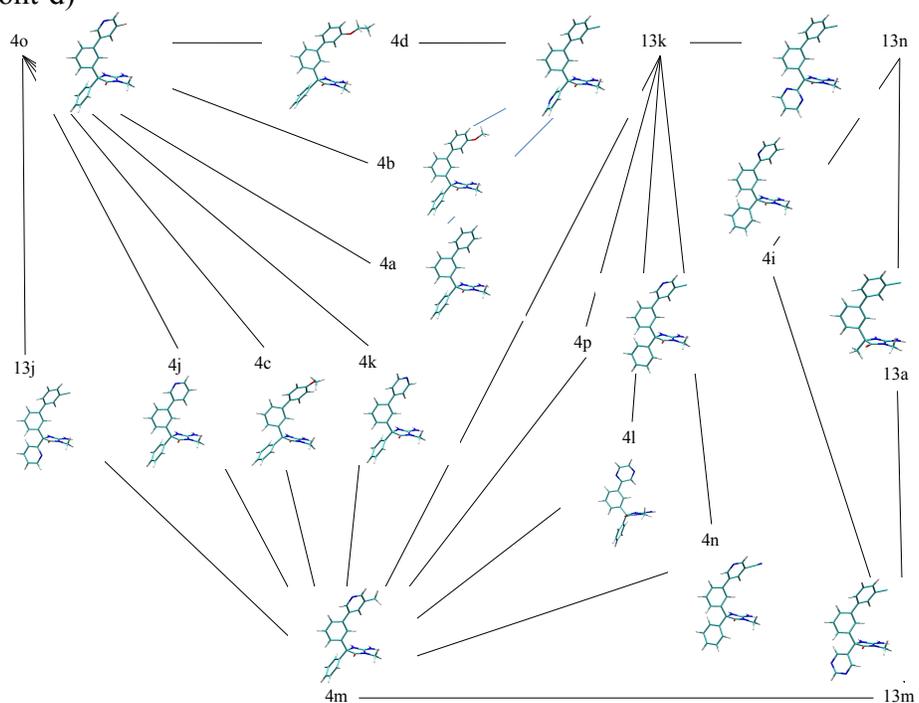


Figure 31 (cont'd)



f) System: MCL1

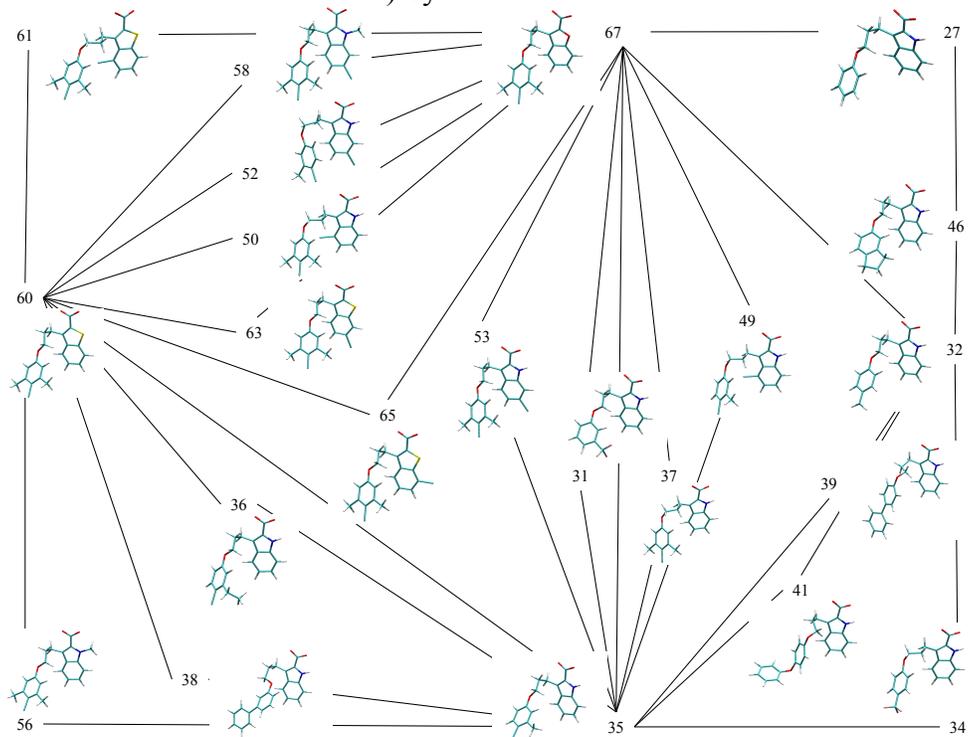
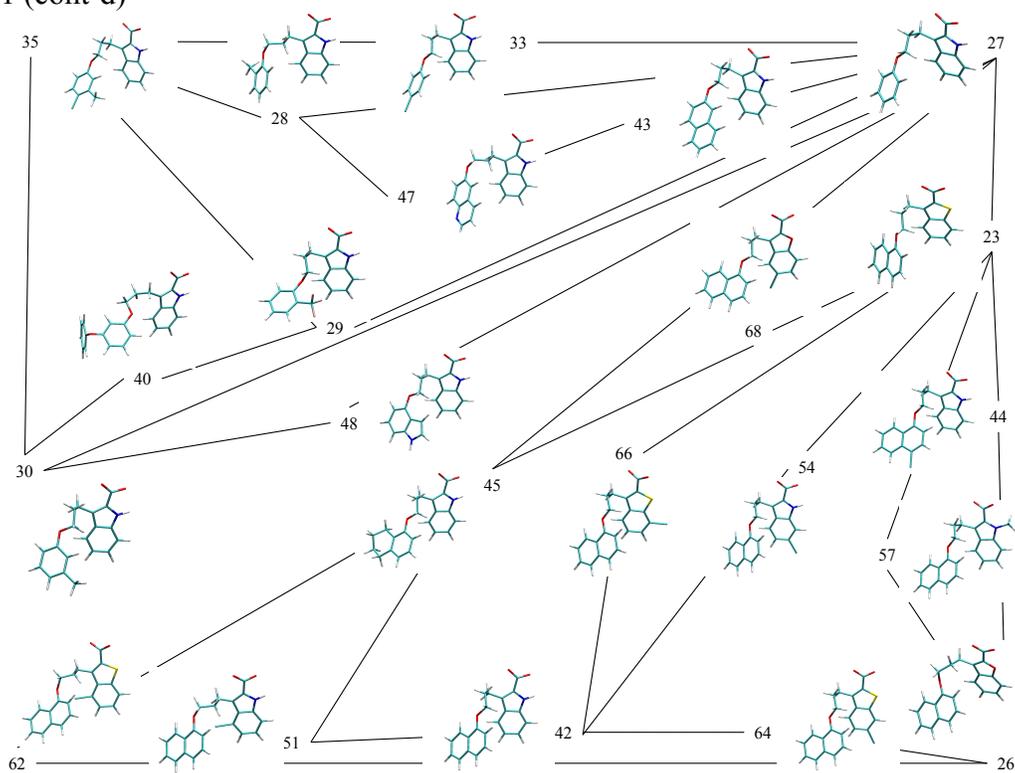


Figure 31 (cont'd)



g) System: P38

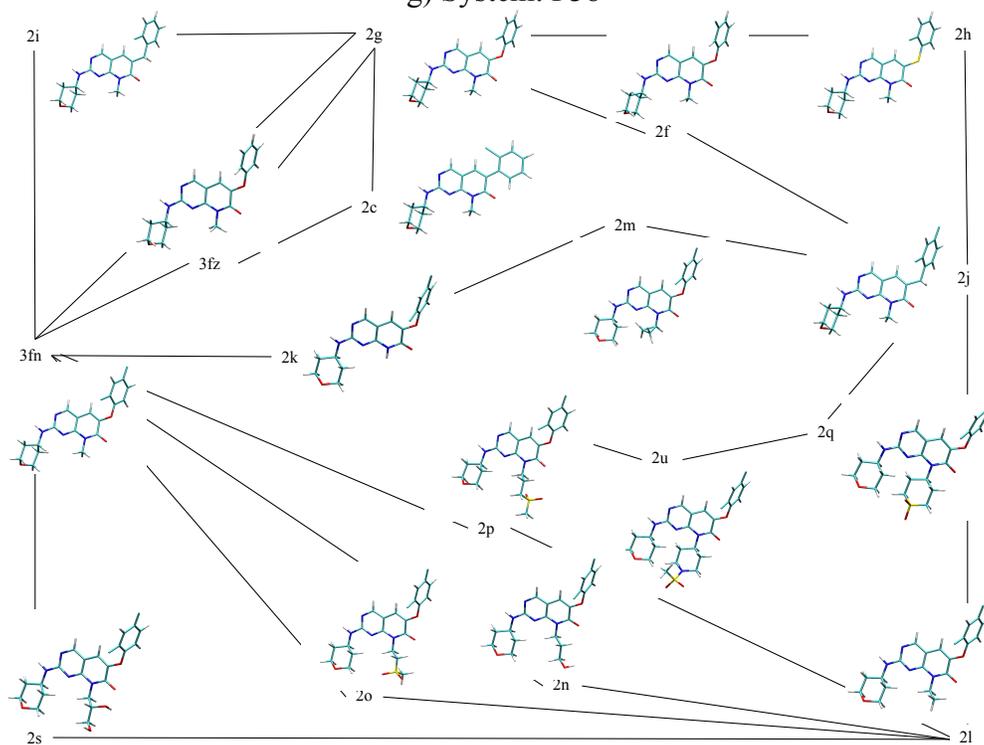


Figure 31 (cont'd)

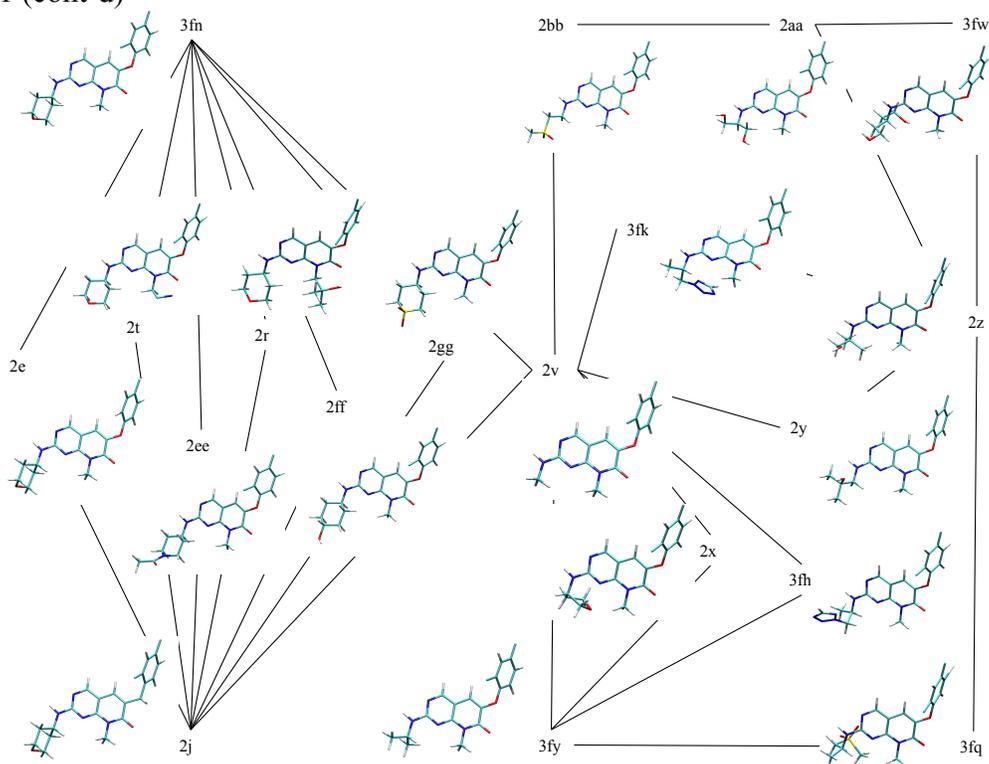
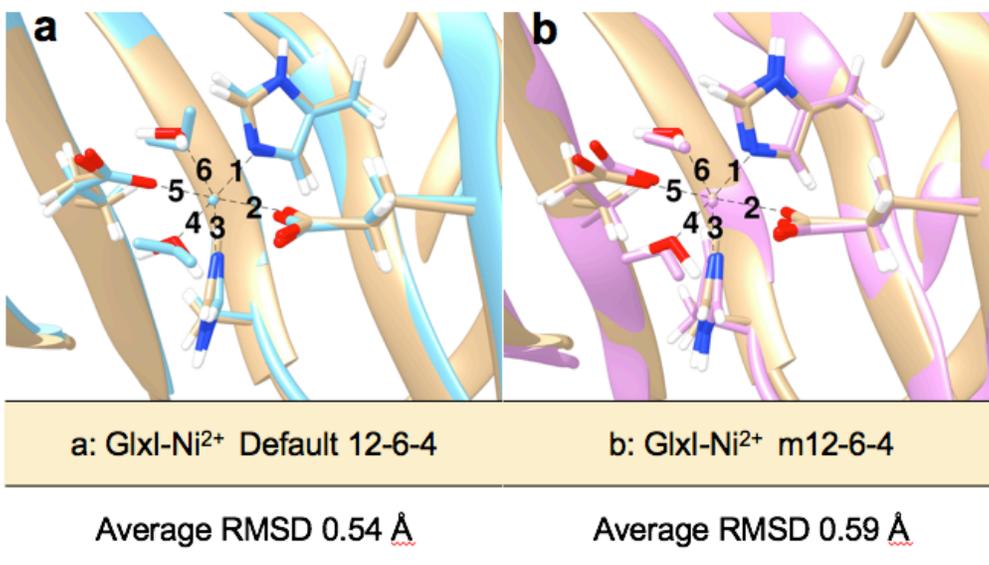


Figure 32. Geometries of Ni^{2+} bound protein obtained after 300ns MD simulations with (a) default 12-6-4 and (b) m12-6-4 potential aligned with crystal structure (light orange). The RMSD measurements are based on the side chain of the two HIS and the two GLU, plus the metal ion and the two water molecules.



BIBLIOGRAPHY

BIBLIOGRAPHY

1. Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S., Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150-160.
2. Wlodawer, A., Rational approach to AIDS drug design through structural biology. *Annu. Rev. Med.* **2002**, *53*, 595-614.
3. Williams, J. A.; Bauman, J.; Cai, H.; Conlon, K.; Hansel, S.; Hurst, S.; Sadagopan, N.; Tugnait, M.; Zhang, L.; Sahi, J., In vitro ADME phenotyping in drug discovery: current challenges and future solutions. *Curr. Opin. Drug. Discov. Devel.* **2005**, *8*, 78-88.
4. Szakacs, G.; Varadi, A.; Ozvegy-Laczka, C.; Sarkadi, B., The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). *Drug Discov. Today* **2008**, *13*, 379-93.
5. Ekins, S.; Nikolsky, Y.; Nikolskaya, T., Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. *Trends Pharmacol. Sci.* **2005**, *26*, 202-9.
6. Caldwell, J.; Gardner, I.; Swales, N., An introduction to drug disposition: the basic principles of absorption, distribution, metabolism, and excretion. *Toxicol. Pathol.* **1995**, *23*, 102-14.
7. Balani, S. K.; Miwa, G. T.; Gan, L. S.; Wu, J. T.; Lee, F. W., Strategy of utilizing in vitro and in vivo ADME tools for lead optimization and drug candidate selection. *Curr. Top. Med. Chem.* **2005**, *5*, 1033-8.
8. Wang, L.; Wu, Y. J.; Deng, Y. Q.; Kim, B.; Pierce, L.; Krilov, G.; Lupyán, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695-2703.
9. Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

10. Mccammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585-590.
11. Merz, K. M.; Kollman, P. A., Free-Energy Perturbation Simulations of the Inhibition of Thermolysin - Prediction of the Free-Energy of Binding of a New Inhibitor. *J. Am. Chem. Soc.* **1989**, *111*, 5649-5658.
12. Mobley, D. L.; Gilson, M. K., Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531-558.
13. Mobley, D. L.; Klimovich, P. V., Perspective: Alchemical free energy calculations for drug discovery. *J. Chem. Phys.* **2012**, *137*.
14. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889-897.
15. Kuhn, B.; Kollman, P. A., Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786-3791.
16. Li, Y.; Liu, Z. H.; Wang, R. X., Test MM-PB/SA on True Conformational Ensembles of Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2010**, *50*, 1682-1692.
17. Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M., Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31*, 797-810.
18. Lazaridis, T.; Masunov, A.; Gandolfo, F., Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins* **2002**, *47*, 194-208.
19. Luo, H.; Sharp, K., On the calculation of absolute macromolecular binding free energies. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 10399-404.
20. Luo, R.; Gilson, M. K., Synthetic adenine receptors: Direct calculation of binding affinity and entropy. *J. Am. Chem. Soc.* **2000**, *122*, 2934-2937.
21. Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A., Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* **1998**, *16*, 671-82.
22. Vorobjev, Y. N.; Hermans, J., ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. *Biophys. Chem.* **1999**, *78*, 195-205.

23. Swanson, J. M.; Henchman, R. H.; McCammon, J. A., Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **2004**, *86*, 67-74.
24. Guitierrez-de-Teran, H.; Aqvist, J., Linear Interaction Energy: Method and Applications in Drug Design. *Methods Mol Biol* **2012**, *819*, 305-323.
25. Bennett, C. H., Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245-268.
26. Kirkwood, J. G., Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300-313.
27. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*.
28. Zwanzig, R. W., High-Temperature Equation of State by a Perturbation Method .1. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420-1426.
29. Zwanzig, R. W.; Kirkwood, J. G.; Oppenheim, I.; Alder, B. J., Statistical Mechanical Theory of Transport Processes .7. The Coefficient of Thermal Conductivity of Monatomic Liquids. *J. Chem. Phys.* **1954**, *22*, 783-790.
30. Kong, X. J.; Brooks, C. L., lambda-Dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414-2423.
31. Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A., Calculations of Antibody Antigen Interactions - Microscopic and Semimicroscopic Evaluation of the Free-Energies of Binding of Phosphorylcholine Analogs to Mcpc603. *Protein Eng.* **1992**, *5*, 215-228.
32. Bhakat, S.; Soderhjelm, P., Resolving the problem of trapped water in binding cavities: prediction of host-guest binding free energies in the SAMPL5 challenge by funnel metadynamics. *J. Comput. Aided Mol. Des.* **2017**, *31*, 119-132.
33. Doudou, S.; Burton, N. A.; Henchman, R. H., Standard Free Energy of Binding from a One-Dimensional Potential of Mean Force. *J. Chem. Theory. Comput.* **2009**, *5*, 909-918.
34. Henriksen, N. M.; Fenley, A. T.; Gilson, M. K., Computational Calorimetry: High-Precision Calculation of Host-Guest Binding Thermodynamics. *J. Chem. Theory. Comput.* **2015**, *11*, 4377-4394.
35. Hsiao, Y. W.; Soderhjelm, P., Prediction of SAMPL4 host-guest binding affinities using funnel metadynamics. *J. Comput. Aided Mol. Des.* **2014**, *28*, 443-454.

36. Lee, M. S.; Olson, M. A., Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys. J.* **2006**, *90*, 864-877.
37. Velez-Vega, C.; Gilson, M. K., Overcoming Dissipation in the Calculation of Standard Binding Free Energies by Ligand Extraction. *J. Comput. Chem.* **2013**, *34*, 2360-2371.
38. Ytreberg, F. M., Absolute FKBP binding affinities obtained via nonequilibrium unbinding simulations. *J. Chem. Phys.* **2009**, *130*.
39. Ucisik, M. N.; Zheng, Z.; Faver, J. C.; Merz, K. M., Bringing Clarity to the Prediction of Protein-Ligand Binding Free Energies via "Blurring". *J. Chem. Theory. Comput.* **2014**, *10*, 1314-1325.
40. Woo, H. J.; Roux, B., Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825-30.
41. Aqvist, J.; Luzhkov, V. B.; Brandsdal, B. O., Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002**, *35*, 358-365.
42. Bansal, N.; Zheng, Z.; Cerutti, D. S.; Merz, K. M., On the fly estimation of host-guest binding free energies using the movable type method: participation in the SAMPL5 blind challenge. *J. Comput. Aided Mol. Des.* **2017**, *31*, 47-60.
43. Torrie, G. M.; Valleau, J. P., Monte-Carlo Free-Energy Estimates Using Non-Boltzmann Sampling - Application to Subcritical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28*, 578-581.
44. Jarzynski, C., Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690-2693.
45. Laio, A.; Parrinello, M., Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562-12566.
46. Bastug, T.; Chen, P. C.; Patra, S. M.; Kuyucak, S., Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. *J. Chem. Phys.* **2008**, *128*.
47. Cuendet, M. A.; Michielin, O., Protein-protein interaction investigated by steered molecular dynamics: The TCR-pMHC complex. *Biophys. J.* **2008**, *95*, 3575-3590.
48. Grater, F.; de Groot, B. L.; Jiang, H. L.; Grubmuller, H., Ligand-release pathways in the pheromone-binding protein of *Bombyx mori*. *Structure* **2006**, *14*, 1567-1576.

49. Vashisth, H.; Abrams, C. F., Ligand Escape Pathways and (Un)Binding Free Energy Calculations for the Hexameric Insulin-Phenol Complex. *Biophys. J.* **2008**, *95*, 4193-4204.
50. Zhang, D. Q.; Gullingsrud, J.; McCammon, J. A., Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain (vol 128, pg 3019, 2006). *J. Am. Chem. Soc.* **2006**, *128*, 4493-4493.
51. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.
52. Torrie, G. M.; Valleau, J. P., Monte-Carlo Study of a Phase-Separating Liquid-Mixture by Umbrella Sampling. *J. Chem. Phys.* **1977**, *66*, 1402-1408.
53. Torrie, G. M.; Valleau, J. P., Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187-199.
54. Kosztin, D.; Izrailev, S.; Schulten, K., Unbinding of retinoic acid from its receptor studied by steered molecular dynamics. *Biophys. J.* **1999**, *76*, 188-197.
55. Li, P. F.; Song, L. F.; Merz, K. M., Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJ-Type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 883-895.
56. Li, P.; Song, L. F.; Merz, K. M., Jr., Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theory. Comput.* **2015**, *11*, 1645-57.
57. Li, P. F.; Merz, K. M., Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory. Comput.* **2014**, *10*, 289-297.
58. Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M., Jr., Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory. Comput.* **2013**, *9*, 2733-2748.
59. Sengupta, A.; Seitz, A.; Merz, K. M., Simulating the Chelate Effect. *J. Am. Chem. Soc.* **2018**, *140*, 15166-15169.
60. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **1995**, *16*, 1339-1350.
61. Wang, J. Y.; Deng, Y. Q.; Roux, B., Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798-2814.

62. Jorgensen, W. L.; Thomas, L. L., Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory. Comput.* **2008**, *4*, 869-876.
63. Peierls, R., On the theory of diamagnetism of conduction electrons. *Z. Phys.* **1933**, *80*, 763-791.
64. Kirkwood, J. G.; Alder, B. J., *Theory of liquids*. Gordon and Breach: New York,, 1968; p xxiv, 140 p.
65. Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A., Free energy calculations by computer simulation. *Science* **1987**, *236*, 564-8.
66. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E., Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087-1092.
67. Metropolis, N.; Ulam, S., The Monte Carlo method. *J. Am. Stat. Assoc.* **1949**, *44*, 335-41.
68. Metropolis, N., Monte-Carlo - in the Beginning and Some Great Expectations. *Lect. Notes. Phys.* **1985**, *240*, 62-70.
69. Tanaka, S.; Scheraga, H. A., Model of Protein Folding - Inclusion of Short-Range, Medium-Range, and Long-Range Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 3802-3806.
70. Tanaka, S.; Scheraga, H. A., Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* **1976**, *9*, 945-950.
71. Levitt, M.; Lifson, S., Refinement of Protein Conformations Using a Macromolecular Energy Minimization Procedure. *J. Mol. Biol.* **1969**, *46*, 269-&.
72. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977**, *267*, 585-90.
73. Levitt, M., The birth of computational structural biology. *Nat. Struct. Biol.* **2001**, *8*, 392-3.
74. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

75. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187-217.
76. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A., An All Atom Force-Field for Simulations of Proteins and Nucleic-Acids. *J. Comput. Chem.* **1986**, *7*, 230-252.
77. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A New Force-Field for Molecular Mechanical Simulation of Nucleic-Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.
78. Jorgensen, W. L.; Tiradorives, J., The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657-1666.
79. van Gunsteren, W. F. B., H. J. C. , GROMOS 86: Groningen Molecular Simulation Program Package. *University of Groningen: Groningen, The Netherlands* **1986**.
80. Mackerell, A. D., Jr., Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* **2004**, *25*, 1584-604.
81. Price, D. J.; Brooks, C. L., Modern protein force fields behave comparably in molecular dynamics simulations. *J. Comput. Chem.* **2002**, *23*, 1045-1057.
82. Edinger, S. R.; Cortis, C.; Shenkin, P. S.; Friesner, R. A., Solvation free energies of peptides: Comparison of approximate continuum solvation models with accurate solution of the Poisson-Boltzmann equation. *J. Phys. Chem. B* **1997**, *101*, 1190-1197.
83. Gilson, M. K.; Davis, M. E.; Luty, B. A.; Mccammon, J. A., Computation of Electrostatic Forces on Solvated Molecules Using the Poisson-Boltzmann Equation. *J. Phys. Chem.* **1993**, *97*, 3591-3600.
84. Warwicker, J.; Watson, H. C., Calculation of the Electric-Potential in the Active-Site Cleft Due to Alpha-Helix Dipoles. *J. Mol. Biol.* **1982**, *157*, 671-679.
85. Honig, B.; Dinur, U.; Nakanishi, K.; Baloghna, V.; Gawinowicz, M. A.; Arnaboldi, M.; Motto, M. G., External Point-Charge Model for Wavelength Regulation in Visual Pigments. *J. Am. Chem. Soc.* **1979**, *101*, 7084-7086.
86. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T., Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.

87. Lu, Y.; Yang, C. Y.; Wang, S., Binding free energy contributions of interfacial waters in HIV-1 protease/inhibitor complexes. *J. Am. Chem. Soc.* **2006**, *128*, 11830-9.
88. Michel, J.; Tirado-Rives, J.; Jorgensen, W. L., Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J. Am. Chem. Soc.* **2009**, *131*, 15403-11.
89. Bernal, J. D.; Fowler, R. H., A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *J. Chem. Phys.* **1933**, *1*, 515-548.
90. Horne, E. R. A., *Structure and Transport Processes in Water and Aqueous Solutions*. Wiley-Interscience, New York . 1972; Vol. 8.
91. Stillinger, F. H.; Rahman, A., Improved Simulation of Liquid Water by Molecular-Dynamics. *J. Chem. Phys.* **1974**, *60*, 1545-1557.
92. Jorgensen, W. L., Quantum and Statistical Mechanical Studies of Liquids .10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers - Application to Liquid Water. *J. Am. Chem. Soc.* **1981**, *103*, 335-340.
93. H. J. C. Berendsen, J. P. M. P., W. F. van Gunsteren, and J. Hermans, In *Intermolecular Forces*, edited by B. Pullman, p. 331. **1981**.
94. Jorgensen, W. L., Quantum and Statistical Mechanical Studies of Liquids .24. Revised Tips for Simulations of Liquid Water and Aqueous-Solutions. *J. Chem. Phys.* **1982**, *77*, 4156-4163.
95. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.
96. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269-6271.
97. Cisneros, G. A.; Wikfeldt, K. T.; Ojamae, L.; Lu, J. B.; Xu, Y.; Torabifard, H.; Bartok, A. P.; Csanyi, G.; Molinero, V.; Paesani, F., Modeling Molecular Interactions in Water: From Pairwise to Many Body Potential Energy Functions. *Chem. Rev.* **2016**, *116*, 7501-7528.
98. Onufriev, A. V.; Izadi, S., Water models for biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*.
99. Ouyang, J. F.; Bettens, R. P. A., Modelling Water: A Lifetime Enigma. *Chimia* **2015**, *69*, 104-111.

100. Jorgensen, W. L.; Tirado-Rives, J., Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6665-6670.
101. Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R., Thermodynamics of Cavity Formation in Water - a Molecular-Dynamics Study. *Faraday Symp. Chem. Soc.* **1982**, 55-67.
102. Jorgensen, W. L.; Ravimohan, C., Monte-Carlo Simulation of Differences in Free-Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050-3054.
103. Jorgensen, W. L.; Madura, J. D.; Swenson, C. J., Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. *J. Am. Chem. Soc.* **1984**, *106*, 6638-6646.
104. Jorgensen, W. L.; Swenson, C. J., Optimized Intermolecular Potential Functions for Amides and Peptides - Hydration of Amides. *J. Am. Chem. Soc.* **1985**, *107*, 1489-1496.
105. Jorgensen, W. L.; Buckner, J. K.; Huston, S. E.; Rossky, P. J., Hydration and Energetics for (Ch₃)₃ccl Ion-Pairs in Aqueous-Solution. *J. Am. Chem. Soc.* **1987**, *109*, 1891-1899.
106. Singh, U. C.; Brown, F. K.; Bash, P. A.; Kollman, P. A., An Approach to the Application of Free-Energy Perturbation-Methods Using Molecular-Dynamics - Applications to the Transformations of Ch₃oh-] Ch₃ch₃, H₃o+]-] Nh₄⁺, Glycine-] Alanine, and Alanine-] Phenylalanine in Aqueous-Solution and to H₃o+(H₂o)₃-] Nh₄⁺(H₂o)₃ in the Gas-Phase. *J. Am. Chem. Soc.* **1987**, *109*, 1607-1614.
107. Tembe, B. L.; Mccammon, J. A., Ligand Receptor Interactions. *Comput. Chem.* **1984**, *8*, 281-283.
108. Lybrand, T. P.; Ghosh, I.; Mccammon, J. A., Hydration of Chloride and Bromide Anions - Determination of Relative Free-Energy by Computer-Simulation. *J. Am. Chem. Soc.* **1985**, *107*, 7793-7794.
109. Lybrand, T. P.; Mccammon, J. A.; Wipff, G., Theoretical Calculation of Relative Binding-Affinity in Host Guest Systems. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 833-835.
110. Merz, K. M., Jr., Limits of Free Energy Computation for Protein-Ligand Interactions. *J. Chem. Theory. Comput.* **2010**, *6*, 1018-1027.
111. Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, K. M., Jr., Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-ligand Complexes. *J. Chem. Theory. Comput.* **2011**, *7*, 790-797.

112. Faver, J. C.; Merz, K. M., Jr., Fragment-based error estimation in biomolecular modeling. *Drug Discov. Today* **2014**, *19*, 45-50.
113. Faver, J. C.; Yang, W.; Merz, K. M., Jr., The Effects of Computational Modeling Errors on the Estimation of Statistical Mechanical Variables. *J. Chem. Theory. Comput.* **2012**, *8*, 3769-3776.
114. Cieplak, P.; Kollman, P. A., Calculation of the Free-Energy of Association of Nucleic-Acid Bases in Vacuo and Water Solution. *J. Am. Chem. Soc.* **1988**, *110*, 3734-3739.
115. Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tiradorives, J., Efficient Computation of Absolute Free-Energies of Binding by Computer-Simulations - Application to the Methane Dimer in Water. *J. Chem. Phys.* **1988**, *89*, 3742-3746.
116. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M., Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535-9551.
117. Wong, C. F.; Mccammon, J. A., Dynamics and Design of Enzymes and Inhibitors. *J. Am. Chem. Soc.* **1986**, *108*, 3830-3832.
118. Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A., Calculation of the Relative Change in Binding Free-Energy of a Protein-Inhibitor Complex. *Science* **1987**, *235*, 574-576.
119. Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A., Free-Energy Perturbation Calculations on Binding and Catalysis after Mutating Asn-155 in Subtilisin. *Nature* **1987**, *328*, 551-554.
120. Lipkowitz, K. B.; Boyd, D. B., Reviews in computational chemistry - Volume 16. *Reviews in Computational Chemistry, Vol 16* **2000**, *16*, V-Viii.
121. Kollman, P., Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395-2417.
122. Pierce, A. C.; Jorgensen, W. L., Computational binding studies of orthogonal cyclosporin-cyclophilin pairs. *Angewandte Chemie-International Edition in English* **1997**, *36*, 1466-1469.
123. Essex, J. W.; Severance, D. L.; TiradoRives, J.; Jorgensen, W. L., Monte Carlo simulations for proteins: Binding affinities for trypsin-benzamidine complexes via free-energy perturbations. *J. Phys. Chem. B* **1997**, *101*, 9663-9669.

124. Ferguson, D. M.; Radmer, R. J.; Kollman, P. A., Determination of the relative binding free energies of peptide inhibitors to the HIV-1 protease. *J. Med. Chem.* **1991**, *34*, 2654-9.
125. Kudalkar, S. N.; Beloor, J.; Chan, A. H.; Lee, W. G.; Jorgensen, W. L.; Kumar, P.; Anderson, K. S., Structural and Preclinical Studies of Computationally Designed Non-Nucleoside Reverse Transcriptase Inhibitors for Treating HIV infection. *Mol. Pharmacol.* **2017**, *91*, 383-U168.
126. Jorgensen, W. L., Computer-aided discovery of anti-HIV agents. *Bioorg. Med. Chem.* **2016**, *24*, 4768-4778.
127. Lee, W. G.; Gallardo-Macias, R.; Frey, K. M.; Spasov, K. A.; Bollini, M.; Anderson, K. S.; Jorgensen, W. L., Picomolar Inhibitors of HIV Reverse Transcriptase Featuring Bicyclic Replacement of a Cyanovinylphenyl Group. *J. Am. Chem. Soc.* **2013**, *135*, 16705-16713.
128. Bollini, M.; Gallardo-Macias, R.; Spasov, K. A.; Tirado-Rives, J.; Anderson, K. S.; Jorgensen, W. L., Optimization of benzyloxazoles as non-nucleoside inhibitors of HIV-1 reverse transcriptase to enhance Y181C potency. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 1110-1113.
129. Bollini, M.; Domoal, R. A.; Thakur, V. V.; Gallardo-Macias, R.; Spasov, K. A.; Anderson, K. S.; Jorgensen, W. L., Computationally-Guided Optimization of a Docking Hit to Yield Catechol Diethers as Potent Anti-HIV Agents. *J. Med. Chem.* **2011**, *54*, 8582-8591.
130. Jorgensen, W. L.; Bollini, M.; Thakur, V. V.; Domoal, R. A.; Spasov, K. A.; Anderson, K. S., Efficient Discovery of Potent Anti-HIV Agents Targeting the Tyr181Cys Variant of HIV Reverse Transcriptase. *J. Am. Chem. Soc.* **2011**, *133*, 15686-15696.
131. Thakur, V. V.; Kim, J. T.; Hamilton, A. D.; Bailey, C. M.; Domoal, R. A.; Wang, L. G.; Anderson, K. S.; Jorgensen, W. L., Optimization of pyrimidinyl- and triazinyl-amines as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5664-5667.
132. Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K. S.; Hamilton, A. D., Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 663-667.
133. Kim, J. T.; Hamilton, A. D.; Bailey, C. M.; Domoal, R. A.; Wang, L. G.; Anderson, K. S.; Jorgensen, W. L., FEP-guided selection of bicyclic heterocycles in lead optimization for non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J. Am. Chem. Soc.* **2006**, *128*, 15372-15373.
134. Ruiz-Caro, J.; Basavapathruni, A.; Kim, J. T.; Bailey, C. M.; Wang, L. G.; Anderson, K. S.; Hamilton, A. D.; Jorgensen, W. L., Optimization of diarylamines as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 668-671.

135. Zeevaart, J. G.; Wang, L. G.; Thakur, V. V.; Leung, C. S.; Tirado-Rives, J.; Bailey, C. M.; Domaoal, R. A.; Anderson, K. S.; Jorgensen, W. L., Optimization of azoles as anti-human immunodeficiency virus agents guided by free-energy calculations. *J. Am. Chem. Soc.* **2008**, *130*, 9492-9499.
136. Barreiro, G.; Kim, J. T.; Guimares, C. R. W.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L., From docking false-positive to active Anti-HIV agent. *J. Med. Chem.* **2007**, *50*, 5324-5329.
137. Leung, C. S.; Zeevaart, J. G.; Domaoal, R. A.; Bollini, M.; Thakur, V. V.; Spasov, K. A.; Anderson, K. S.; Jorgensen, W. L., Eastern extension of azoles as non-nucleoside inhibitors of HIV-1 reverse transcriptase; cyano group alternatives. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2485-2488.
138. Nichols, S. E.; Domaoal, R. A.; Thakur, V. V.; Tirado-Rives, J.; Anderson, K. S.; Jorgensen, W. L., Discovery of Wild-Type and Y181C Mutant Non-nucleoside HIV-1 Reverse Transcriptase Inhibitors Using Virtual Screening with Multiple Protein Structures. *J. Chem. Inf. Model.* **2009**, *49*, 1272-1279.
139. Gray, W. T.; Frey, K. M.; Laskey, S. B.; Mislak, A. C.; Spasov, K. A.; Lee, W. G.; Bollini, M.; Siliciano, R. F.; Jorgensen, W. L.; Anderson, K. S., Potent Inhibitors Active against HIV Reverse Transcriptase with K101P, a Mutation Conferring Rilpivirine Resistance. *ACS Med. Chem. Lett.* **2015**, *6*, 1075-1079.
140. Lovering, F.; Aevazelis, C.; Chang, J.; Dehnhardt, C.; Fitz, L.; Han, S.; Janz, K.; Lee, J.; Kaila, N.; McDonald, J.; Moore, W.; Moretto, A.; Papaioannou, N.; Richard, D.; Ryan, M. S.; Wan, Z. K.; Thorarensen, A., Imidazotriazines: Spleen Tyrosine Kinase (Syk) Inhibitors Identified by Free-Energy Perturbation (FEP). *Chemmedchem* **2016**, *11*, 217-233.
141. Keranen, H.; Perez-Benito, L.; Ciordia, M.; Delgado, F.; Steinbrecher, T. B.; Oehlich, D.; van Vlijmen, H. W. T.; Trabanco, A. A.; Tresadern, G., Acylguanidine Beta Secretase 1 Inhibitors: A Combined Experimental and Free Energy Perturbation Study. *J. Chem. Theory. Comput.* **2017**, *13*, 1439-1453.
142. Hukushima, K.; Nemoto, K., Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604-1608.
143. Wang, L.; Friesner, R. A.; Berne, B. J., Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2) (vol 115, pg 9431, 2011). *J. Phys. Chem. B* **2011**, *115*, 11305-11305.
144. Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J., Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13749-13754.

145. Wang, L.; Berne, B. J.; Friesner, R. A., On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1937-1942.
146. Lee, T. S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M., GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58*, 2043-2050.
147. Lee, T. S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M., Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theory. Comput.* **2017**, *13*, 3077-3084.
148. Bergdorf, M. K., E. T.; Rendleman, C. A.; Shaw, D. E. , Desmond GPU Performance as of November 2014; Technical Report, DESRESTR-2014-01. **2014**.
149. Song, L. F.; Lee, T. S.; Zhu, C.; York, D. M.; Merz, K. M., Using AMBER18 for Relative Free Energy Calculations. *J. Chem. Inf. Model.* **2019**, *59*, 3128-3135.
150. Gapsys, V.; Perez-Benito, L.; Aldeghi, M.; Seeliger, D.; Van Vlijmen, H.; Tresadern, G.; de Groot, B. L., Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci.* **2020**, *11*, 1140-1152.
151. He, X.; Liu, S.; Lee, T. S.; Ji, B.; Man, V. H.; York, D. M.; Wang, J., Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein-Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **2020**, *5*, 4611-4619.
152. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
153. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D., CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671-690.
154. Harder, E.; Damm, W.; Maple, J.; Wu, C. J.; Reboul, M.; Xiang, J. Y.; Wang, L. L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory. Comput.* **2016**, *12*, 281-296.
155. Roos, K.; Wu, C. J.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L. L.; Abel, R.; Friesner, R. A.; Harder, E. D., OPLS3e: Extending

Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory. Comput.* **2019**, *15*, 1863-1874.

156. Wang, J., A Snapshot of GAFF2 Development.

157. Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K., Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory. Comput.* **2018**, *14*, 6076-6092.

158. Liu, S.; Wu, Y. J.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L., Lead optimization mapper: automating free energy calculations for lead optimization. *J. Comput. Aided Mol. Des.* **2013**, *27*, 755-770.

159. [https://www.cresset-group.com/media/uploads/files/Automated Assessment of Binding Affinity via Free Energy Perturbation.pdf](https://www.cresset-group.com/media/uploads/files/Automated_Assessment_of_Binding_Affinity_via_Free_Energy_Perturbation.pdf).

160. Yu, H. B.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D.; Roux, B., Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory. Comput.* **2010**, *6*, 774-786.

161. Lamoureux, G.; Orabi, E. A., Molecular modelling of cation-pi interactions. *Mol. Simulat.* **2012**, *38*, 704-722.

162. Patel, S.; Brooks, C. L., Fluctuating charge force fields: Recent developments and applications from small molecules to macromolecular biological systems. *Mol. Simulat.* **2006**, *32*, 231-249.

163. Russo, M. F.; van Duin, A. C. T., Atomistic-scale simulations of chemical reactions: Bridging from quantum chemistry to engineering. *Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms* **2011**, *269*, 1549-1554.

164. Li, P.; Merz, K. M., Jr., Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117*, 1564-1686.

165. Li, P.; Merz, K. M., Jr., MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56*, 599-604.

166. Kottalam, J.; Case, D. A., Dynamics of Ligand Escape from the Heme Pocket of Myoglobin. *J. Am. Chem. Soc.* **1988**, *110*, 7690-7697.

167. Case, D. A.; McCammon, J. A., Dynamic simulations of oxygen binding to myoglobin. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 222-33.
168. Case, D. A.; Karplus, M., Dynamics of ligand binding to heme proteins. *J. Mol. Biol.* **1979**, *132*, 343-68.
169. Gelin, B. R.; Karplus, M., Mechanism of tertiary structural change in hemoglobin. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 801-5.
170. Hoops, S. C.; Anderson, K. W.; Merz, K. M., Force field design for metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262-8270.
171. Aqvist, J.; Warshel, A., Computer simulation of the initial proton transfer step in human carbonic anhydrase I. *J. Mol. Biol.* **1992**, *224*, 7-14.
172. Aqvist, J.; Warshel, A., Free-Energy Relationships in Metalloenzyme-Catalyzed Reactions - Calculations of the Effects of Metal-Ion Substitutions in Staphylococcal Nuclease. *J. Am. Chem. Soc.* **1990**, *112*, 2860-2868.
173. Jiang, Y.; Zhang, H.; Tan, T., Rational Design of Methodology-Independent Metal Parameters Using a Nonbonded Dummy Model. *J. Chem. Theory. Comput.* **2016**, *12*, 3250-60.
174. Jiang, Y.; Zhang, H. Y.; Feng, W.; Tan, T. W., Refined Dummy Atom Model of Mg²⁺ by Simple Parameter Screening Strategy with Revised Experimental Solvation Free Energy. *J. Chem. Inf. Model.* **2015**, *55*, 2575-2586.
175. Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Aqvist, J.; Kamerlin, S. C. L., Force Field Independent Metal Parameters Using a Nonbonded Dummy Model. *J. Phys. Chem. B* **2014**, *118*, 4351-4362.
176. Masetti, M.; Musiani, F.; Bernetti, M.; Falchi, F.; Cavalli, A.; Ciurli, S.; Recanatini, M., Development of a multisite model for Ni(II) ion in solution from thermodynamic and kinetic data. *J. Comput. Chem.* **2017**, *38*, 1834-1843.
177. Liao, Q. H.; Kamerlin, S. C. L.; Strodel, B., Development and Application of a Nonbonded Cu²⁺ Model That Includes the Jahn-Teller Effect. *J. Phys. Chem. Lett.* **2015**, *6*, 2657-2662.
178. Lu, S. Y.; Huang, Z. M.; Huang, W. K.; Liu, X. Y.; Chen, Y. Y.; Shi, T.; Zhang, J., How calcium inhibits the magnesium-dependent kinase GSK3 beta: A molecular simulation study. *Proteins-Structure Function and Bioinformatics* **2013**, *81*, 740-753.

179. Hirschfelder, J. O.; Ewell, R. B.; Roebuck, J. R., Determination of intermolecular forces from the Joule-Thomson coefficients. *J. Chem. Phys.* **1938**, *6*, 205-218.
180. Born, M.; Mayer, J. E., For the Lattice theory of Ionic crystals. *Physica B.* **1932**, *75*, 1-18.
181. Li, P.; Song, L. F.; Merz, K. M., Jr., Parameterization of highly charged metal ions using the 12-6-4 LJ-type nonbonded model in explicit water. *J. Phys. Chem. B* **2015**, *119*, 883-95.
182. Marzinek, J. K.; Bond, P. J.; Lian, G.; Zhao, Y.; Han, L.; Noro, M. G.; Pistikopoulos, E. N.; Mantalaris, A., Free energy predictions of ligand binding to an alpha-helix using steered molecular dynamics and umbrella sampling simulations. *J. Chem. Inf. Model.* **2014**, *54*, 2093-104.
183. Panteva, M. T.; Giambasu, G. M.; York, D. M., Force Field for Mg²⁺, Mn²⁺, Zn²⁺, and Cd²⁺ Ions That Have Balanced Interactions with Nucleic Acids. *J. Phys. Chem. B* **2015**, *119*, 15460-15470.
184. Liao, Q.; Pabis, A.; Strodel, B.; Kamerlin, S. C. L., Extending the Nonbonded Cationic Dummy Model to Account for Ion-Induced Dipole Interactions. *J. Phys. Chem. Lett.* **2017**, *8*, 5408-5414.
185. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177-6196.
186. 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.; Shenkin, P. S., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739-1749.
187. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750-1759.
188. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623-1641.
189. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132-146.

190. D.A. Case, R. M. B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,; A.W. Goetz, N. H., S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C.; Lin, T. L., R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I.; Omelyan, A. O., D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails,; R.C. Walker, J. W., R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman, AMBER 2016, University of California, San Francisco. **2016**.
191. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
192. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
193. Biedermann, F.; Scherman, O. A., Cucurbit[8]uril Mediated Donor-Acceptor Ternary Complexes: A Model System for Studying Charge-Transfer Interactions. *J. Phys. Chem. B* **2012**, *116*, 2842-2849.
194. Liu, S. M.; Ruspic, C.; Mukhopadhyay, P.; Chakrabarti, S.; Zavalij, P. Y.; Isaacs, L., The cucurbit[n]uril family: Prime components for self-sorting systems. *J. Am. Chem. Soc.* **2005**, *127*, 15959-15967.
195. Faver, J. C.; Zheng, Z.; Merz, K. M., Jr., Model for the fast estimation of basis set superposition error in biomolecular systems. *J. Chem. Phys.* **2011**, *135*, 144110.
196. D.A. Case, V. B., J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E.; Duke, H. G., A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko,; T.S. Lee, S. L., T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui,; R. Salomon-Ferrer, G. S., C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X.; Kollman, W. a. P. A., AMBER 14, University of California, San Francisco. **2014**.
197. M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09 (Gaussian, Inc., Wallingford CT, 2009).

198. Joung, I. S.; Cheatham, T. E., Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020-9041.
199. D.A. Case, R. M. B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,; A.W. Goetz, N. H., S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C.; Lin, T. L., R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I.; Omelyan, A. O., D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails,; R.C. Walker, J. W., R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman (2016), , AMBER 2016, University of California, San Francisco.
200. Ucisik, M. N.; Hammes-Schiffer, S., Relative Binding Free Energies of Adenine and Guanine to Damaged and Undamaged DNA in Human DNA Polymerase ϵ : Clues for Fidelity and Overall Efficiency. *J. Am. Chem. Soc.* **2015**, *137*, 13240-13243.
201. Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammon, J. A., Improving the Efficiency of Free Energy Calculations in the Amber Molecular Dynamics Package. *J. Chem. Theory. Comput.* **2013**, *9*, 4131-4139.
202. Steinbrecher, T.; Joung, I.; Case, D. A., Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.* **2011**, *32*, 3253-3263.
203. Steinbrecher, T.; Mobley, D. L.; Case, D. A., Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.* **2007**, *127*.
204. Wang, L. L.; Deng, Y. Q.; Knight, J. L.; Wu, Y. J.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R., Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. *J. Chem. Theory. Comput.* **2013**, *9*, 1282-1293.
205. Loeffler, H. H.; Bosisio, S.; Duarte Ramos Matos, G.; Suh, D.; Roux, B.; Mobley, D. L.; Michel, J., Reproducibility of Free Energy Calculations across Different Molecular Simulation Software Packages. *J. Chem. Theory. Comput.* **2018**, *14*, 5567-5582.
206. Prier, C. K.; Arnold, F. H., Chemomimetic biocatalysis: exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts. *J. Am. Chem. Soc.* **2015**, *137*, 13992-4006.
207. Hammer, S. C.; Kubik, G.; Watkins, E.; Huang, S.; Mingos, H.; Arnold, F. H., Anti-Markovnikov alkene oxidation by metal-oxo-mediated enzyme catalysis. *Science* **2017**, *358*, 215-218.

208. Handel, T. M.; Williams, S. A.; DeGrado, W. F., Metal ion-dependent modulation of the dynamics of a designed protein. *Science* **1993**, *261*, 879.
209. Steel, E. A.; Merz, K. M.; Selinger, A.; Castleman, A. W., Mass spectral and computational free energy studies of alkali metal ion-containing water clusters. *J. Phys. Chem.* **1995**, *99*, 7829-7836.
210. Cook, T. R.; Zheng, Y. R.; Stang, P. J., Metal-organic frameworks and self-assembled supramolecular coordination complexes: comparing and contrasting the design, synthesis, and functionality of metal-organic materials. *Chem. Rev.* **2013**, *113*, 734-77.
211. Wang, Q.; Astruc, D., State of the Art and Prospects in Metal-Organic Framework (MOF)-Based and MOF-Derived Nanocatalysis. *Chem. Rev.* **2019**.
212. Bignozzi, C. A.; Argazzi, R.; Boaretto, R.; Busatto, E.; Carli, S.; Ronconi, F.; Caramori, S., The role of transition metal complexes in dye sensitized solar devices. *Coord. Chem. Rev.* **2013**, *257*, 1472-1492.
213. Osmieri, L.; Escudero-Cid, R.; Armandi, M.; Ocon, P.; Videla, A. H. A. M.; Specchia, S., Effects of using two transition metals in the synthesis of non-noble electrocatalysts for oxygen reduction reaction in direct methanol fuel cell. *Electrochim. Acta* **2018**, *266*, 220-232.
214. Fernandez-Moreira, V.; Thorp-Greenwood, F. L.; Coogan, M. P., Application of d6 transition metal complexes in fluorescence cell imaging. *Chem. Commun. (Camb.)* **2010**, *46*, 186-202.
215. Firman, T. K.; Landis, C. R., Valence bond concepts applied to the molecular mechanics description of molecular shapes. 4. Transition metals with pi-bonds. *J. Am. Chem. Soc.* **2001**, *123*, 11728-11742.
216. Landis, C. R.; Cleveland, T.; Firman, T. K., Valence bond concepts applied to the molecular mechanics description of molecular shapes. 3. Applications to transition metal alkyls and hydrides. *J. Am. Chem. Soc.* **1998**, *120*, 2641-2649.
217. Nielson, K. D.; van Duin, A. C. T.; Oxgaard, J.; Deng, W. Q.; Goddard, W. A., Development of the ReaxFF reactive force field for describing transition metal catalyzed reactions, with application to the initial stages of the catalytic formation of carbon nanotubes. *J. Phys. Chem. A* **2005**, *109*, 493-499.
218. Comba, P.; Martin, B.; Sanyal, A., An efficient fluctuating charge model for transition metal complexes. *J. Comput. Chem.* **2013**, *34*, 1598-1608.

219. Yang, Z. Z.; Wang, J. J.; Zhao, D. X., Valence State Parameters of All Transition Metal Atoms in Metalloproteins-Development of ABEEM sigma pi Fluctuating Charge Force Field. *J. Comput. Chem.* **2014**, *35*, 1690-1706.
220. Li, P. F.; Merz, K. M., Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117*, 1564-1686.
221. Keskin, S.; Liu, J.; Rankin, R. B.; Johnson, J. K.; Sholl, D. S., Progress, Opportunities, and Challenges for Applying Atomically Detailed Modeling to Molecular Adsorption and Transport in Metal-Organic Framework Materials. *Industrial & Engineering Chemistry Research* **2009**, *48*, 2355-2371.
222. Flood, E.; Boiteux, C.; Lev, B.; Vorobyov, I.; Allen, T. W., Atomistic Simulations of Membrane Ion Channel Conduction, Gating, and Modulation. *Chem. Rev.* **2019**, *119*, 7737-7832.
223. Cieplak, P.; Lybrand, T. P.; Kollman, P. A., Calculation of free energy changes in ion-water clusters using nonadditive potentials and the Monte Carlo method. *J. Chem. Phys.* **1987**, *86*, 6393-6403.
224. Lybrand, T. P.; Ghosh, I.; McCammon, J. A., Hydration of chloride and bromide anions: determination of relative free energy by computer simulation. *J. Am. Chem. Soc.* **1985**, *107*, 7793-7794.
225. Marrone, T. J.; Merz, K. M., Molecular recognition of potassium ion by the naturally occurring antibiotic ionophore nonactin. *J. Am. Chem. Soc.* **1992**, *114*, 7542-7549.
226. Miyamoto, S.; Kollman, P. A., Molecular dynamics studies of calixspherand complexes with alkali metal cations: calculation of the absolute and relative free energies of binding of cations to a calixspherand. *J. Am. Chem. Soc.* **1992**, *114*, 3668-3674.
227. Baştuğ, T.; Kuyucak, S., Energetics of Ion Permeation, Rejection, Binding, and Block in Gramicidin A from Free Energy Simulations. *Biophys. J.* **2006**, *90*, 3941-3950.
228. Aqvist, J.; Fothergill, M.; Warshel, A., Computer simulation of the carbon dioxide/bicarbonate interconversion step in human carbonic anhydrase I. *J. Am. Chem. Soc.* **1993**, *115*, 631-635.
229. Dudev, T.; Lim, C., Competition among Metal Ions for Protein Binding Sites: Determinants of Metal Ion Selectivity in Proteins. *Chem. Rev.* **2014**, *114*, 538-556.
230. Rao, L.; Cui, Q.; Xu, X., Electronic Properties and Desolvation Penalties of Metal Ions Plus Protein Electrostatics Dictate the Metal Binding Affinity and Selectivity in the Copper Efflux Regulator. *J. Am. Chem. Soc.* **2010**, *132*, 18092-18102.

231. Nechay, M. R.; Gallup, N. M.; Morgenstern, A.; Smith, Q. A.; Eberhart, M. E.; Alexandrova, A. N., Histone Deacetylase 8: Characterization of Physiological Divalent Metal Catalysis. *J. Chem. Phys. B.* **2016**, *120*, 5884-5895.
232. Tsui, V.; Case, D. A., Calculations of the Absolute Free Energies of Binding between RNA and Metal Ions Using Molecular Dynamics Simulations and Continuum Electrostatics. *J. Chem. Phys. B.* **2001**, *105*, 11314-11325.
233. Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P., Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand–Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory. Comput.* **2007**, *3*, 1960-1986.
234. Hermida-Ramón, J. M.; Brdarski, S.; Karlström, G.; Berg, U., Inter- and intramolecular potential for the N-formylglycinamide-water system. A comparison between theoretical modeling and empirical force fields. *J. Comput. Chem.* **2003**, *24*, 161-176.
235. Jing, Z.; Liu, C.; Qi, R.; Ren, P., Many-body effect determines the selectivity for Ca²⁺ and Mg²⁺ in proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E7495-E7501.
236. Macchiagodena, M.; Pagliai, M.; Andreini, C.; Rosato, A.; Procacci, P., Upgrading and Validation of the AMBER Force Field for Histidine and Cysteine Zinc(II)-Binding Residues in Sites with Four Protein Ligands. *J. Chem. Inf. Model.* **2019**, *59*, 3803-3816.
237. Li, P.; Merz, K. M., Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory. Comput.* **2014**, *10*, 289-297.
238. Liao, Q. H.; Pabis, A.; Strodel, B.; Kamerlin, S. C. L., Extending the Nonbonded Cationic Dummy Model to Account for Ion-Induced Dipole Interactions. *J. Phys. Chem. Lett.* **2017**, *8*, 5408-5414.
239. Parkin, G., Synthetic analogues relevant to the structure and function of zinc enzymes. *Chem. Rev.* **2004**, *104*, 699-767.
240. Holm, R. H.; Kennepohl, P.; Solomon, E. I., Structural and Functional Aspects of Metal Sites in Biology. *Chem. Rev.* **1996**, *96*, 2239-2314.
241. DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A., De novo design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **1999**, *68*, 779-819.
242. Tebo, A. G.; Pecoraro, V. L., Artificial metalloenzymes derived from three-helix bundles. *Curr. Opin. Chem. Biol.* **2015**, *25*, 65-70.

243. Petrik, I. D.; Liu, J.; Lu, Y., Metalloenzyme design and engineering through strategic modifications of native protein scaffolds. *Curr. Opin. Chem. Biol.* **2014**, *19*, 67-75.
244. Bou-Abdallah, F.; Giffune, T. R., The thermodynamics of protein interactions with essential first row transition metals. *Biochim. Biophys. Acta.* **2016**, *1860*, 879-891.
245. Clugston, S. L.; Yajima, R.; Honek, J. F., Investigation of metal binding and activation of Escherichia coli glyoxalase I: kinetic, thermodynamic and mutagenesis studies. *Biochem. J* **2004**, *377*, 309-316.
246. He, M. M.; Clugston, S. L.; Honek, J. F.; Matthews, B. W., Determination of the structure of Escherichia coli glyoxalase I suggests a structural basis for differential metal activation. *Biochemistry-Us* **2000**, *39*, 8719-8727.
247. Sjoberg, S., Critical evaluation of stability constants of metal-imidazole and metal-histamine systems. *Pure Appl. Chem.* **1997**, *69*, 1549-1570.
248. Bunting, J. W.; Thong, K. M., Stability Constants for Some 1-1 Metal-Carboxylate Complexes. *Can. J. Chem.* **1970**, *48*, 1654-&.
249. Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221-234.
250. D.A. Case, I. Y. B.-S., S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden,; R.E. Duke, D. G., M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang,; S. Izadi, A. K., T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J.; Mermelstein, K. M. M., Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R.; Qi, D. R. R., A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman, AMBER 2018, University of California, San Francisco. . **2018**.
251. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory. Comput.* **2015**, *11*, 3696-713.
252. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
253. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

254. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics & Modelling* **1996**, *14*, 33-38.
255. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M., Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Chem. Phys. B.* **2003**, *107*, 9535-9551.
256. Golan, Y., Alhadeff, R., Warshel, A., & Assaraf, Y. G., ZnT2 is an electroneutral proton-coupled vesicular antiporter displaying an apparent stoichiometry of two protons per zinc ion. *PLoS Comput. Biol.* **2019**, *15*(3), e1006882.
257. Simonson, T.; Carlsson, J.; Case, D. A., Proton binding to proteins: pK(a) calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167-4180.
258. Walba, H.; Isensee, R. W., Acidity Constants of Some Arylimidazoles and Their Cations. *J. Org. Chem.* **1961**, *26*, 2789-&.