

COMPUTATIONAL DRUG DESIGN AND DISCOVERY FOR *TRYPANOSOMA cruzi*
TRANS-SIALIDASE

By

BILL MILLER III

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2013

© 2013 Bill Miller III

To my parents, Robin and Bill Jr, my wife, Ryan, and children, Rylee and Billy

ACKNOWLEDGEMENTS

The work I present in this dissertation would not have been possible without the help, support, and encouragement of many people. My advisor, Dr. Adrian Roitberg, gave me the opportunity to succeed in his lab and gave me a project that was a perfect fit for me. He allowed me to become a better scientist by guiding me and always giving great advice. More importantly, I have to thank Dr. Roitberg for being very lenient and accommodating with my personal needs, allowing me to have a flexible work schedule so that I could spend more time with my children. Becoming a father twice while in graduate school is not ideal for most doctoral students, but Dr. Roitberg gave me the opportunity to not put my personal plans on hold and allowed me the chance to more comfortably balance life as a Ph.D. student and life as a father and husband. Over the last five years, Dr. Roitberg has helped me develop professionally, while at the same time allowing me to spend quality time with the people that are most important in my life.

Many other University of Florida (UF) faculty have also been instrumental in helping me along the way. I want to thank Dr. Kennie Merz for being a great scientific collaborator that was always pushing me to do better—and for being available to talk sports. I would like to thank Dr. Erik Deumens for being a great resource as a computer expert. I also want to thank Dr. Nicole Horenstein for her patience in helping teach me about the process of making protein and running kinetic assays in a biochemistry lab. I acknowledge the rest of my committee, Dr. Gail Fanucci and Dr. Art Edison, for being resources to me for my oral defense all the way through my doctoral dissertation.

My early transition from undergraduate to graduate school was aided by many people in the Quantum Theory Project (QTP), including Dr. Dan Sindhikara, Dr. Özlem Demir, and Dr. Gustavo Seabra. They helped me become acclimated to conducting

graduate-level research in the Roitberg group. I also could not have survived graduate school without all the help provided by the QTP staff, especially Judy Parker and Coralu Clements—they were always willing to help me with any problems.

My life in graduate school would not have been complete without so many graduate students and post-docs at UF helping me along with way, both professionally and personally. I want to give a special thanks to Dwight McGee and Jason Swails for with me since the beginning. I want to thank Johan Cruz, Dr. Juan Bueren Calabuig, Natali di Russo, Matias Pascualini, and Shuai Liang for many great memories in the lab. Dr. Thomas Watson helped recruit me to UF and has been a great friend ever since.

Finally, I want to acknowledge the help and support of my family. I attribute my intelligence and work ethic from my mom and dad, Robin and Bill Jr. They have been nothing but supportive of me despite the fact that I moved halfway across the country from them. My children, Rylee (3 years) and Billy IV (6 months), are the best kids any father could ask for. Thanks to them, I have enjoyed nearly every minute that I have spent away from work for the past three years. My final acknowledgement goes to my loving wife, Ryan. She agreed to pick up and move away from her family so I could go to school. She has allowed me the time I needed to graduate so that I can make our family move across the country once again. She has proofread papers and posters, and listened to practice presentations—all to help me—and I could not thank her enough for that.

TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGEMENTS	4
LIST OF TABLES.....	9
LIST OF FIGURES.....	10
LIST OF ABBREVIATIONS.....	14
ABSTRACT	15
CHAPTER	
1 INTRODUCTION	17
1.1 Prologue	17
1.2 Chagas' Disease.....	17
1.3 <i>Trypanosoma cruzi</i>	21
1.4 <i>Trypanosoma cruzi</i> Trans-sialidase	24
1.4.1 TcTS Structure.....	25
1.4.2 TcTS Reaction Mechanism.....	29
1.4.3 TcTS Dynamics	34
1.4.4 TcTS Drug Design and Discovery Efforts	37
2 METHODS.....	42
2.1 Molecular Structure and Function	42
2.2 Molecular Docking	42
2.2.1 Receptor Grid Generation.....	45
2.2.2 Generating Ligand Conformations.....	47
2.2.3 Search Algorithms	48
2.2.4 Scoring Functions.....	52
2.3 Pharmacophore Models.....	60
2.4 Molecular Dynamics.....	64
2.5 Binding Free Energy Calculations.....	70
3 DESIGN OF e-PHARMACOPHORE MODELS USING COMPOUND FRAGMENTS FOR the <i>TRANS</i> -SIALIDASE of <i>TRYPANOSOMA cruzi</i> : SCREENING FOR NOVEL INHIBITOR SCAFFOLDS	71
3.1 Preliminaries	71
3.2 Methods	72
3.2.1 Receptor Structures and Grid Generation	72
3.2.2 Docking Fragments to TcTS Structures.....	74
3.2.3 E-pharmacophore Generation	75

3.2.4	Pharmacophore Screening	75
3.2.5	Molecular Dynamics and Free Energy Calculations	76
3.3	Results and Discussion.....	78
3.3.1	Pharmacophore Hypotheses	78
3.3.1.1	1S0I pharmacophore model.....	78
3.3.1.2	Holo pharmacophore model.....	81
3.3.1.3	Apo pharmacophore model.....	84
3.3.2	Screening Results Using TcTS Pharmacophore Models	87
3.3.2.1	Screening results for the 1S0I pharmacophore.....	87
3.3.2.2	Screening results for the holo pharmacophore	91
3.3.2.3	Screening results for the apo pharmacophore	93
3.3.3	Analysis of the Most Promising TcTS Inhibitors.....	96
3.3.3.1	ZINC13359679	96
3.3.3.2	ZINC02576132	100
3.4	Concluding Remarks.....	104
4	COMBINED MOLECULAR DOCKING, MOLECULAR DYNAMICS AND BINDING FREE ENERGY CALCULATIONS TO DESIGN AND DISCOVER NOVEL INHIBITORS FOR <i>TRYPANOSOMA cruzi</i> TRANS-SIALIDASE.....	106
4.1	Preliminaries	106
4.2	Methods	107
4.2.1	Acquisition of Compounds	108
4.2.1.1	Molecular libraries.....	108
4.2.1.2	<i>De novo</i> design.....	109
4.2.1.3	Rational drug design.....	109
4.2.2	Docking Potential Ligands with Glide.....	109
4.2.2.1	Receptor grid generation	109
4.2.2.2	Molecular docking.....	110
4.2.3	<i>De novo</i> Drug Design	110
4.2.4	Molecular Dynamics	111
4.2.5	Free Energies of Binding	111
4.3	Results.....	111
4.4	Discussion	116
4.4.1	Best Potential Inhibitors	117
4.4.1.1	DANA Result 034.....	117
4.4.1.2	Oseltamivir Result 70.....	122
4.4.1.3	PubChem 54353125.....	125
4.4.1.4	ZINC16247520 Rank 10	129
4.4.1.5	Sialic Acid Rank 1	135
4.4.1.6	ZINC04939766	137
4.4.1.7	Zanamivir phosphonate	142
4.4.2	<i>De novo</i> Design Candidate Combining ZINC Fragments	147
4.5	Concluding Remarks.....	149

5	<i>MMPBSA.py</i> : AN EFFICIENT PROGRAM FOR END-STATE FREE ENERGY CALCULATIONS	151
5.1	Preliminaries	151
5.2	Theory and Methods	152
5.2.1	Stability and Binding Free Energy Calculations	153
5.2.2	Free Energy Decomposition	155
5.2.3	Entropy Calculations.....	156
5.3	Results and Discussion.....	157
5.3.1	General Workflow	157
5.3.2	Poisson-Boltzmann: MM-PBSA	160
5.3.3	Generalized Born: MM-GBSA.....	161
5.3.4	Reference Interaction Site Model: MM/3D-RISM.....	163
5.3.5	Free Energy Decomposition	164
5.3.6	Alanine Scanning.....	165
5.3.7	Solute Entropy	166
5.3.7.1	Normal Mode Analysis.....	166
5.3.7.2	Quai-harmonic Approximation	167
5.3.8	Running in Parallel.....	168
5.3.9	Differences to mm_pbsa.pl	169
5.4	Concluding Remarks.....	170
6	CONCLUSIONS	172
	APPENDIX: SUPPLEMENTARY MATERIAL.....	175
A.1	Hit Compounds From e-pharmacophore Screening	175
A.2	Potential Inhibitors of TcTS From Molecular Docking and <i>de novo</i> Design	191
	LIST OF REFERENCES	210
	BIOGRAPHICAL SKETCH.....	224

LIST OF TABLES

<u>Table</u>	<u>Page</u>
3-1 Pharmacophore sites of the 1S0I model.....	79
3-2 Pharmacophore sites of the TcTS holo model.....	82
3-3 Pharmacophore sites of the TcTS apo model.	85
3-4 The best five compounds from screening results of the 1S0I pharmacophore according to MM-GBSA binding free energies.	88
3-5 The best five hit compounds from screening results of the holo pharmacophore according to MM-GBSA binding free energies..	92
3-6 The best five hit compounds from screening of the TcTS apo pharmacophore according to MM-GBSA binding free energies.	95
4-1 List of ZINC molecular libraries and their corresponding number of compounds used for virtual screening.	108
4-2 The best potential inhibitors of TcTS from virtual screening and <i>de novo</i> drug methods.....	112
A-1 All hit compounds from screening results of the 1S0I pharmacophore according to MM-GBSA binding free energies..	175
A-2 All hit compounds from screening results of the holo pharmacophore according to MM-GBSA binding free energies..	180
A-3 All hit compounds from screening results of the apo pharmacophore according to MM-GBSA binding free energies..	187
A-4 Compounds from molecular docking and de novo design considered as potential inhibitors of TcTS.	191

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1-1	Romana's sign, a typical sign of Chagas' disease, occurs on the eyelid after being bitten by an infected Triatomine bug.....	19
1-2	Structures of drugs currently available as treatments of Chagas' disease. Specifically, a) benznidazole and b) nifurtimox	20
1-3	Life cycle of <i>T. cruzi</i> as it progresses through triatomine bugs and humans.	22
1-4	Image of <i>T. cruzi</i> in the trypomastigote form taken from a blood smear of an infected patient.	23
1-5	The general structure of sialic acid, where R=H is neuraminic acid, R=CH ₃ -CO— is N-acetyl neuraminic acid, and R=HOCH ₂ -CO— is N-glycolyl neuraminic acid..	24
1-6	Structure of TcTS showing the three major domains.....	26
1-7	Depictions of the different regions in the TcTS active site with sialyl-lactose bound..	28
1-8	Chemical mechanism of <i>T. cruzi</i> <i>trans</i> -sialidase, showing the Tyr342/Glu230 hydrogen transfer and the nucleophilic attack of Tyr342 on the anomeric carbon of the substrate.....	31
1-9	Conformation of sialyl-lactose at the transition state showing the planarity of the carbocation of sialic acid, as well as the positions of the catalytic residues Tyr342, Glu230, and Asp59.....	32
1-10	Overlay of TcTS structures displaying the flexibility of the Tyr119 (right) and Trp312 (left) flaps.	35
1-11	Hydrogen bonding network of the Trp312 loop (red)	36
2-1	Schematic diagram of common models for a substrate binding to the active site of an enzyme.	44
2-2	Depiction of how Glide scores a hydrogen bond based on the distance between the heavy atom (X) and hydrogen atom (H).	56
2-3	An example of a pharmacophore model.....	61
2-4	Workflow summary of e-pharmacophore generation from docked fragments using Schrödinger.....	63

2-5	The workflow used in this study using a combination of molecular docking, MD, and binding free energy calculations to discovery new potential inhibitors of TcTS.....	70
3-1	The required matches for the 1S0I pharmacophore used for screening lead compounds.....	80
3-2	The required sites for the holo pharmacophore model used for screening lead compounds.....	83
3-3	The required sites for the TcTS apo pharmacophore model used for screening lead compounds.....	86
3-4	The RMSD of TcTS and ZINC13359679 using the initial frame as reference. ...	97
3-5	The MM-GBSA binding free energy of ZINC13359679 bound to TcTS as a function of time for a 10-ns MD simulation.	98
3-6	The protein-ligand interactions between TcTS and ZINC13359679.....	100
3-7	The RMSD of TcTS and ZINC02576132 using the initial frame as reference. .	101
3-8	The binding free energy of ZINC02576132 bound to TcTS as a function of time for a 10-ns MD simulation.....	102
3-9	The main interactions between TcTS and ZINC02576132 observed during a 10-ns MD simulation.....	103
4-1	A plot of MM-GBSA binding free energies compared to Glide XP docking free energies.....	116
4-2	The RMSD of TcTS and DANA Result 034 using the initial frame as reference.	118
4-3	Depiction of the flexibility of the naphthalene region (green) of DANA Result 034 compared to the rest of the ligand.	118
4-4	The MM-GBSA binding free energy of DANA Result 034 bound to TcTS as a function of time for a 50-ns MD simulation.	119
4-5	Modifications made to DANA Result 034.....	120
4-6	The RMSD of TcTS and Oseltamivir Result 70 using the initial frame as reference.	123
4-7	Depiction of the rotation conformational change observed for Oseltamivir Result 070 during MD around 4 ns.....	123

4-8	The MM-GBSA binding free energy of Oseltamivir Result 70 bound to TcTS as a function of time for a 10-ns MD simulation.....	124
4-9	The 2D structure of chEMBL 198898 and PubChem 54353125.....	125
4-10	The RMSD of TcTS and PubChem 54353125 using the initial frame as reference.	126
4-11	The MM-GBSA binding free energy of PubChem 54353125 bound to TcTS as a function of time for a 110-ns MD simulation.....	126
4-12	Protein-ligand interactions between TcTS and PubChem 54353125.	127
4-13	Representations of the conformational changes for two TcTS residues during MD of the TcTS-PubChem 54353125 complex.	128
4-14	The 2D structures of ZINC16247520 and a derivative.....	130
4-15	Depiction of ZINC16247520 interacting with the arginine triad of TcTS.	131
4-16	The RMSD of TcTS and ZINC16247520 Rank 10 using the initial frame as reference.	132
4-17	The MM-GBSA binding free energy of ZINC16247520 Rank 10 bound to TcTS as a function of time for a 10-ns MD simulation.	132
4-18	The hydrogen bonding network between ZINC16247520 Rank10 and TcTS residues Arg93, Asp59, and Ser118.....	134
4-19	The RMSD of TcTS and Sialic Acid Rank 1 using the initial frame as reference.	135
4-20	Overlay of 250 evenly spaced conformations of Sialic Acid Rank 10 bound to TcTS during a 10-ns MD simulations.....	136
4-21	The RMSD of TcTS and ZINC04939766 using the initial frame as reference. .	139
4-22	The MM-GBSA binding free energy of ZINC04939766 bound to TcTS as a function of time for a 10-ns MD simulation.	139
4-23	The binding mode of ZINC04939766 with the arginine triad of TcTS.	140
4-24	The hydrogen bonding network between ZINC04939766 and the Asp96/Glu230 motif on TcTS.....	141
4-25	The RMSD plots from TcTS-ZP1 and TcTS-ZP62 using the initial frame as reference.	143

4-26	The MM-GBSA binding free energy of ZP1 (blue) and ZP62 (purple) bound to TcTS as a function of time for a 10-ns MD simulation.	143
4-27	The hydrogen bonding network between the phosphonate group of ZP62 and the arginine triad of TcTS.	145
4-28	The RMSD of TcTS and the Super fragment using the initial frame as reference.	148
4-29	The MM-GBSA binding free energy of the Super fragment bound to TcTS as a function of time for a 72-ns MD simulation.....	148
4-30	The position of the Super fragment relative to the TcTS active site at 0 ns (red) and 72 ns (blue) showing the dissociation of the ligand during MD.	149
5-1	Thermodynamic cycles for common free energy calculations.	152
5-2	General workflow for performing end-state calculations with <i>MMPBSA.py</i>	158
5-3	<i>MMPBSA.py</i> scaling comparison for MM-PBSA and MM-GBSA calculations on 200 frames of a 5910-atom complex.	168

LIST OF ABBREVIATIONS

TERM:	Definition
CI	Covalent intermediate
CDC	Centers for Disease Control and Prevention
DANA	2,3-dehydro-3-deoxy-N-acetylneuraminic acid
GA	Genetic algorithm
GB	Generalized Born
MC	Michaelis complex
MD	Molecular dynamics
MM-GBSA	Molecular Mechanics/Generalized Born Surface Area
MM-PBSA	Molecular Mechanics/Poisson Boltzmann Surface Area
MTP	Multiple Trajectory Protocol
NAB	Nucleic Acid Builder
PB	Poisson-Boltzmann
QM	Quantum Mechanics
QM/MM	Quantum-mechanical/molecular-mechanical
RISM	Reference Interaction Site Model
RMSD	Root-mean-square deviation
SAPA	Shed acute phase antigen
SASA	Solvent accessible surface area
SP	Soft-Precision
STP	Single Trajectory Protocol
TcTS	<i>Trypanosoma cruzi</i> trans-sialidase
VS	Virtual screening
XP	Extra-Precision

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

COMPUTATIONAL DRUG DESIGN AND DISCOVERY FOR *TRYPANOSOMA cruzi*
TRANS-SIALIDASE

By

Bill Miller III

August 2013

Chair: Adrian E. Roitberg
Major: Chemistry

Chagas' disease is a lethal condition that affects millions of people worldwide, although the highest prevalence of Chagas' is found in rural regions of Central and South America. The World Health Organization has recognized Chagas' as a neglected tropical disease due to its low prevalence in developed countries. Although two drug treatments are available, they only work during the early phase of infection and incur unwanted side effects for patients. Thus, there is a critical need for new drugs in the treatment of Chagas' disease.

Trypanosoma cruzi, the protozoa that causes Chagas' disease, expresses a *trans*-sialidase (TcTS) enzyme that is vital to the parasitic life cycle due to its role in host cell invasion and defiance of the host immune system. TcTS specifically catalyzes the transfer of sialic acids from host glycoconjugates to glycoconjugates attached to the parasitic cell surface. Thus, TcTS is an ideal biological target for the design of drugs against Chagas' disease.

Despite high structural and sequence similarities with neuraminidase enzymes, there are currently no known strong inhibitors of TcTS. In this study, computational techniques are employed to identify novel lead compounds in the development of

inhibitors against TcTS. Pharmacophore models were developed using molecular docking of fragment compounds to TcTS. A virtual screening of the ZINC Leads molecular database using the pharmacophore models resulted in 82 novel inhibitor scaffolds. Molecular dynamics (MD) simulations followed by free energy of binding calculations were used to further refine the binding modes and relative theoretical binding affinity of the new scaffolds. Furthermore, virtual screening and *de novo* drug design procedures were utilized that involved docking of millions of compounds to TcTS in the pursuit of stronger inhibitors. The most promising candidates from docking were also subjected to further MD and free energy of binding calculations.

Computational drug design would be useless without the development of methods to estimate the free energies of binding of compounds. A newly developed program, *MMPBSA.py*, is introduced that was designed to quickly approximate the relative binding free energies of compounds from MD simulations.

CHAPTER 1 INTRODUCTION

1.1 Prologue

In this chapter, a description of Chagas' disease (American trypanosomiasis) will be introduced, including its transmission, epidemiology, clinical symptoms, and current prevention and treatment methods. The parasitic cause of Chagas' disease, *Trypanosoma cruzi* (*T. cruzi*), will be described in the following section. The subsequent sections will introduce the structural and mechanistic details of *T. cruzi* trans-sialidase (TcTS), which plays several key roles in the parasitic life cycle. Finally, the last section will include a review of the current efforts to design and discover strong, specific inhibitors of TcTS.

1.2 Chagas' Disease

Chagas' disease¹ is a lethal disease that currently affects approximately 10 million people, while the death toll each year is estimated at 12,500 people worldwide.² Although the disease is found mostly in underprivileged regions of Central and South America, recent immigration and international travel has caused Chagas' to spread to other parts of the world leaving nearly 25 million more people at risk of infection.³ In fact, estimates suggest there may be as many as 1 million cases of Chagas' in the United States alone,⁴ with thousands of other confirmed cases in Canada, Europe, Australia, and Japan.⁵ Chagas' relatively low prevalence in developed countries has left it mostly ignored by the international community, prompting the World Health Organization to consider it one of 17 neglected tropical diseases.⁶ The globalization of Chagas' has created more attention and research recently to combat the disease.⁷

In 1909, Brazilian physician Dr. Carlos Chagas fully described the disease named after him, including identifying and detailing the complete life cycle of the flagellated parasite *Trypanosoma cruzi*, determine by him as the lone cause of the disease.¹ *T. cruzi* can be transferred between humans by many mechanisms—blood⁸ or organ transfusions,^{9,10} from mother to baby during pregnancy,¹¹ and ingesting contaminated food or drink.^{12,13} But the most common method of infection is being bitten by the insect vector known as triatomines (members of the family Reduviidae and subfamily Triatominae), which are blood-sucking insects more commonly known as “kissing bugs” in some countries that live in the walls and roofs of poor buildings.

Chagas’ disease has two clinical stages—acute¹⁴ and chronic.¹⁵ The acute stage is often symptom-free, but can be characterized by the presence of swelling and redness at the infection site (typically the cut in the skin caused by the triatomine feeding) that lasts for weeks or months after infection. When the swelling is located on the eyelid, it is known as Romaña’s sign (Figure 1-1). The acute phase can also be associated with headaches, fever, rash, fatigue, vomiting, and diarrhea.¹⁶ The initial symptoms are all either nonexistent or not specific to only Chagas’, making the disease extremely difficult for physicians to diagnose. Following the acute phase, the symptoms (if present) fade and the condition appears to become dormant in the patient for years or even decades. However, during this period *T. cruzi* is hidden in specific tissues of the body, typically the heart or intestinal tract muscles. The chronic stage can be characterized by much more severe symptoms, usually arising from swelling of the tissues where *T. cruzi* resides. Parasites in the intestinal tract typically result in damage of the digestive system, such as megacolon or megaesophagus, while heart damage is

associated with arrhythmia, cardiomyopathy, and complete heart failure.¹⁷ However, the heart and intestinal problems are not mutually exclusive, as many in the chronic stage actually suffer from both sets of symptoms. The tissue enlargement is the result of an inflammatory response to cell death induced by the parasite, and is often fatal if untreated.^{18,19}



Figure 1-1. Romana's sign, a typical sign of Chagas' disease, occurs on the eyelid after being bitten by an infected Triatomine bug. This image is taken with permission from the CDC website (www.cdc.gov), which is an online source for credible health information and is the official Web site of the Centers for Disease Control and Prevention (CDC).

Currently, benznidazole²⁰ and nifurtimox²¹ (Figure 1-2) are the only two administered drug treatments for Chagas' disease, although the details of exactly how they work are unknown. Both drugs, which are only distributed by the Centers for Disease Control and Prevention (CDC), can be effective when correctly administered, but are most beneficial when treatment occurs in the acute phase of the disease, which is problematic given the difficulty in diagnosing Chagas' early. Treatment is lengthy and there can be many potential side effects to the drugs, often resulting in treatment being ended prematurely either by the patient or physician.²² The potential side effects from nifurtimox include anorexia, psychic alterations, excitability or sleepiness, and digestive

or intestinal problems (such as nausea, vomiting constipation, or diarrhea).

Benznidazole side effects mostly include skin manifestations (dermatitis, hypersensitivity, and skin blisters) or reduction in bone marrow. Some strains of *T. cruzi* are also known to be resistant to nifurtimox or benznidazole, and often when it is resistant to one of the drugs, it is resistant to the other.²³

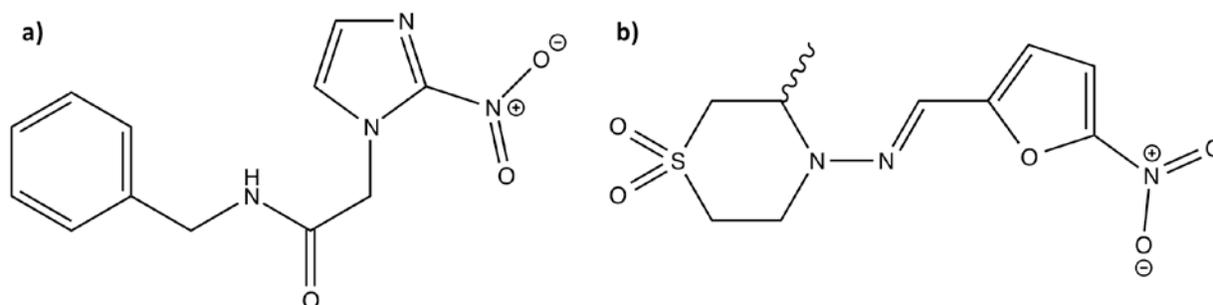


Figure 1-2. Structures of drugs currently available as treatments of Chagas' disease. Specifically, a) benznidazole and b) nifurtimox

Benznidazole and nifurtimox are anti-fungal treatments that combat the parasite directly, but recent progress has been made in the discovery of specific biochemical enzymes and pathways to target that are important to the *T. cruzi* life cycle. Studies indicate triazole derivatives can inhibit ergosterol synthesis and make promising drug candidates to treat patients in both the acute and chronic phases of Chagas' disease.²⁴ Cruzipain, a cysteine protease responsible for most of the proteolytic activity of *T. cruzi*, is an ideal drug target that can help prolong the lifespan of infected individuals with the proper inhibitor.^{25,26} Enzymes that have no mammalian equivalent make perfect drug targets for the design of selective *T. cruzi* treatments. One example of such an enzyme is trypanothione reductase that can induce oxidative stress on the parasite if disabled by an inhibitor.²⁷ Studies have also had success attempting to inhibit hypoxanthine-guanine phosphoribosyl transferase,²⁸ the synthesis pathway of phosphatidylcholine,²⁹ the

isoprenoid biosynthetic pathway,³⁰ and inositol metabolism³¹ in *T. cruzi* in the pursuit of novel treatments for Chagas' disease.

The lack of effective treatments has only helped increase the focus on prevention measures against Chagas' disease. There are currently no vaccines approved for use against the disease, but promising results have been found recently on infected mice using *Trypanosoma rangeli* epimastigotes to prevent *T. cruzi* infection.³² Another means of prevention available to the public directly targets the triatomines using insecticides.³³ Most preventative measures impact the life style of people in infected regions, and are intended to prevent infection from the insect vector. People are encouraged to solidify their home with concrete floors, plastered walls, and corrugated galvanized iron roofing to prevent the insect from entering the home. A still experimental method invented by Spanish chemist Pilar Mateo currently being tested in Bolivia uses paint infused with insecticides in microcapsules on the walls of homes to prevent triatomines from entering the living areas of the house.³⁴ More personal protection involves using nets covering the bed while sleeping and proper hygienic measures to prevent the consumption of contaminated food or liquid. Government regulation has helped with the implementation of laws to screen donated blood and tissue for *T. cruzi* infection prior to permitting its use in blood transfusions and other medical procedures.³⁵ Furthermore, organizations, such as the American Red Cross that screens all donated blood, are beginning to test more regularly for Chagas' disease. Despite the recent scientific advancements, there is still no single vaccine or drug approved to treat Chagas' disease.

1.3 *Trypanosoma cruzi*

The etiological agent of Chagas' disease is the flagellated protozoa *Trypanosoma cruzi* (*T. cruzi*), which can be generally described as physically consisting

of a DNA-containing organelle connected to a flagellum. Although some details of the evolution of the parasitic life cycle (Figure 1-3) remain unknown, it is generally accepted that *T. cruzi* has three distinct life cycle stages that can be observed using microscopy—trypomastigote, amastigote, and epimastigote.

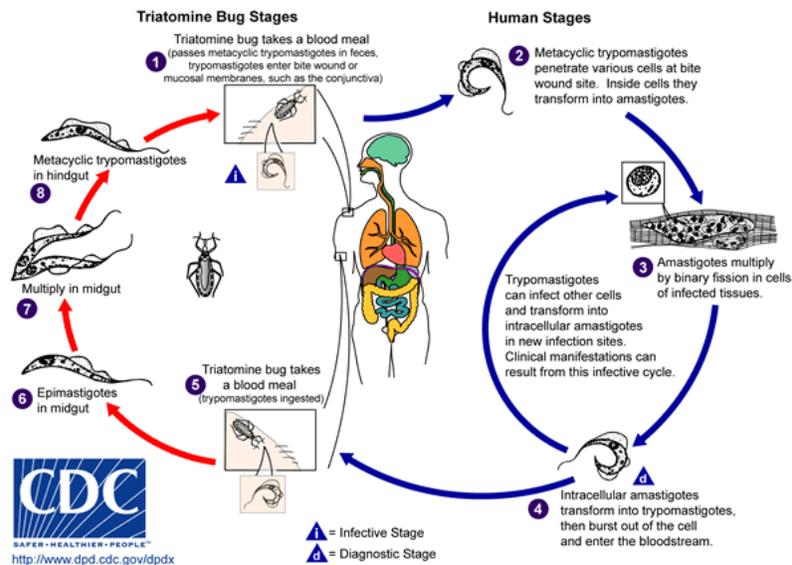


Figure 1-3. Life cycle of *T. cruzi* as it progresses through triatomine bugs and humans. Source: CDC website (www.cdc.gov).

When the blood-sucking triatomines feed on an infected mammal, *T. cruzi* is primarily in the trypomastigote form.^{36,37} The trypomastigote stage cannot replicate and is characterized by the parasite with a large flagellum connected to the nucleus (Figure 1-4). Upon entering the midgut of the insect vector, the trypomastigotes differentiate into the amastigote form, which are round, have a very short flagellum, and can undergo replication. The amastigotes quickly undergo differentiation into the epimastigote form, although this transition is reversible and *T. cruzi* is likely to be found in both forms within the insect vector. Epimastigotes can reproduce and are significantly longer than amastigotes with a clear and distinct flagellum. The epimastigotes proliferate as they progress through the intestinal tract of the insect and attach to the inside wall of the

rectum where they differentiate into metacyclic trypomastigotes,^{38,39} which are infective in mammals. During the triatomine's next meal, the trypomastigotes are deposited on the skin of the mammal in the insect's feces and make their way into the blood stream of the mammalian host through the skin abrasion created by the insect during feeding.

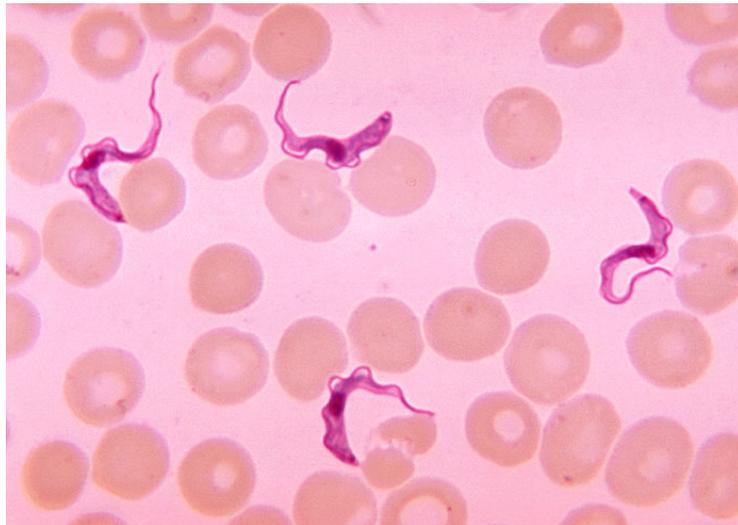


Figure 1-4. Image of *T. cruzi* in the trypomastigote form taken from a blood smear of an infected patient. Source: CDC website (www.cdc.gov).

Upon entering the mammalian blood, the infective trypomastigotes attempt to enter the host cells by interacting with the host cell surface. Invasion of the mammalian cells is mediated by the interaction of *trans*-sialidase and neuraminidase enzymes on the trypomastigote surface with the host cell surface. Once the trypomastigotes enter the host cell and reach the intracellular cytoplasm they differentiate back into the amastigote form to reproduce.⁴⁰ The amastigotes divide and multiply repeatedly followed by differentiation back into the trypomastigote form triggered by a yet unknown cause. *T. cruzi* multiplication fills the interior of the host cells until the pressure forces the cell walls to rupture, releasing infective trypomastigotes back into the host blood stream. This process of host cell invasion and induced cell death by *T. cruzi* continues

for years inside the mammalian host as the infected individual progresses into the chronic stage of Chagas' disease. The host immune system is able to create a balance with the parasite that slows the progression of the disease into the chronic stage. The level of trypomastigotes in the blood stream also makes the mammal a reservoir of infective parasites waiting to be ingested during a meal by another insect vector.

1.4 *Trypanosoma cruzi* Trans-sialidase

Sialic acids (Figure 1-5) are neuraminic acid (5-amino-3,5-dideoxy-D-glycero-D-galacto-2-nonulosonic acid) derivatives that have many important cellular functions.⁴¹ In particular, their location at the terminal position of cell surface glycoconjugates makes sialic acids ideally suited for intercellular communication, such as intercellular transport and immune system recognition.⁴² *T. cruzi* is unable to synthesize its own sialic acids *de novo*,⁴³ which prevents parasitic cells from invading the host mammalian cells and leaves it vulnerable to attacks from the host immune system. To combat this dilemma, prior to host cell invasion the parasite expresses a glycoposphatidylinositol (GPI)-anchored surface enzyme called *trans*-sialidase (TcTS).

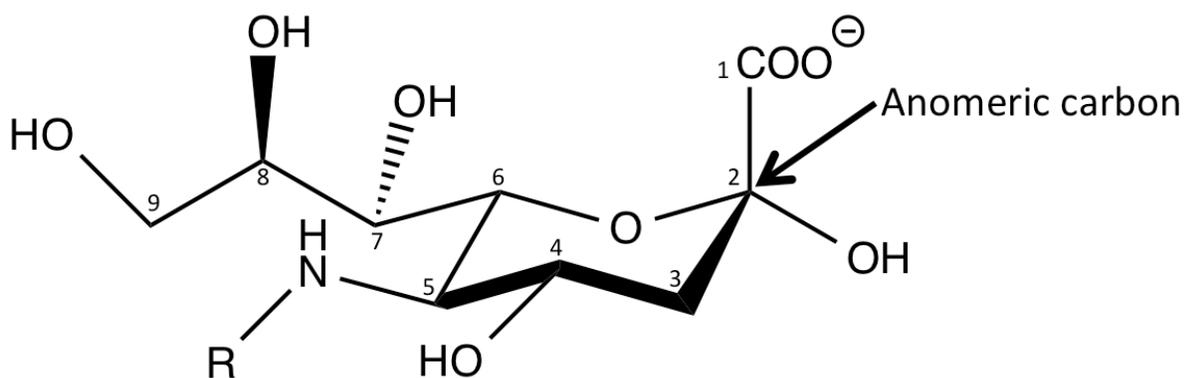


Figure 1-5. The general structure of sialic acid, where R=H is neuraminic acid, R=CH₃-CO— is N-acetyl neuraminic acid, and R=HOCH₂-CO— is N-glycolyl neuraminic acid. Traditional carbon atom numbering is shown along with the location of the anomeric carbon.

TcTS, expressed only in the trypomastigote stage of the *T. cruzi* life cycle, efficiently catalyzes the transfer of sialic acid from the host cell surface to glycoconjugates attached to the parasitic cell surface.^{44,45} The importance of TcTS, which has no human analog, to *T. cruzi*'s survival and ability to invade mammalian cells makes it an interesting drug target to combat Chagas' disease.

1.4.1 TcTS Structure

The X-ray crystal structure of free and liganded TcTS was resolved more than a decade ago by Buschiazzo *et al.*⁴⁶ For the holo complex (defined as the structure of the enzyme with substrate bound), the crystal structure had a resolution of 1.58 Å, with 94.3% completeness, and an R factor of 16.3%. The structure with DANA bound to TcTS was resolved to 2 Å with 85.9% completeness and an R factor of 16.3%. Finally, for the apo conformation (defined as the structure of TcTS with no ligand in the active site), the crystal structure had a resolution of 2.25 Å with 99.7% completeness and an R factor of 21.8%.

TcTS has a 70 kDa globular core connected to a variable length (100-500 residues) hydrophilic region called the shed acute phase antigen (SAPA) consisting of repeating 12-residue motifs.⁴⁷ The globular core consists of the N-terminal catalytic domain (residues 1-371) connected to a C-terminal lectin-like domain (residues 395-632) by an α -helix linker region (residues 372-394) (Figure 1-6). Finally, the C-terminal lectin-like domain is connected to the SAPA by a kinked α -helix region (residues 614-626). The catalytic and lectin-like domains are both similar in structure and sequence to related sialidases,^{48,49} with the tertiary structure of the catalytic domain consisting of a six-bladed β -propeller architecture. Compared to other sialidases, the α -helix linker region between the two domains is more hydrophobic, resulting in significantly more

buried surface area (2550 Å² for TcTS, 1300-1600 Å² for sialidases).⁴⁶ This observation suggests the protein core is more rigid, implying the lectin-like domain is not directly involved in catalysis given its spatial proximity to the catalytic domain. Although it is unlikely to be directly involved in transglycosylation, the C-terminal lectin-like domain has the potential to have a role in allosteric binding for substrate specificity.

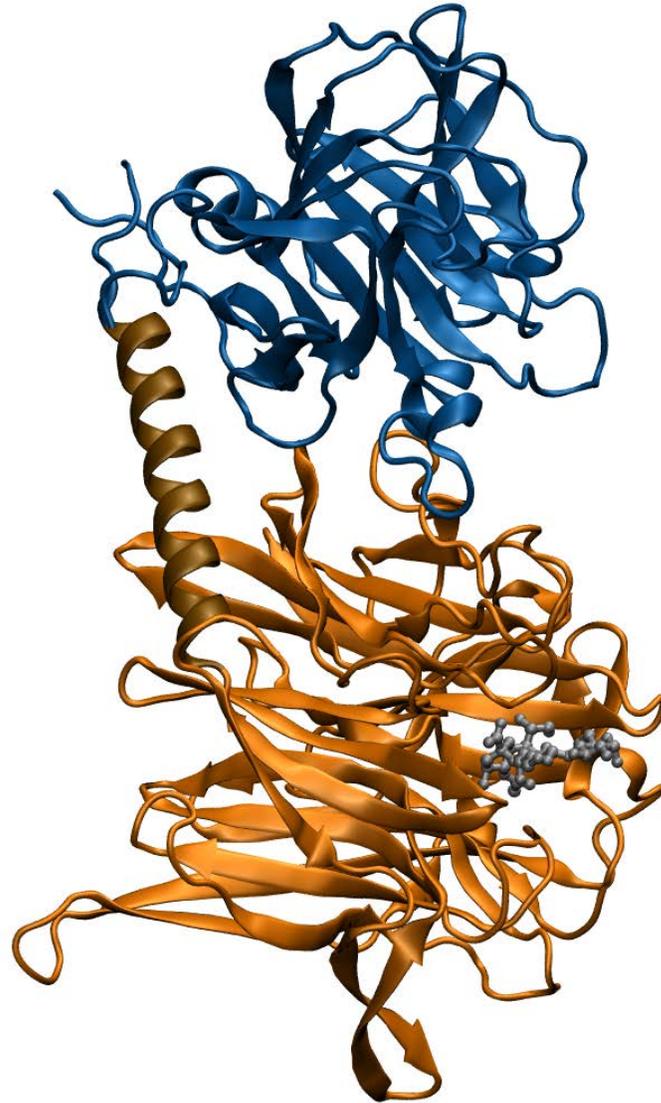


Figure 1-6. Structure of TcTS showing the three major domains. The catalytic domain is shown in orange ribbons, the lectin-like domain is shown in blue, and the alpha-helix linker region is shown in brown. Sialic acid is shown in silver balls and sticks within the active site of the catalytic domain. Coordinates are from PDB code 1S0I.

The TcTS active site (Figure 1-7) also shares several conserved residues and structural similarities with the sialidase family.^{48,50} In particular, the arginine triad (Arg35, Arg245, and Arg314) stabilizing the substrate carboxylic acid is conserved in TcTS compared to related neuraminidases, along with Glu357 that helps stabilize the side chain of Arg35. The catalytic residues—Asp59, Glu230, and Tyr342—were proposed based on their similarities to the corresponding residues in sialidases. Other residues involved in substrate binding were also conserved in TcTS—Val95, Trp120, and Leu176—that form the hydrophobic pocket where the N-acetyl group of the substrate binds. Compared to related sialidases, the binding site of TcTS is narrower and more hydrophobic, primarily due to the presence of two aromatic residues (Tyr119 and Tyr248) at the mouth of the binding pocket not present in sialidases. These residues (along with Trp312) are vital for positioning the β -galactose acceptor molecule in the binding pocket of TcTS, which at least in part prevents sialidases from performing transglycosylation because sialidases lack these aromatic residues.⁴⁶

The binding pocket of TcTS (Figure 1-7) can be divided into two main regions—the sialic acid binding site and the galactose acceptor/donor binding site. The substrate sialic acid binds deep into the active site, with the arginine triad motif stabilizing the carboxylic acid. Glu230 and Tyr342 of the catalytic motif are positioned below the anomeric carbon of the sugar. The N-acetyl group of sialic acid occupies the hydrophobic region near Val95 and Leu176, while the glycerol side chain resides in a more hydrophilic area of the active site near Trp120, Thr121, and Ser229. The galactose binding site is a narrow, hydrophobic region near the mouth of the active site located between the side chains of Tyr119 and Trp312 that helps position the acceptor

and donor β -galactose molecules for catalysis. Tyr119 and Trp312 lie parallel to one another with galactose situated between the two aromatic rings.

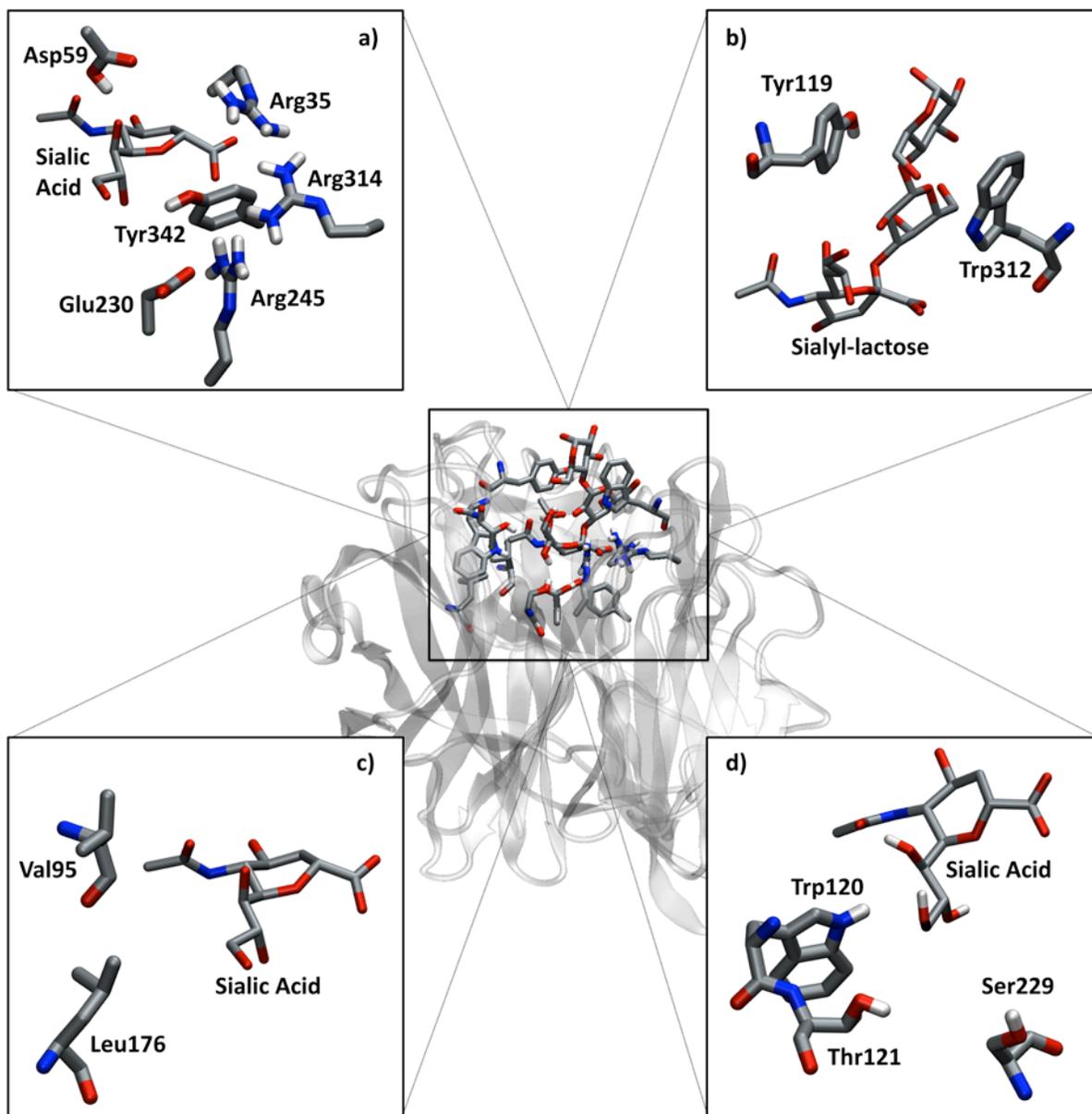


Figure 1-7. Depictions of the different regions in the TcTS active site with sialyl-lactose bound. The TcTS active site regions include a) the sialic acid binding site, b) the galactose binding site, c) the N-acetyl binding region, and d) the glycerol tail binding region.

1.4.2 TcTS Reaction Mechanism

The details of the TcTS catalytic mechanism have also been under investigation for more than a decade. Interestingly, TcTS differs from sialyltransferases because TcTS transfers sialic acid with retention of configuration,⁴⁵ suggesting a unique mechanism. Kinetic isotope studies were the first to suggest a nucleophilic enzyme residue,⁵¹ but elucidation of the X-ray crystal structure of TcTS showed Tyr342 as the only possibility based on proximity of the side chain with the anomeric carbon of the substrate⁴⁶ and that was deemed unlikely because the pK_a value of free tyrosine is too high to make Tyr342 a suitable nucleophile. However, Tyr342 was soon shown to not only be the nucleophilic residue, but was also part of a covalent intermediate formation between the enzyme and sialic acid.⁵² The hydroxy side chain of tyrosine typically has a pK_a value near 10, so the environment of the active site was clearly affecting the protonation state of Tyr342. Specifically, the side chain of Glu230 is positioned perfectly to accept the Tyr342 proton, thus altering the pK_a of the tyrosine and facilitating nucleophilic attack of the substrate by deprotonated Tyr342.⁵³ Recent computational results suggest this proton transfer is stable within the TcTS active site only after substrate binding because of the stabilization of the resulting charge density on the nucleophilic oxygen atom of Tyr342.⁵⁴ Before the substrate binds, Arg245 of the arginine triad forms a hydrogen bond with the side chain of Glu230, making Glu230 unlikely to accept a hydrogen transfer from Tyr342. However, upon substrate binding the arginine triad motif, including Arg245, hydrogen bonds with the carboxylate of sialic acid, creating a more suitable environment for Glu230 to accept a proton. Upon deprotonation, the nucleophilic oxygen atom of Tyr342 contains a larger electron

density when the substrate is bound, making the interacting of tyrosinate with the side chains of Arg35 and Arg314 more favorable with sialyl-lactose in the active site.

The attack of Tyr342 to form a covalent intermediate does not fully explain the reaction. The β -lactose leaving group must also have a corresponding residue responsible for its deprotonation prior to its departure from the active site. Protonated Asp59 performs the role of proton donor for TcTS catalysis because it is ideally situated for proton donation to the leaving group due to a conformational distortion of sialyl-lactose upon binding that leaves β -galactose in a pseudo-axial position.⁵³ Additionally, the entrance and exit of the donor and acceptor molecules, respectively, is aided by the flexible loop of Trp312 that positions and orients the sugar in the acceptor/donor binding region of the active site.⁵⁵

In general, the TcTS catalytic mechanism can now be rationalized as depicted in Figure 1-8. Pierdominici-Sottile *et al.* recently described the pathway and energetics of the mechanism from the Michaelis complex (MC)⁵³ to the covalent intermediate (CI)⁵² via nucleophilic attack of Tyr342 using computational methods.⁵⁶ The computed free energies showed a reaction barrier of $+20.80 \pm 0.70$ kcal/mol and an overall free energy change of -0.89 ± 0.43 kcal/mol from the MC to the CI. A negative free energy change for this reaction helped confirm the stability of the CI during the reaction. A carbocation was assumed to be formed during the reaction, but it was unknown if this state was a transition state or intermediate state until Pierdominici *et al.*⁵⁶ confirmed the carbocation state as the transition state (Figure 1-8). The carbocation arises from the dissociation of the lactose leaving group from sialic acid prior to the nucleophilic attack of Tyr342 on the anomeric carbon of sialic acid. The elucidation of the transition state is vital for drug

design efforts because it is well-known that the best inhibitor for an enzyme will mimic the transition state conformation.^{57,58}

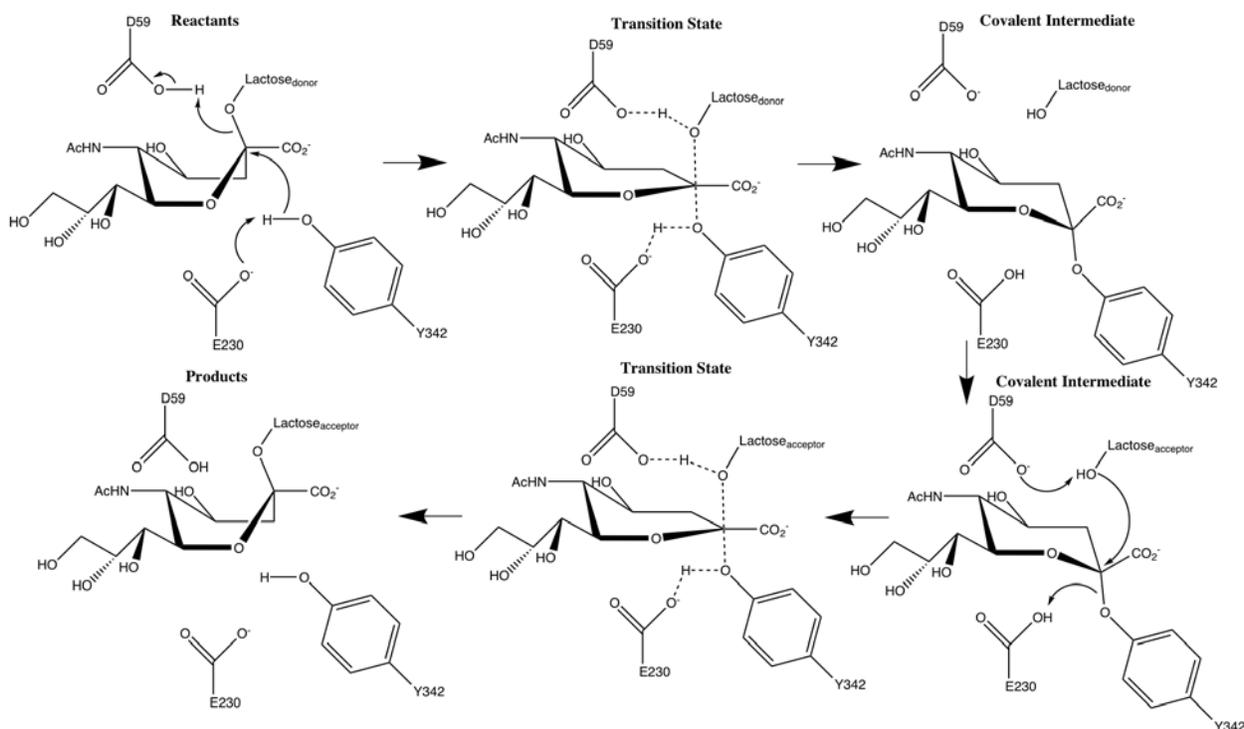


Figure 1-8. Chemical mechanism of *T. cruzi* trans-sialidase, showing the Tyr342/Glu230 hydrogen transfer and the nucleophilic attack of Tyr342 on the anomeric carbon of the substrate. Asp59 is also shown as a hydrogen donor for the leaving group. Note that sialic acid is rotated 180° compared to its orientation in Figure 1-5.

The TcTS transition state conformation is shown in Figure 1-9, where sialic acid is shown as a carbocation and the sugar ring is planar at the anomeric carbon. Although the transition state is not a pure carbocation because full dissociation of the leaving group does not occur prior to nucleophilic attack, the confirmation of little partial positive charge on the anomeric carbon is still helpful for drug design. The presence of an atom with partial positive charge character on an inhibitor near the hydroxyl group of Tyr342 would be logical. Furthermore, the transition state depicts the sialic acid sugar ring in a semi-chair conformation around the anomeric carbon compared to the traditional boat or

chair conformations found in the MC and CI, respectively. This observation suggests an aromatic ring of an inhibitor may be preferable for binding in the sialic acid region of the active site as it would be a better mimic of the transition state conformation. This is ideal for medicinal chemists because an aromatic ring is preferred over a sugar moiety in an inhibitor because it is more rigid with less rotatable bonds.

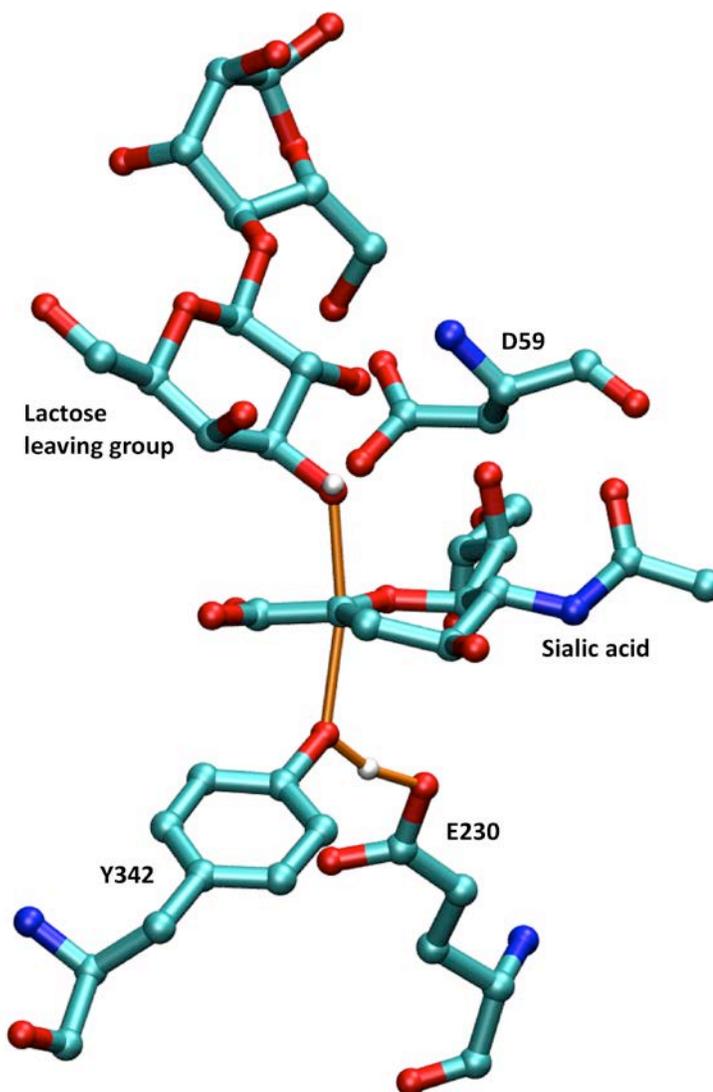


Figure 1-9. Conformation of sialyl-lactose at the transition state showing the planarity of the carbocation of sialic acid, as well as the positions of the catalytic residues Tyr342, Glu230, and Asp59. Note that the hydrogen transfer from Asp59 to the leaving group has already occurred, while all bonds in transition are

shown in orange. For clarity, all hydrogen atoms not involved in catalysis have been removed.

Further calculations by Pierdominici-Sottile *et al.*⁵⁶ showed the active site provided 21.09 kcal/mol more stabilization for the transition state than the MC, reemphasizing the idea that inhibitors should be designed based on the transition state conformation of the substrate. The free energy was decomposed by residue to determine exactly which residues were most important for transition state stabilization. Six residues—Arg35, Arg53, Asp96, Asp247, Arg311, and Glu230/Tyr342—provided the most stabilization (greater than 5 kcal/mol), where the Glu230/Tyr342 motif was considered a single residue because of the proton shuttle between them during catalysis. Arg35, Asp96, and Glu230/Tyr342 all directly stabilize the substrate in the transition state, while Arg53, Glu247, and Arg311 indirectly help stabilize the complex by forming favorable interactions with enzymatic residues involved in catalysis or positioning of the substrate. Previously, only residues that were directly involved in the reaction mechanism were known to be important for stability of the transition state.

Detailed analysis of the stabilization of the transition state is important for drug design efforts targeting TcTS. The importance of the Glu230/Tyr342 motif was already well understood because of its role in catalysis, but other residues important for transition state stabilization—namely Arg35 and Asp96—provide medicinal chemists with TcTS residues to target with rational drug design. Future inhibitors can be designed to interact directly with these residues, while secondary consideration should also be given to the other major transition state stabilizing residues—Arg53, Asp247, and R311. This information is vital for medicinal chemists when designing specific inhibitors of TcTS. For example, the transition state structure of TcTS and related sialidases led to

the design of 2,3-dehydro-3-deoxy-N-acetylneuraminic acid (DANA), a strong transition state analog inhibitor of sialidases that only weakly inhibits TcTS.

1.4.3 TcTS Dynamics

The structures derived from theory, NMR, and X-ray crystallography are vital to drug design, but proteins are dynamic molecules that cannot be fully described by a single conformation. Recently, Demir *et al.*⁵⁹ used classical molecular dynamics (MD) computer simulations to investigate the dynamics of TcTS in the bound (holo) and unbound (apo) state. The ligands used in the holo complexes included the covalent intermediate, the substrate sialyl-lactose, and DANA. A comparison of the dynamics from these simulations showed significant variability of the acceptor binding site regions, specifically the Trp312 loop, compared to the conformations proposed by the available X-ray crystal structures. This region of TcTS is vital for *trans*-sialidase activity because Tyr119 and Trp312 help orient the acceptor lactose in the active site for catalysis. The Tyr119 loop was mostly rigid throughout the simulations, but the Trp312 loop consisting of residues 310-314 showed varying flexibility depending on the particular ligand present in the active site (Figure 1-10). The Tyr119-Trp312 distance was 14.7-14.8 Å (closed conformation) for the crystal structures, which was consistent with the sialyl-lactose-bound TcTS simulation results. However, MD simulations of the DANA-bound TcTS and TcTS covalent intermediate complexes showed a more open protein conformation with Tyr119-Trp312 distances of 16-20 Å (semi-open conformation). The difference in loop flexibility between the X-ray crystal structure and the conformations generated from MD is likely the result of a glycerol molecule present in the acceptor binding site of these crystal structures (PDB codes 1MS1 and 2AH2), forming hydrogen bonds with the loop regions and preventing the active site from relaxing to the more

open conformation observed with MD. The Trp312 loop was even more flexible in the TcTS apo simulation, with Tyr119 and Trp312 reaching as much as 25 Å apart (open conformation). Crystal contacts and unit cell packing likely prevented this open conformation from being resolved using X-ray crystallography.

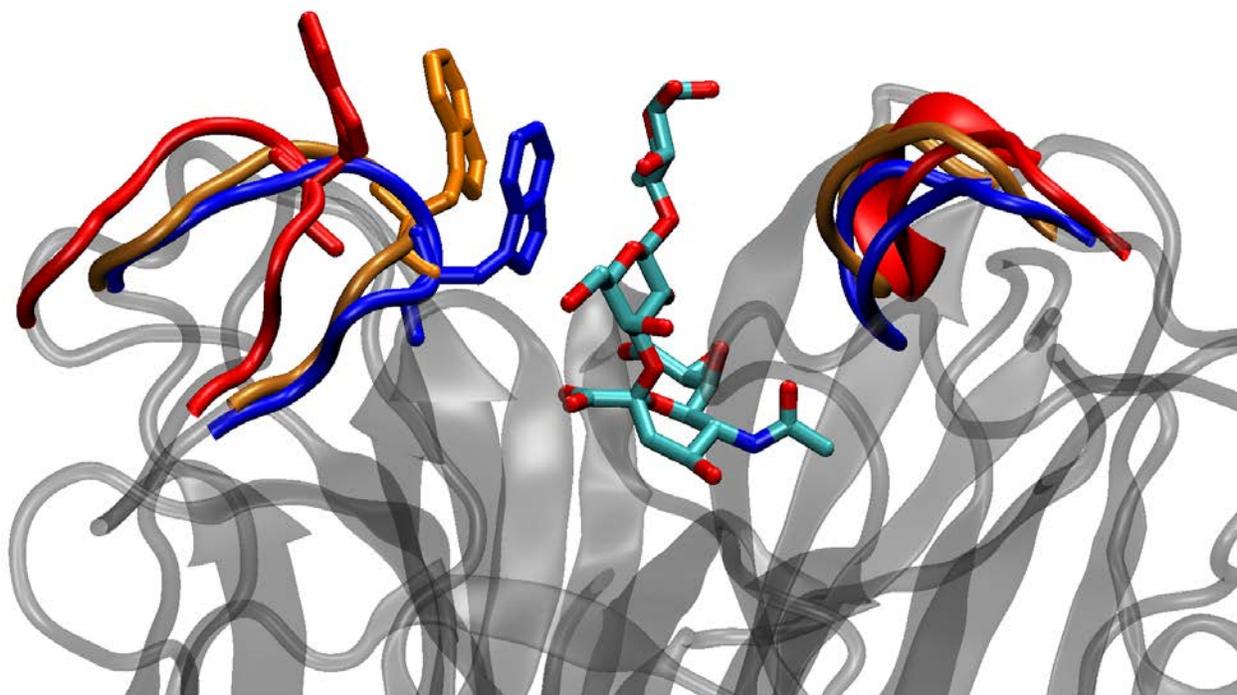


Figure 1-10. Overlay of TcTS structures displaying the flexibility of the Tyr119 (right) and Trp312 (left) flaps. Shown are representative flap orientations of the closed (blue), semi-open (orange), and open (red) conformations from MD. The sialyl-lactose substrate is shown for reference to depict the location of the binding site.

A hydrogen bonding network between residues Lys309, Asp337, Glu338, Asn339, and Arg7 help anchor the Trp312 loop while the interior of the Trp312 loop remains free to move in the apo form of TcTS (Figure 1-11). The difference between the semi-open and open conformation is related to the presence of a ligand in the sialic acid binding site of TcTS that interacts with the arginine triad (Arg35, Arg245, and Arg314), such as DANA and the TcTS covalent intermediate. The carboxylic acid interacting with

the Arg314 guanidinium group reduces the flexibility of the Trp312 loop by fixing the position of Arg314, preventing the loop from accessing the open conformation observed in the apo simulation. The side chain conformation of Tyr119 also appeared to affect the flexibility of the Trp312 loop. With the acceptor site occupied, Tyr119 prefers an outward conformation that leaves the side chain interacting with the acceptor lactose. Conversely, simulations with the acceptor site unoccupied showed a preferred conformation with the phenol positioned inward towards the catalytic cleft of the protein. This conformation of Tyr119 disrupts the water network in the active site enough to weaken the hydrogen bonds between the side chain of Trp312 and water in the active site, allowing the Trp312 loop to become more mobile.

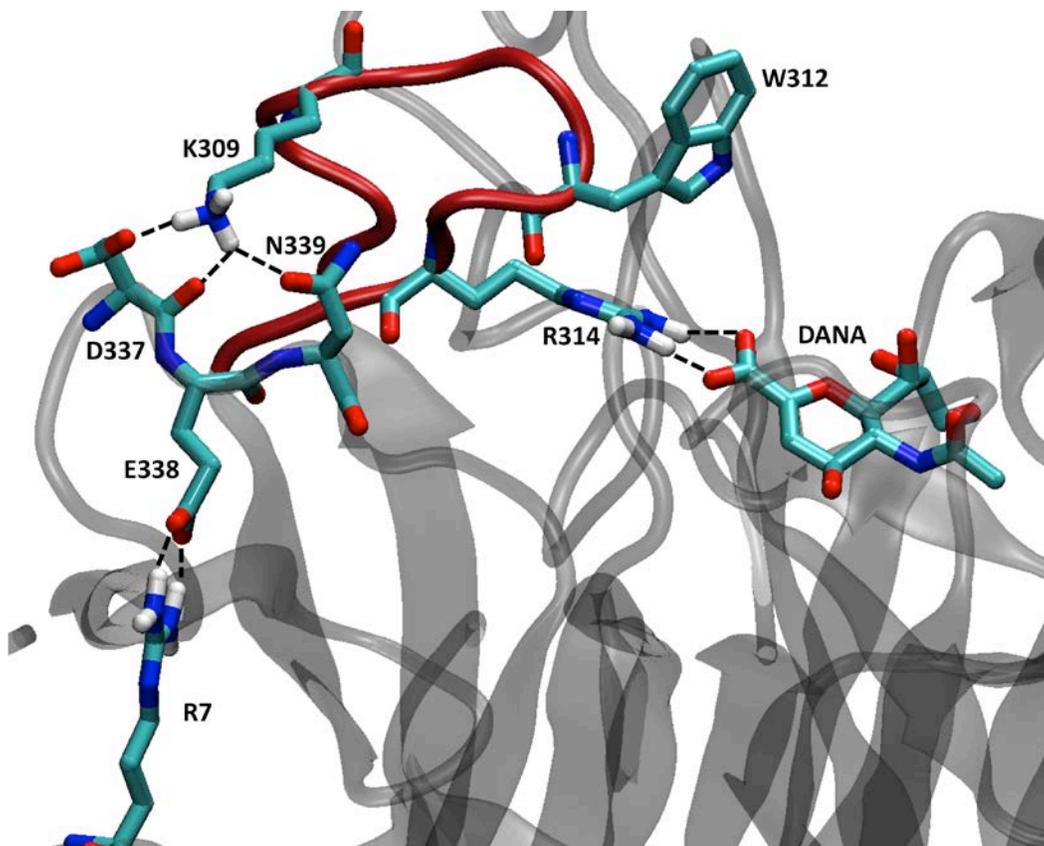


Figure 1-11. Hydrogen bonding network of the Trp312 loop (red). The loop becomes less flexible when Arg314 hydrogen bonds to a ligand, such as DANA.

Enzyme mobility is becoming increasingly recognized for its importance in the drug design process,^{60,61} thus the discovery of the W312 loop flexibility on TcTS using MD has major implications for the future design of inhibitors. Previously, the crystal structures were the only structures available, but the dynamics on the apo, DANA-bound, and covalent intermediate-bound structures of TcTS provide more conformations to consider for drug design. Furthermore, the simulation results show the importance of targeting the acceptor binding site with potential inhibitors of TcTS. The presence of DANA, a weak inhibitor of TcTS, in the active site diminishes the mobility of the W312 loop compared to the apo dynamics, but DANA does not stabilize the loop as much as sialyl-lactose because DANA is not able to directly interact with the Tyr119/Trp312 motif. Thus, a compound is more likely to be a strong inhibitor of TcTS if it interacts with both the sialic acid binding site *and* the acceptor binding region of the catalytic cleft.

1.4.4 TcTS Drug Design and Discovery Efforts

The structural, mechanistic, and dynamics work on *T. cruzi trans*-sialidase has been performed with one ultimate goal in mind: design and discovery of a strong inhibitor that can be used as a drug to treat Chagas' disease. Significant effort—both computational and experimentally—has gone into searching for a TcTS inhibitor, but still no strong inhibitor exists. Nevertheless, drug design and discovery on TcTS has uncovered several interesting lead molecules.

Being structurally and sequentially similar to related sialidases such as neuraminidase, naturally led to the consideration of DANA as a potential inhibitor of TcTS given its strong inhibition of influenza neuraminidase as a transition state analog.⁶² Surprisingly, DANA is a very poor inhibitor of TcTS, with an inhibition constant

(K_i) of only 12.3 mM.^{63,64} Regardless of the poor inhibitory activity, many derivatives of DANA have been designed attempting to inhibit sialidase and *trans*-sialidase activity with little success.⁶⁵ Similarly, successful inhibitors of other enzymes with sugar-moiety substrates have been assayed for inhibition of TcTS, such as the common glycosidase inhibitor Neu5Ac α 2-3-S-Gal β -O-octyl,⁶⁶ with no observed inhibition even at mM concentrations. The lack of affinity of TcTS for compounds that inhibit similar enzymes reveals the extreme difficulty in designing a strong inhibitor of TcTS.

Significant research has also been done on designing an irreversible inhibitor that would form a covalent bond with TcTS, preventing further catalysis. The first breakthrough for irreversible inhibitor design of TcTS was 2,3-difluorosialic acid, which forms a covalent bond with the nucleophilic oxygen of Tyr342 equivalent to the covalent intermediate in the catalytic mechanism.⁵² Despite inactivation of the enzyme in a time-dependent manner, TcTS fully regained catalytic activity in the presence of lactose unless very high concentrations (20 mM) of 2,3-difluorosialic acid were applied. Buchini *et al.*⁶⁷ investigated this observation further, and determined that 2,3-difluorosialic acid could be altered at the C9 position to accommodate the large and hydrophobic region in the lactose binding site, potentially making the inhibitor more specific for TcTS. Addition of an aromatic ring at the C9 position never created an inhibitor worse than 2,3-difluorosialic acid, and one substituent was significantly better at inactivating TcTS. Furthermore, although reactivation of TcTS catalysis still occurred in the presence of lactose, higher concentrations of lactose were necessary compared to using 2,3-difluorosialic acid. This observation was consistent with the hypothesis that the ligand aromatic ring occupied the lactose binding site, which was confirmed using X-ray

crystallography, and that TcTS inhibitors should target both the sialic acid and lactose binding sites.

A unique irreversible inhibitor, 2-difluoromethyl-4-nitrophenyl-3,5-dideoxy-D-glycero- α -D-galacto-2-nonulopyranosid acid (NeuNAcFNP), was designed by Carvalho *et al.*⁶⁸ that surprisingly results in the formation of a covalent bond between two enzymatic residues—Asp247 and Arg245. Unlike the high concentrations necessary for 2,3-difluorosialic acid inhibition to be observed, only 0.6 mM of NeuNAcFNP was necessary to reach 50% inhibition of *trans*-sialidase activity. Furthermore, 0.1 mM and 10 mM concentrations of NeuNAcFNP were able to prevent mammalian cell invasion of *T. cruzi* by 34.0% and 90.0%, respectively. This inhibitor has proven to be a strong lead candidate as a drug against Chagas' disease, but more optimization is necessary to increase the inhibition affinity and specificity for TcTS, such as substitution of an aromatic group at the C9 position of sialic acid as suggested with 2,3-difluorosialic acid.

Noncovalent inhibitors have been designed to mimic the lactose binding mode of the substrate. The first such inhibitor was lactitol (and similar lactose derivatives), which directly competes with sialic acid acceptors *in vitro* and *in vivo*.⁶⁹ In fact, sialic acid is preferentially transferred to lactitol 74.3% of the time over typical *in vitro* acceptor lactose and *N*-acetyllactosamine. The same group consequently synthesized and assayed derivatives of lactose covalently bound to polyethylene glycol (PEG) or oseltamivir (a sialid acid mimic), but the derivatives showed lower inhibition than lactitol.^{70,71} A different study used 'click chemistry'⁷² to synthesize and test a library of 1,4-disubstituted 1,2,3-triazole derivatives of galactose modified at the C1 or C6 positions designed to act as acceptor substrates of TcTS.⁷³ The compounds in the

library showed weak inhibition of TcTS (i.e. less than 40% inhibition at 1 mM inhibitor concentration), but interestingly did show trypanocidal activity against the trypomastigote form of *T. cruzi* Y strain at concentrations in the μM range. Another study on galactose derivatives examined the ability of TcTS to tolerate various functional groups substituted at the C2, C4, and C6 positions.⁷⁴ Finally, substituted C-glycosides have shown some promise as lead drug candidates with dissociation constant (K_D) values as low as 0.16 mM.⁷⁵

The lack of immediate success in designing strong, specific substrate analog inhibitors of TcTS naturally led to attempts to discover novel compound scaffolds with inhibitory activity. The first reported inhibitors not mimicking the TcTS substrate were benzoic acid and pyridine derivatives with poor K_i values in the high μM range discovered using virtual screening of the Asinex molecular library using DOCK.⁷⁶ The same group performed another virtual screening of the Evotec database using the GOLD docking software, resulting in the discovery of over 20 unique compound scaffolds yielding inhibitory activity.⁷⁷ A fluorimetric assay was used to measure the percent inhibition for analogs of two of the best scaffolds, resulting in IC_{50} values in the sub-mM range. Sub- μM competitive inhibitors were found in a separate study on the synthesis and testing of sulfonamide-containing chalcones and quinoline derivatives.⁷⁸ Furthermore, a QM/MM study by Lima *et al.*⁷⁹ on two of the quinoline derivatives, DHQ and THQ, was reported detailing the binding modes and inhibitory mechanism of these inhibitors in the TcTS active site. However, the most promising novel scaffolds reported to-date—flavonoids and anthraquinones—involved screening of a natural products library.⁸⁰ The strongest inhibitor analyzed, 6-chloro-9,10-dihydro-4,5,7-trihydroxy-9,10-

dioxo-2- anthracenecarboxylic acid, had a reported IC₅₀ value of 0.58 μM, which is the best inhibitor of TcTS discovered so far. Additionally, this compound was specific to TcTS, with a >170-fold specificity for TcTS over human neuraminidase.

Despite the discovery of several novel inhibitor scaffolds of TcTS, we still have not found a compound with inhibitory activity strong enough to be considered a good drug candidate. However, given the vast size of chemical space, the likelihood that a strong inhibitor exists is a distinct possibility. The task is difficult and unlikely to succeed without the efforts and collaboration of many researchers, especially those studying other sialidases, neuraminidases, and *trans*-sialidases because of the similarities between those enzymes and TcTS.

CHAPTER 2 METHODS

2.1 Molecular Structure and Function

The primary structure of a biological macromolecule (e.g. proteins or nucleic acids) describes the sequence of components—amino acids or bases—that comprise the molecule as a whole, but gives no details about their three-dimensional (3D) structure. The 3D structure of macromolecules is of great interest to structural biologists because the precise folded structure of a macromolecule determines its function. More than simply the approximation of the overall structure, the details of the orientation for all atoms in the molecule help structural biologists elucidate the details of chemical mechanisms for biomolecules, such as proteins and enzymes. These structures are equally important to medicinal chemists designing compounds to inhibit the function of these macromolecules. Many computational methods and techniques are available to medicinal chemists to design and discover novel inhibitors, such as molecular docking, *de novo* design, molecular dynamics, and free energy calculations.

2.2 Molecular Docking

Molecular docking attempts to orient a compound (usually a small molecule, or ligand) into a favorable conformation within the confines of macromolecular binding site typically using known three-dimensional coordinates for the macromolecule, or receptor. Additionally, an empirical scoring function is used to estimate the activity of the ligand based on its intramolecular atomic interactions with the receptor. A brief description of molecular docking methods, with specific emphasis on Schrödinger's Glide software, is presented here.

Structural drug design can attribute its start to Nobel laureate Dr. Emil Fischer who proposed the “lock and key” model for enzymes, describing how the enzyme active site is structurally complementary to the substrate structure as a key would fit directly into a lock (Figure 2-1).⁸¹ Drug design efforts using molecular docking traditionally used the “lock and key” model to describe the binding process, and the theory still has a prominent role in many molecular docking algorithms used by current computational software. The binding process involves not only the conformations of the complex, but also the structures of the receptor and ligand separately as they would appear in solution. By considering the receptor and ligand as rigid molecules, the number of degrees of freedom involved in the binding process is greatly reduced, which simplifies docking algorithms. The model of rigid molecules here is unphysical, but despite its limitations the assumption has led to many successful molecular docking applications attempting to predict and reproduce experimental protein-ligand bound structures.^{82–85} Unfortunately, for many receptors and especially small molecule ligands a rigid model is not appropriate and structural flexibility must be considered during the docking process.⁸⁶ In particular, sophisticated molecular docking methods should strive more to fit the induced-fit binding model, which suggests that an enzyme will adopt a different conformation upon ligand binding in order to maximize favorable protein-ligand interactions (Figure 2-1).⁸⁷ As the availability of computational resources has increased, so has the development of more sophisticated docking methods and algorithms based on our understanding of flexible molecular recognition, both for the receptor and the ligand.

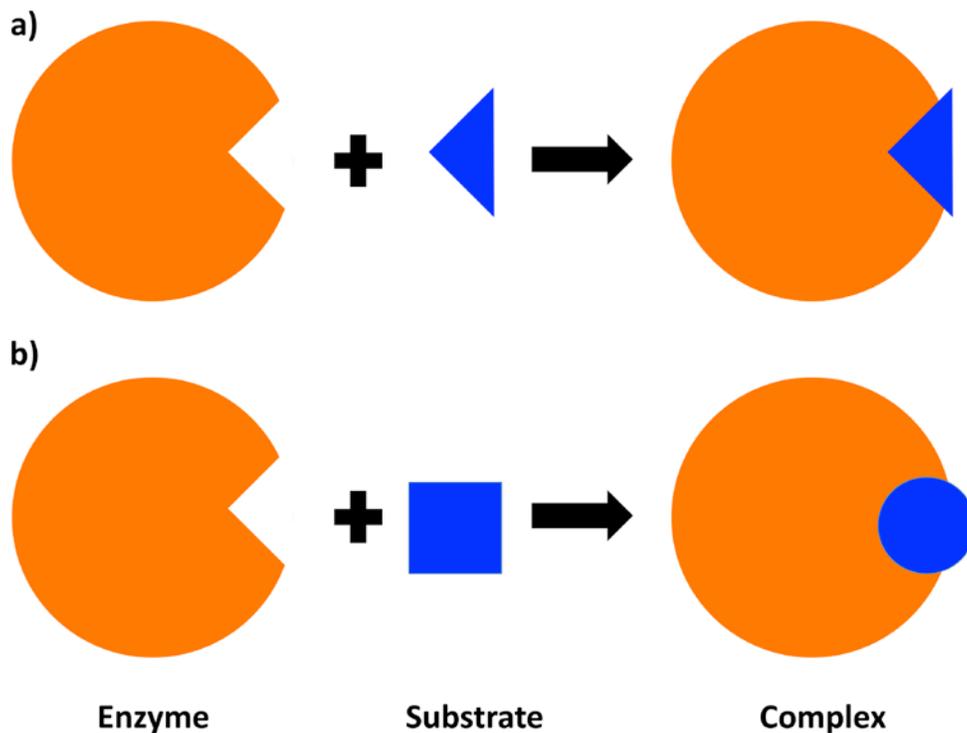


Figure 2-1. Schematic diagram of common models for a substrate binding to the active site of an enzyme. Specifically, a) the "lock and key" model and b) the induced fit model

The first docking software, DOCK,⁸⁵ released in 1982 used rigid spheres to represent both the receptor and ligand atoms, and all internal degrees of freedom were removed for docking. The receptor structure was actually depicted as a collection of spheres in grooves and pockets within the binding site instead of explicitly representing the receptor atoms. To account for ligand structure flexibility, in 1986 DOCK introduced an algorithm that divided the ligand into fragments and docked those fragments to the receptor binding site.⁸⁸ When the fragments were finally joined together, flexibility was integrated into the connection between the fragments. While DOCK developers were working on ligand flexibility, others were redesigning the way docking software handles the receptor structure by creating grid files that contain physical and chemical

information about the receptor. Grid files increase the efficiency of docking methods by pre-calculating information that can be re-used for each docking ligand. DOCKER used the grid files to define extended van der Waals radii of the receptor atoms,⁸⁹ while GRID pre-calculated the potential energy of receptor-ligand interactions with various functional groups at each point of the grid file.⁹⁰ These grids were further improved by including estimates to directly calculate the protein-ligand interaction energies by summing the van der Waals, Coulombic, and hydrogen bonding contributions.^{91,92} These grids set the precedent for automated grid file usage by many future docking programs.

The development of docking algorithms has progressed significantly since the release of the first docking program more than 30 years ago. Currently, the docking procedure generally involves four main steps—receptor grid creation, ligand conformation generation, ligand docking, and scoring of docked ligand poses. Although most docking programs use these steps, their details may vary (including combining certain steps) depending on the docking method and software. What follows is a brief description of the options and details of each step in the docking process.

2.2.1 Receptor Grid Generation

Many docking software packages now utilize a receptor grid file that contains structural and chemical information about the receptor and potential interactions with ligand atoms. The grid file contains the coordinates and charges of all receptor atoms within the defined binding region. Additionally, protein-ligand interaction energies are typically pre-computed by considering the possible interactions the receptor atoms can make with ligand atoms within the active site and then these values are stored in a table within the grid file. During the docking process, these energies are taken directly from this table instead of being computed on-the-fly during each docking run. When

performing a virtual screening on millions of compounds, a grid file with pre-computed interaction energies saves an enormous amount of computational time. In addition to interaction energies, the grid file also specifies the binding site dimensions and location with respect to the receptor structure. The binding site location is typically chosen based on prior knowledge of the protein, such as the location of a substrate or inhibitor compound within the active site from an X-ray crystal structure, but software is available to help predict binding sites of macromolecules if their location is unknown.^{93–96}

Furthermore, the size of the grid can also be adjusted to accommodate the goal of the docking run. For example, the grid size can be small when attempting to reproduce a crystal structure conformation because the binding mode is known *a priori*, but the grid size can be larger when screening for new inhibitors because the size and binding mode of each ligand may vary substantially within the binding pocket.

Many docking programs can also account for receptor flexibility using various methods. In 1991, the first method to include protein flexibility made the receptor atoms “soft” by reducing the van der Waals repulsion of overlapping atoms,⁹⁷ which is still being used by docking programs today.^{98,99} Explicit protein flexibility was implemented a few years later by creating different receptor structures that explore the flexibility of rotatable bonds on protein side chains.^{100–102} Rotatable bonds are identified and the torsions are systematically sampled to create different receptor structures that are used for ligand docking, although this method is clearly more computationally demanding since docking must occur to multiple receptor conformations instead of a single structure. The rotatable bonds can include all possible rotatable bonds of specified side chains, or simpler methods only consider the rotation of polar hydrogen atoms on

receptor side chains. More computationally expensive algorithms involve simultaneous optimization of the ligand and receptor active site atoms during the docking process to include protein flexibility during docking.^{103–105} The Induced Fit Docking protocol used by the Schrödinger program suite¹⁰⁶ involves docking ligands to a receptor structure with a soft van der Waals potential and relaxing the protein iteratively until the protein structure stops changing.⁹⁸ Molecular dynamics (MD) can be used for a more comprehensive sampling of protein conformations, and representative structures from the simulations can be used for ensemble-based docking methods,¹⁰⁷ such as the relaxed complex scheme.^{108–110} Ensemble-based docking methods use receptor conformations from MD simulations as the receptors in ligand docking, and the different conformations represent the global flexibility of the protein with or without a ligand in the binding site in contrast to the local flexibility probed by exploring rotatable bonds of protein side chains.

2.2.2 Generating Ligand Conformations

Ligand conformation generators are significantly more utilized than flexible receptor methods, especially when the ligand binding mode is unknown as in the case of virtual screening for drug design. When the bound conformation of a ligand has not been determined experimentally, docking programs often create a collection of many different ligand conformations using several different methods, although it is important to note that bond distances are typically not varied during this process. The first developed method broke the ligand into fragments and then joined the fragments after docking, creating flexibility at the location where the fragments were joined back together.⁸⁸ More recently, docking programs began using energy optimization techniques that minimize the docked structure in the active site to maximize favorable protein-ligand interactions.^{104,111,112} A more common approach initially proposed by

Judson *et al.*¹¹³ and used by programs such as AutoDock,¹¹⁴ DARWIN,¹¹⁵ and GOLD¹¹⁶ among others¹¹⁷⁻¹¹⁹ utilizes genetic algorithms (GA) to generate ligand conformations. GAs use evolutionary ideas to solve search and optimization problems, such as determining the optimal binding mode of a ligand in an active site. Initially, a population of ligand conformations are generated and docked to the receptor binding site, and then the torsions and coordinates of the lowest energy ligand structures are used to generate new conformations for the next population of structures. The torsions and coordinates are passed down from population to population, although “mutations” randomly create new torsion or coordinate values to ensure conformational diversity of the ligand as new populations are generated. This process is repeated successively for a determined number of generations to determine the lowest energy binding modes for each ligand. Stochastic ligand conformer generators, such as Schrödinger’s LigPrep program, generate different ligand poses by using a random number to vary the internal torsions for the rotatable bonds present. As a result, stochastic docking programs will yield different results with repeated runs regardless of the input ligand structure.¹²⁰ Deterministic docking programs use a systematic approach to create ligand conformers, which leads to identical results if the same docking calculation is performed repeatedly on the same processor. However, even deterministic docking programs have results that can vary significantly depending on the input ligand coordinates, especially for algorithms that use structure minimization methods during conformer generation.¹²¹

2.2.3 Search Algorithms

Docking programs use many different search algorithms to dock ligands into the binding sites of receptors, of which several have already been introduced as part of the ligand conformer generation, such as genetic, stochastic, and deterministic algorithms.

These algorithms all attempt to place the ligand into the proper binding mode within the defined receptor binding region. Stochastic models typically involve using random numbers to place and/or alter the ligand within the active site of the receptor. An example of a stochastic method is the previously mentioned genetic algorithm, which places the ligand within the active site and then makes random changes to the ligand translations, rotations, and torsional angles. Additionally, a few programs such as AutoDock¹²² perform an energy minimization of the structure followed by a Monte Carlo¹²³ evaluation of the Metropolis criterion¹²⁴ to determine which ligand conformations are used to create the next generation of structures.¹²⁵ The deterministic methods—energy minimization and MD—are also available using programs such as CHARMM,¹²⁶ but these algorithms often suffer from being unable to overcome large energy barriers, which can keep the ligand trapped in a local minimum conformation and not sampling all conformations within the receptor binding site. In order to surmount this dilemma, several MD algorithms have attempted using biasing potentials to lower energy barriers^{126–128} or sampling at increased temperatures^{129,130} to enhance ligand sampling and improve docking results.

Schrödinger's docking program Glide¹³¹ uses a unique docking process compared to other available software that combines several of the already described methods. Ligand conformations are generated by Glide using an exhaustive search of torsion-angle space, which is rare among docking programs. Glide performs a quick initial screening of all generated ligand conformations to determine promising ligand conformations, followed by a minimization of the ligand within the active site using the standard molecular mechanics OPLS-AA force field combined with a distance-

dependent dielectric model.¹³² After minimization, a Monte Carlo procedure is applied to the internal dihedral angles of three to six best ligand poses to sample the local torsional minima and optimize the structure within the receptor binding region. This procedure is thorough because it screens a large number of ligand conformations without being too computationally demanding, which is ideal for any docking procedure.

Incremental construction (i.e. *de novo* docking or anchor-and-grow method) is an alternative docking algorithm available for computational drug design that typically places a small molecule anchor or fragment into the binding site, and then attempts to 'grow' that molecule to maximize protein-ligand interactions. However, the first anchor-and-grow methods published actually docked multiple fragments of a single ligand into the docking site and later tried to join those fragments together.^{88,133} In 1992, DOCK became the first docking software to dock a single ligand anchor fragment into the active site and then systematically extend the anchor by exploring the torsions of all additions systematically,¹³⁴ which was later replaced with a minimization procedure.¹³⁵

Multiple Copy Simultaneous Search (MCSS)¹³⁶ and LUDI¹³⁷ were the pioneer programs that began using incremental construction algorithms as *de novo* drug design tools. MCSS randomly places 1,000 to 5,000 copies of many functional groups into a receptor binding site followed by minimization to obtain a map of the energetically favorable interaction sites of a receptor surface for ligand drug design. On the other hand, LUDI directly designs new enzyme inhibitors specific to the active site starting from an initial anchor fragment and adding new functional groups to interact with the protein.

More recently, another *de novo* software, AutoGrow, was released that uses AutoDock and a genetic algorithm to design new inhibitions.¹³⁸ A previously docked fragment provides the anchor in the active site, while hydrogen atoms are randomly substituted for other fragments from a database of small molecules (AutoGrow provides its own small and large fragment library). The genetic algorithm combines and mutates the functional groups from the best fitting ligands of the initial population to create a new population of compounds that are docked and re-scored used AutoDock. This process is repeated many times to remove undesirable functional groups, and properly orient favorable groups to evolve a novel compound that inhibits the target enzyme.

LigBuilder is a *de novo* program written in C++ that gives the flexibility to 'grow' compounds from an initial anchor compound or 'link' multiple docked compounds within the confines of a receptor binding site.^{139,140} The POCKET module of LigBuilder initially analyzes the receptor binding site from the provided 3D structure for key interaction sites. Unlike AutoGrow, LigBuilder does not have the ability to dock compounds, even though the initial anchor compounds must be docked into the active site frame of reference prior to using LigBuilder. But similar to AutoGrow, the growing and linking strategies of LigBuilder are controlled by a genetic algorithm designed to efficiently design new inhibitors, although the new population of ligands is generated using the elitism algorithm,¹⁴¹ where the top 10% of the original population is directly copied into the new population while the rest of the new population is created by combining the remaining ligands from the old population. The genetic algorithm significantly speeds up the design process compared to a systematic approach that could generate an impossibly enormous number of molecules after just a few growing rounds. Each newly

created compound is compared to all other compounds in the same population to ensure compounds are different and promote the analysis of unique molecules each generation. Furthermore, each structure is checked for chemical validity using built-in rules based on current drug molecules to guarantee only compounds with reasonable chemical connectivity are generated.

Despite their obvious ability to design novel drug compounds, *de novo* docking programs have several limitations. The number of chemically feasible compounds is enormous, which means it is highly unlikely that these programs will be able to find the absolute ideal drug molecule. Incremental growth programs also do not include receptor flexibility during the design process, which as mentioned previously can be a major contributor for ligand binding. Additionally, because the anchor fragment is static in the binding site, most programs do not minimize the final compounds, so it is suggested to optimize the ligand within the active size using an external structure minimization procedure.

2.2.4 Scoring Functions

After a ligand pose is generated from the searching algorithm of docking programs, that pose must be accurately scored in order to properly rank the compounds of interest. Regardless of how exhaustive or correct the docking algorithm is for a program, if the program fails to accurately predict the binding affinity of compounds then the results are useless. Meanwhile, the calculation of protein-ligand binding affinities must be quick and efficient because docking programs must be flexible enough to screen millions of potential drugs accurately in a reasonable time frame to be useful for drug design and discovery. Unlike more sophisticated binding affinity calculations such as thermodynamic integration¹⁴² or free energy perturbation,¹⁴³ molecular docking

programs use a simplified equation (or set of equations) called a scoring function to estimate a binding score or binding affinity. Each docking program uses a different scoring function determined to be a match for the docking algorithm. In general, scoring functions can be divided into three main categories—force-field based, empirical, and knowledge-based.¹⁴⁴

Force-field based scoring functions are the most computationally expensive scoring functions because of the complicated nature of the equations, but nevertheless the functions are built on first-principles and correspond to the non-bonded interactions (i.e. van der Waals and electrostatic) as they are described by classical MD simulations. The van der Waals contributions are described by a Lennard-Jones potential function¹⁴⁵ designed to estimate the attractive and repulsive nonpolar interactions between the receptor and ligand. The electrostatic interactions are estimated using a Coulombic interaction term based on the charge and distance between two atoms. Because of the computational time required to calculate these interactions, a distance cut-off is usually applied to reduce the number of interactions for each atom, although this clearly reduces the accuracy of the scoring function. The functional forms of additional energy terms—hydrogen bonding, solvation energies, and entropic contributions—have also been considered and implemented by several programs using force-field based scoring functions, such as GOLD,¹⁴⁶ DOCK,¹³⁵ and AutoDock.¹¹⁴

Empirical scoring functions are relatively simple, quick to evaluate, and based on a fitting of data from receptor-ligand complexes with known binding affinities. The equations do not reflect classical non-bonded interaction functionals, but can typically still be separated into distinct energy terms, such as hydrogen bonding, solvation free

energies, entropic contributions, ionic interactions, and nonpolar interactions. The functions are easy to solve, but the accuracy is based entirely on the chosen test set of complexes used to determine the coefficients in the equations. A test set can have a number of problems, such as the number and variety of complexes analyzed, leading to limitations of scoring function accuracy when the program is confronted with a complex not in the test set. Empirical scoring functions can vary significantly by the software package since the functions depend significantly on the test set chosen, but examples of empirical scoring functions can be found with LUDI,¹³⁷ FlexX,¹³³ and ChemScore.¹⁴⁷

Knowledge-based scoring functions are derived from the frequency of certain atom-atom interactions found in known protein-ligand complexes. Experimental protein-ligand structures are analyzed for the distance and frequency of each interaction; the higher the occurrence of a particular interaction, the more favorable that interaction is considered. These atom-atom contact populations are translated into a scoring function describing the interaction between two atom types where favorable interactions are favored over repulsive interactions. Similar to the empirical approach, knowledge-based scoring functions are simple and quick to compute. Furthermore, they are similarly limited by the chosen test set of protein-ligand structures, but knowledge-based approaches have the extra ability to specifically account for unique atom-atom interactions (e.g. π -stacking contacts and interactions with sulfur, phosphorous, halogens, or metals) that might not be described using an empirical method. Several knowledge-based scoring functions are available, including the PMF score,¹⁴⁸ DrugScore,¹⁴⁹ and SMOG¹⁵⁰ scoring functions.

Schrödinger's Glide docking program¹³¹ utilizes an empirical approach to describe its two most rigorous scoring functions—Standard-Precision (SP)¹³¹ and Extra-Precision (XP)¹⁵¹—that are based on the ChemScore function,¹⁴⁷ as shown in Equation 2-1. The SP function is suited for discovering ligands that bind well within a time frame suitable for screening large numbers of compounds. Consequently, the Glide SP scoring function is softer and allows some imperfections in the protein-ligand docked pose. The XP scoring function is stricter and applies a significant penalty for any docked pose that violates any known physical chemistry properties (protein/ligand strain, loss of entropy, and desolvation effects), which makes Glide XP ideal for obtaining more accurate docked poses and binding affinities. The increased accuracy of the XP scoring function comes at the cost of computational time, limiting the applicability of Glide XP to smaller numbers of compounds that can be reasonably analyzed.

$$\Delta G_{bind} = C_0 + C_{lipo} \sum f(r_{lr}) + C_{hbond} \sum g(\Delta r)h(\Delta \alpha) + C_{metal} \sum f(r_{lm}) + C_{rota} H_{rota} \quad (2-1)$$

In Equation 2-1, the constants, C , are derived from the experimental data obtained from the protein-ligand parameterization set. The first two summation terms extend over all atom-atom pairs for hydrophobic (i.e. lipophilic) and hydrogen bonding interactions between the ligand and receptor, while the final summation terms extends over all pairwise interaction between receptor metal atoms and the ligand atoms. The functions f , g , and h are distance (r_{lr} for ligand-receptor distances and r_{lm} for ligand-metal distances) and/or angle (α) dependent functions that yield a maximum value of 1.00 when the interaction between two appropriate atoms is within an ideal range. The functions are linear between 1.00 and 0.00 for distances and angles just outside the ideal range, and 0.00 beyond the cutoff for a favorable interaction (see Figure 2-2 for

the example of a hydrogen bond distance). The final term in Equation 2-1, $C_{rotb}H_{rotb}$, applies an entropic flexibility penalty for frozen ligand bonds, defined as a rotatable bond where the atoms on either side of the bond are in contact with the receptor.

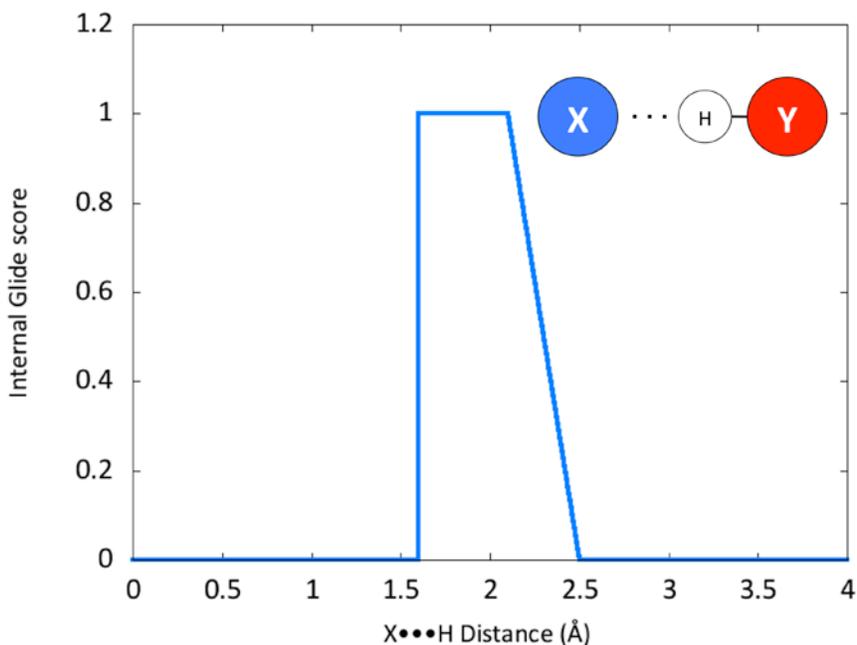


Figure 2-2. Depiction of how Glide scores a hydrogen bond based on the distance between the heavy atom (X) and hydrogen atom (H). A maximum score of 1.00 is given for distances between 1.60 Å and 2.10 Å, with a linear decrease from 1.00 to 0.00 for distances from 2.10 Å to 2.25 Å, and the score is zero beyond this region.

Glide expands the ChemScore function by dividing the summations into more specific categories, as shown in Equation 2-2.

$$\begin{aligned}
 \Delta G_{bind} = & C_{lipo-lipp} \sum f(r_{lr}) + C_{hbond-neut-neut} \sum g(\Delta r)h(\Delta \alpha) + \\
 & C_{hbond-neut-charged} \sum g(\Delta r)h(\Delta \alpha) + C_{hbond-charged-charged} \sum g(\Delta r)h(\Delta \alpha) + \\
 & C_{max-metal-ion} \sum f(r_{lm}) + C_{rotb}H_{rotb} + C_{polar-phob}V_{polar-phob} + \\
 & C_{coul}E_{coul} + C_{vdW}E_{vdW} + solvation
 \end{aligned} \tag{2-2}$$

The form of the pairwise lipophilic term remains unchanged from the initial ChemScore function. The hydrogen bonding term from ChemScore has been replaced

by three hydrogen bonding terms that describe type-specific hydrogen bonding interactions, allowing the three functions to be weighted differently, providing more flexibility than treating all hydrogen bonds equally. Fits to experiment recognized the hydrogen bond between two neutral functional groups as the most favorable, while the interaction between two charged species was the weakest. The fifth term in Equation 2-2 describes the interaction of a charged ligand functional group with a receptor metal ion in order to model the extreme favorability of metal-anionic coordination sites compared to metal-neutral interactions, which differs from the original ChemScore function that considers all metal-ligand interactions equal. The entropic cost of restricting ligand conformational motion upon binding is described by the sixth term in the scoring function. The seventh term represents the favorable interaction energy of a polar group without hydrogen bonding capability in a hydrophobic region. The Coulomb and van der Waals interactions (the eighth and ninth terms, respectively) required adjustments of atomic descriptors to improve the estimated binding affinities; reduction of the net ionic charge and van der Waals radii of formally charged groups such as carboxylates or guanidinium groups significantly improved the Glide SP binding energies compared to experiment by allowing the resulting hydrogen bonds to be in their ideal conformations. Lastly, solvation terms describing solvated charged and polar groups on the ligand and receptor are typically computationally expensive to calculate and are often neglected in docking scoring functions, but Glide estimates the solvation effects of the most energetically promising poses by explicitly docking water molecules into the binding site. The presence of explicit water molecules in the active site has a distinct advantage over using a continuum solvation model because the compact and

limited space in the active site dictates that the specific conformation of a water molecule is meaningful. Overall, this Glide SP function showed improved enrichment over previous scoring functions for docking a diverse test set of protein-ligand complexes.¹⁵²

The Glide XP scoring function¹⁵¹ is intended to be a significant improvement of accuracy over Glide SP, with the expected consequence of increasing the necessary computational time. The general form of the Glide XP scoring function is shown in Equation 2-3, while the details of represented functions are similar to those for Glide SP.

$$XP\ GlideScore = E_{coul} + E_{vdW} + E_{bind} + E_{penalty} \quad (2-3)$$

The interactions that are favorable for ligand binding are represented by E_{bind} , which is defined in Equation 2-4.

$$E_{bind} = E_{hyd_enclosure} + E_{hb_nm_motif} + E_{hb_cc_motif} + E_{PI} + E_{hb_pair} + E_{phobic_pair} \quad (2-4)$$

The first term, $E_{hyd_enclosure}$, estimates the favorable interaction energy of a ligand lipophilic group surrounded (i.e. enclosed) by receptor hydrophobic regions arising from the displacement of water molecules that were in a lipophilic environment in the active site, as well as a loss of the unfavorable relationship between bulk water molecules and the lipophilic group of the ligand upon binding. Theoretical and experimental analyses of pharmaceutical drugs and their targets illustrate the increased significance of hydrogen bonding between neutral groups of the receptor and ligand for drug potency, which is estimated in the $E_{hb_nm_motif}$ term of Equation 2-4. Another hydrogen bonding term, $E_{hb_cc_motif}$, estimates the energetic contribution to binding of special charged-charged hydrogen bond motifs. E_{PI} estimates the favorable π - π stacking and π -cation

interactions attributed to ligand binding. The final two terms that contribute to ligand binding, E_{hb_pair} and E_{phobic_pair} , represent the placement of ligand halogen atoms in a hydrophobic region of the receptor and a correction term to improve the binding affinity of small molecules that cannot make as many protein-ligand interactions as larger ligands, respectively.

The $E_{penalty}$ term in Equation 2-3 describes the unfavorable energetic contributions of protein-ligand binding, and is defined in Equation 2-5.

$$E_{penalty} = E_{desolv} + E_{ligand_strain} \quad (2-5)$$

As was the case for the Glide SP scoring function, the desolvation energy term, E_{desolv} , is calculated by docking explicit water molecules into the binding pocket instead of using conventional continuum solvation models. The efficiency of the calculation is improved by pre-calculating the energies during the receptor grid generation. However, the Glide XP scoring function is more heavily weighted and rigorous than SP using increased sampling techniques for docking the water molecules. Finally, the E_{ligand_strain} term does not contribute significantly to the overall binding energy, but accounts for the internal strain of a ligand structure and appropriately penalizes high-energy docked conformations of the ligand with close internal contacts.

Overall, docking programs aim to provide adequate assessment of the propensity for a ligand to bind to a particular receptor conformation, which is useful for drug design and discovery efforts but there are many limitations of these programs that should be considered. The algorithms and scoring functions are intended to be quick and relatively accurate allowing the screening of millions of compounds within a reasonable timeframe, which means the docking results should not be considered blindly. Docked

poses and energies should be considered the initial step in the drug design process, whereby more sophisticated theoretical and experimental methods should follow. Among the clear limitations of docking methods, the resulting poses are static structures that do not describe the dynamic nature of protein-ligand interactions. Other theoretical approaches such as molecular dynamics (as described later in Section 2.3) along with a more sophisticated binding free energy calculation should be considered after obtaining a docked protein-ligand conformation in order to more accurately represent the complex. Finally, a recent review showed little consistency among the results of several docking algorithms, and more importantly observed virtually no correlation between the docking scores and *in vitro* binding affinities.¹⁵³ This conclusion highlights the danger of utilizing just molecular docking during the drug design process.

2.3 Pharmacophore Models

Monty Kier originally proposed the idea of pharmacophore models in the late 1960s as a novel drug design tool.^{154,155} Kier's idea was to create a 3-D arrangement of chemical functional groups that are necessary for a molecule to initiate or hinder a receptor's biological activity. For drug design, these pharmacophore features represent the location and type of functional group that a drug molecule should possess to inhibit a particular target protein (Figure 2-3). Pharmacophore models are traditionally generated using the chemical features and structures of compounds with known experimental inhibitory activity. The six main chemical features in pharmacophore models—hydrogen bond donor, hydrogen bond acceptor, negative charged group, positive charged group, aromatic ring, and hydrophobic group—are used to represent the common functional groups of known active molecules. Furthermore, many software

packages are available to create pharmacophore hypotheses, including HypoGen,¹⁵⁶ HipHop,¹⁵⁷ Pocket v.2,¹⁵⁸ and Phase.^{159,160}

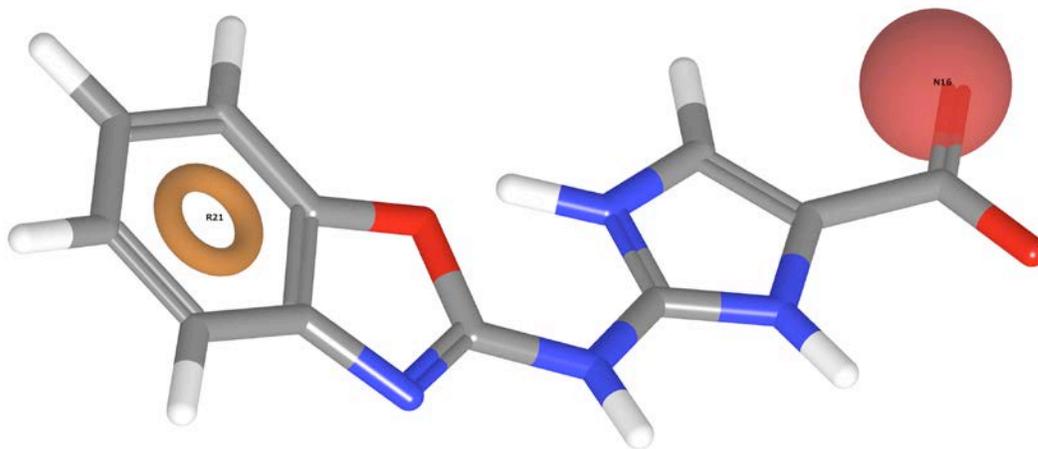


Figure 2-3. An example of a pharmacophore model. Here the red sphere represents an important negative ionizable group, while the orange donut represents the location of a key aromatic group.

Pharmacophore models can be created using either a ligand-based approach or structure-based approach.¹⁶¹ When structural data is not available depicting the binding modes of the known active molecules (i.e. inhibitors) in complex with the target receptor, then a ligand-based approach must be applied. Since the conformation of the bound ligand is unknown, many conformers of each active molecule are initially generated followed by an alignment of all conformers for the active molecules. Once aligned, the features of the pharmacophore model are extracted from the repeated occurrence of similar functional groups at the same spatial location. This approach has two main limitations that reduce its efficiency drastically. First, the results of any ligand-based method is linked directly to breadth of ligand conformations generated. Creating too few ligand conformations likely prevents the pharmacophore from using the proper ligand structure, while too many conformations will make the alignment process

extremely time-consuming and unlikely to converge. Second, the alignment process of the ligand poses is not well established, allowing for different pharmacophore models generated depending on the aligning algorithm. The ligand conformations can be aligned by maximizing atomic, fragment, or functional group overlap between different structures, and each will likely result in a different pharmacophore outcome. Nevertheless, ligand-based approaches can produce useful pharmacophore models for drug design.

The structure-based pharmacophore approach circumvents both limitations of the ligand-based approach by using the conformations of ligands bound to the target receptor as resolved from experiment (e.g. X-ray crystallography, nuclear magnetic resonance, etc.) or theory (e.g. molecular docking, molecular dynamics, etc). Furthermore, because the experimental ligand conformations are bound to the receptor, the conformations are already located in the same frame of reference (the receptor active site) and do not need to be aligned against one another. Instead, the ligands can be directly overlaid in their bound conformations, and statistical analysis is performed to extract the pharmacophore features for the model. The obvious disadvantage to this method is it requires available 3-D protein-ligand structures and cannot be performed otherwise.

The ligand-based and structure-based approaches as described both create pharmacophore models based on structural information of known inhibitors that hinder biological activity, which prevents pharmacophore models from being created for target molecules that have no strong inhibitors. Recently, an algorithm was implemented in Schrödinger's Phase that is able to generate pharmacophore models using fragments

(with no known binding affinity for the receptor) docked to a target receptor, as summarized in Figure 2-4.¹⁶² Initially, fragment molecules are docked into the active site of the molecule using Glide XP¹⁵¹ with an enhanced sampling procedure intended to allow the fragment ample probability to find the lowest energy conformation in the binding site. After docking, the docked fragments (up to 2,000 total fragments) are clustered into distinct groups to ensure the pharmacophore features represent a variety of locations within the receptor active site. Phase extracts the pairwise decomposable receptor-ligand interaction energies from the Glide XP output and a representative fragment structure of each cluster to generate features for energy-optimized pharmacophore models (e-pharmacophores), and then Phase ranks those features based on their perceived importance from the XP energetic terms.¹⁶³ These e-pharmacophores provide chemical features important for drug design for biological targets that currently have no known strong inhibitors.

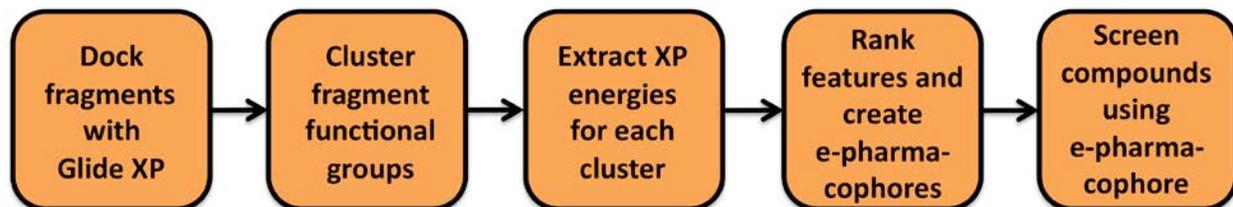


Figure 2-4. Workflow summary of e-pharmacophore generation from docked fragments using Schrödinger.

Besides containing important chemical features for biological activity, pharmacophore models are used for virtual screening of large molecular databases in the pursuit of novel inhibitors. The models provide a template of chemical features spatially arranged within the active site that is ideal for screening. Pharmacophore screening algorithms attempt to align the appropriate functional groups of potential

inhibitors in molecular databases to the chemical features of the pharmacophore. Hit compounds are molecules that match the required pharmacophore features within a specified tolerance distance of the feature location of the pharmacophore. Furthermore, pharmacophore models, such as those generated by Phase, use excluded volumes, spheres corresponding to the space occupied by receptor atoms surrounding the ligand, to prevent potential inhibitors from occupying the same space as the receptor atoms. Large database screening of pharmacophore models has resulted in the discovery of many novel enzyme inhibitors.^{164–166}

2.4 Molecular Dynamics

Unlike docking calculations, molecular dynamics (MD) simulations provide atomistic details about the motions of a system over time. First introduced in the 1950s,^{167,168} MD propagates Sir Isaac Newton's laws of motion describing the forces acting on atoms (Equation 2-6) to create successive configurations of a system depicting the movement of those atoms.

$$a = \frac{F}{m} \quad (2-6)$$

In particular, the differential equations corresponding to Newton's 2nd law of motion, as shown in Equation 2-7 using Cartesian coordinates, describe this atomic motion exactly, but they can only be solved analytically for simple systems.

$$\begin{aligned} \frac{d^2x_i}{dt^2} &= \frac{F_{x_i}}{m_i} \\ \frac{d^2y_i}{dt^2} &= \frac{F_{y_i}}{m_i} \\ \frac{d^2z_i}{dt^2} &= \frac{F_{z_i}}{m_i} \end{aligned} \quad (2-7)$$

Unfortunately, assumptions and techniques must be applied to estimate the motions of large systems where the coordinates of all atoms are relevant to the system dynamics. In numerical analysis, the solutions to differential equations are often approximated using finite-difference methods that divide the calculation into discrete infinitesimal time steps (δt) and assume the forces remain constant between time steps. The Verlet algorithm¹⁶⁹ is a common finite-difference method implemented by MD engines that uses a Taylor expansion to calculate the atomic coordinates and velocities as a function of time, as summarized in Equation 2-8. The Verlet algorithm can be applied using the leapfrog integration method¹⁷⁰ that calculates the current coordinates at time t using the velocity information from the following half-step ($t + 0.5\delta t$). Meanwhile, the velocity information at $t + 0.5\delta t$ is updated using the forces present at $t + \delta t$ in the leapfrog scheme. Consequently, the coordinates and velocities are always a half-step apart during a traditional MD simulation.

$$\begin{aligned}
 \mathbf{r}(t + \delta t) &= \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) + \dots \\
 \mathbf{r}(t - \delta t) &= \mathbf{r}(t) - \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) - \dots \\
 \mathbf{v}\left(t + \frac{1}{2} \delta t\right) &= \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t)}{\delta t}
 \end{aligned}
 \tag{2-8}$$

The time step size used during MD simulations must be chosen carefully to balance computational performance and the potential of causing instabilities in the simulation. A small time step requires more computational time to complete a simulation, while a large time step can break energy conservation rules and destabilize the system by allowing atoms to travel a relatively long distance between force updates. As a compromise for traditional MD simulations, the time step size is suggested to be one-tenth of the fastest motion of the molecules involved, which for most biological

systems is defined as the C-H stretching motion that occurs over 10 fs.¹⁷¹ Thus, many classical simulations utilize a 1 fs time step, unless the bond distances between all heavy atoms and hydrogen are fixed at their equilibrium distance using the SHAKE algorithm,¹⁷² which allows a 2 fs time step.

To be consistent with most experiments, MD simulations are often performed using constant (i.e. average) temperature conditions. Thermostat algorithms are modifications of an MD engine that regulates the internal temperature of the system according to the equipartition theorem that relates the average kinetic energy, $\langle K \rangle$, of a system with its temperature (T), as shown in Equation 2-9, where k_B is Boltzmann's constant and N_{dof} is the internal number of degrees of freedom for the system.

$$\langle K \rangle = \frac{3}{2} k_B N_{dof} T \quad (2-9)$$

A common practice to control the internal temperature of a simulation is to scale the atomic velocities up or down accordingly at each step, adjusting the total kinetic energy of the system to reach the desired target temperature. Furthermore, stochastic thermostat algorithms, such as the Andersen thermostat,¹⁷³ additionally reassign the velocities of atoms by occasionally choosing a new velocity randomly using a Maxwell-Boltzmann probability distribution, which is intended to mimic the propensity of a system to have irregular atomic collisions.

A system obtained from experiment (e.g. X-ray crystallography or NMR) or theory (e.g. molecular docking or pharmacophore screening) does not contain initial velocities for the atoms, so these values must be assigned. A common approach is to assign these velocities based on a random value chosen from a Gaussian or Maxwell-Boltzmann distribution of attainable velocities for a given temperature. The distribution is

designed to give reasonable velocities for the desired simulation temperature, although the random number must also be chosen wisely to prevent synchronization of trajectories that can lead to corruption of simulations using stochastic thermostats.¹⁷⁴ The velocities themselves are typically altered to force the overall net linear and angular momentum to zero for the system prior to any dynamics calculations.

Experiments with biological complexes are typically performed in solution, thus MD simulations must account for interactions with the solvent using a solvation model. Two general types of solvation models—implicit and explicit—can be chosen to describe the solvent. Implicit solvation models use a continuum dielectric medium to represent the solvent, while explicit models contain discrete solvent molecules surrounding the solute. Implicit models such as Generalized Born (GB),¹⁷⁵ Poisson-Boltzman,¹⁷⁵ and Reference Interaction Site Model (RISM)¹⁷⁶ are generally computationally efficient, but typically are unable to capture proper timescale motions and are more prone to system instabilities than explicit models. Explicit solvent models such as TIP3P water¹⁷⁷ and TIP4P water¹⁷⁸ generally result in more realistic solute conformations, but are significantly more computationally expensive than their implicit model counterparts.

When explicit water models are used to represent the solvent, periodic boundary conditions are often used to treat the interactions of the unit cell with its surroundings. For periodic boundaries, identical unit cells effectively surround the primary unit cell to mimic the presence of the solute in bulk solvent and remove interactions of the solvent or solute atoms with a solid surface (i.e. the edge of the unit cell). Although the atoms can move from one unit cell to another freely, the only coordinates that are recorded are

for atoms in the primary unit cell. Since all the unit cells are identical, as an atom from the primary cell enters a neighboring ‘imaginary’ cell, the corresponding atom in the opposite ‘imaginary’ unit cell will simultaneously enter the primary unit cell, ensuring that the number of atoms in each cell remains exactly the same. The Particle Mesh Ewald method uses an Ewald summation that converges quickly to efficiently calculate the long-range electrostatic interaction energies and forces between atoms in neighboring cells.¹⁷⁹

Classical MD uses molecular mechanics to calculate the energy of the system using only the atomic coordinates (excluding any electron information) generated by integrating Newton’s laws of motion. Molecular mechanics uses force fields—empirical mathematical functions that describe atomic interactions—to calculate the bonded and non-bonded energy terms for the system. Force fields primarily treat interactions like electrostatics or changes to bond distances, angles, and dihedrals, but are unable to treat processes that involve bond formation or bond breaking. Several different force fields have been developed that can produce reasonable dynamics and conformational energetics, such as the AMBER ff99SB,¹⁸⁰ parmbsc0,¹⁸¹ and gaff¹⁸² force fields, CHARMM27,¹⁸³ GROMOS,¹⁸⁴ and OPLS.¹³²

The MD performed throughout this study utilizes the AMBER force field, which is shown in its mathematical functional form in Equation 2-10. The first two terms on the right hand side of Equation 2-10 determine the energetic penalties associated with bond distance and angle deviations from their equilibrium values. The third term represents the energy associated with the dihedral angles. The last two terms represent the non-bonded van der Waals and electrostatic interactions, respectively, between atoms i and

j. Specifically, the fourth term is a 6-12 Lennard-Jones potential function¹⁴⁵ commonly used for computational reasons to describe the van der Waals interactions, while the final term uses Coulomb's law to calculate the interaction energy between two charged particles.

$$\begin{aligned}
 U(R) = & \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \sum_n \frac{V_n}{2} (1 + \cos[n\phi - \gamma]) + \\
 & \sum_{i < j}^{atoms} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right) + \sum_{i < j}^{atoms} \frac{q_i q_j}{R_{ij}}
 \end{aligned}
 \tag{2-10}$$

In addition to the functional form of the force field, the values of the constants (e.g. equilibrium bond distances and angles, spring constants, etc.) are also dependent on the chosen force field. These parameters are empirically derived to reproduce either experiment or high-level quantum mechanics calculations. Although every molecule has unique properties, the force field constants are chosen to be transferable to a variety of interactions in many diverse compounds. For example, enzymes are all structurally and sequentially different, but their amino acid building blocks are the same so many of the interactions can be described using similar parameters. Similarly, many force fields, including the AMBER force fields, define atom types that help group atoms into categories based on their atom number, expected hybridization, and connectivity to other atom types. Atom types allow force fields to be transferrable from one system to another by assuming that atoms in different systems with similar connectivity will behave the same. Many force fields define specific atom types for use with families of molecules, such as the AMBER ff99sb force field for proteins,¹⁸⁰ the AMBER parmbsc0 force field for nucleic acid biomolecules,¹⁸¹ or the general AMBER force field (gaff) for

simple organic molecules.¹⁸⁵ The atom types are often stored in a parameter or topology file that is referenced during simulations for force and energy calculations.

2.5 Binding Free Energy Calculations

The ensemble of conformations generated from MD simulations provides the necessary information to estimate the binding free energies of ligands bound to receptor molecules. These end-state free energy calculations are independent of the binding pathway, but have shown the capability to accurately predict relative binding free energies for drug discovery.^{186–188} The use of molecular docking using Glide combined with MD and binding free energy calculations to discover new potential inhibitors is depicted in Figure 2-5. The theory of free energy calculations will be discussed in more detail in Chapter 5 alongside the discussion of a recently developed program—*MMPBSA.py*¹⁸⁹—for performing these calculations.

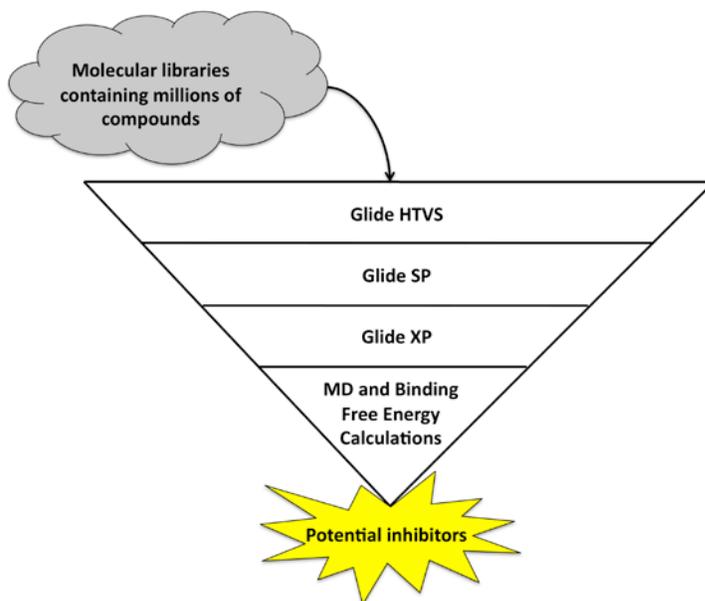


Figure 2-5. The workflow used in this study using a combination of molecular docking, MD, and binding free energy calculations to discovery new potential inhibitors of TcTS.

CHAPTER 3
DESIGN OF e-PHARMACOPHORE MODELS USING COMPOUND FRAGMENTS
FOR the *TRANS*-SIALIDASE of *TRYPANOSOMA cruzi*: SCREENING FOR NOVEL
INHIBITOR SCAFFOLDS

3.1 Preliminaries

Molecular docking has become a highly useful tool for drug discovery.¹⁹⁰ In particular, docking molecular fragments has aided in the discovery of new lead compounds for drug design.¹⁹¹ Due to their relatively small size, fragments help identify new regions of an enzyme's active site that were previously untargeted.¹⁹² Although most docking scoring functions are not robust, a recent study showed that Schrödinger's Glide¹³¹ can be a useful tool for docking small molecular fragments.¹⁹³

Structure-based pharmacophore models have also become increasingly important in computational drug design.¹⁹⁴ Pharmacophore models are used to identify the key features vital to inhibition activity for drug molecules. Traditionally, pharmacophore models are designed using a training set of active molecules. Thus, the question naturally arises what to do prior to any actives being identified for a given enzyme. Recently, Loving *et al.* developed a method that uses docked fragments to create structure-based energy-optimized pharmacophore (e-pharmacophore) hypotheses.¹⁶² This eliminates the need for known active molecules when generating pharmacophore models and helps identify new regions of the active site to target for drug design. Once developed, these e-pharmacophore models may also be used to screen for new lead drug molecules using the key features of the pharmacophore, which is a fast technique typically used for screening large databases of compounds.¹⁹⁵

Unlike docking that usually provides a single conformation, molecular dynamics (MD) provides insights into the dynamical behavior of protein-ligand complexes. MD is

also used to refine biological structures using classical laws of motion to describe the movement of atoms in time. It naturally arises that molecular dynamics can be used to further refine a protein-ligand structure from molecular docking calculations. Additionally, the MD simulations can be used to calculate the relative free energies of binding for a ligand bound non-covalently to a receptor, using such methods as Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) or Molecular Mechanics-Generalized Born Surface Area (MM-GBSA).¹⁹⁶ Finally, it has been shown that using these free energy calculations from MD simulations on docked protein-ligand poses can aid the virtual screening process of drug discovery.¹⁹⁷

We created the first e-pharmacophore models of *T. cruzi trans*-sialidase using fragments docked to the active site. Pharmacophore models were generated for the holo crystal structure (PDB code 1S0I), as well as holo and apo conformations of TcTS from previous MD simulations.⁵⁹ Using the e-pharmacophore models, we characterized the key chemical features important for strong inhibition of the enzyme for each conformation and used those features to screen the Clean Leads database from ZINC,¹⁹⁸ followed by MD and relative binding free energies calculations on all hit compounds. The best compounds from binding free energy calculations are presented in detail and proposed as novel lead drug candidates of TcTS.

3.2 Methods

3.2.1 Receptor Structures and Grid Generation

Pharmacophore models were created using three different structures of TcTS—one X-ray crystal structure and two structures from MD. These structures were chosen because they represent three unique and available conformations of TcTS for drug design. The crystal structure represents the most common conformation for analysis

because it corresponds to the only experimental structure of TcTS with the substrate bound, which is often considered the best conformation for inhibitor design. The structures from MD simulations represent relaxed conformations of the protein with the potential to reveal novel inhibitors that a pharmacophore model from the crystal structure is unable to discover. The X-ray crystal structure of TcTS with the substrate (sialyl-lactose) bound was obtained from the Protein Data Bank (PDB code 1S0I⁵³) for e-pharmacophore model generation (called the '1S0I' structure from now on), but the unliganded crystal structure (PDB code 1MS3) was not used because of its inability to capture the flexibility of the catalytic cleft.⁵⁹ Instead, a structure of TcTS in the apo conformation (defined as the structure of an enzyme without ligand present in the active site) was obtained from a MD simulation previously published by Demir *et al.*⁵⁹ Because the simulation produced thousands of conformations of the protein, clustering was performed with GROMACS¹⁹⁹ to obtain a more manageable number of statistically different conformations. The snapshots from the apo simulation were clustered with a root-mean-square deviation (RMSD) cutoff of 0.125 Å for all heavy atoms in the active site (defined as any residue within 15 Å of the nucleophilic Tyr342 residue). The representative conformation of the most populated cluster (62% of the total structures) was chosen as the apo structure of TcTS for pharmacophore model creation, and will be referred to as the 'apo' structure henceforth. For comparison purposes, clustering was also performed on a MD trajectory of TcTS bound to the substrate with a RMSD cutoff of 0.085 Å to maintain approximately the same number of structures in the most populated cluster between the two simulations. The most populated cluster corresponded to 56% of the total structures from the simulation, and the representative

structure from this cluster was chosen for pharmacophore hypothesis generation, and will be referred to as the 'holo' structure (defined as the conformation of an enzyme when the substrate is present in the active site).

All three TcTS structures were prepared using the Protein Preparation Wizard (PrepWizard) in Schrödinger, which is shown to improve virtual screening enrichment by fixing common problems with initial protein structures.²⁰⁰ The PrepWizard added any necessary hydrogen atoms and removed all water molecules in preparation for the receptor grid generation. The substrate location in the 1S0I structure was used to choose the center and size of the receptor grid, which was generated using Schrödinger's Glide with default settings for all parameters. The grid size was chosen sufficiently large to include all active site residues involved in substrate binding, including residues at the mouth of the active site used to position the acceptor/donor lactose molecule during catalysis. The grids for the apo and holo structures were chosen to have the same size and corresponding centroid point of the active site for consistency.

3.2.2 Docking Fragments to TcTS Structures

The Clean Fragments library of 504,074 compounds was downloaded in structure-data file (SDF) format from the ZINC¹⁹⁸ website. LigPrep,²⁰¹ which prepared ligands for docking with Glide, was used to generate reasonable conformers of each fragment using the default settings. All ligand conformers were docked to each of the three receptor grid files (1S0I, apo, and holo) using Schrödinger's Glide¹³¹ in High-Throughput Virtual Screening (HTVS) mode with default settings to filter out completely unreasonable ligands from being considered. The highest-ranking 120,000 compounds from virtual screening were re-docked using Glide SP with default settings, followed by

re-docking of the top 6,513 fragments from Glide SP using Glide XP¹⁵¹. The Glide XP settings were chosen based on the protocol of Loving *et al.*¹⁶² determined to be ideal for generating e-pharmacophores.

3.2.3 E-pharmacophore Generation

The results from fragment docking with Glide XP were used to generate e-pharmacophore hypotheses for each TcTS conformation using the E-pharmacophore script¹⁶³ available from the Schrödinger Script Center. The E-pharmacophore script works by mapping the Glide XP energies for the top 2,000 docked fragment poses onto the corresponding ligand atoms, aligned to the receptor conformations, followed by volume clustering of the fragments into 15 distinct clusters. Phase¹⁵⁹ was used to generate the pharmacophore hypotheses from the docked poses of the clusters using the default set of six chemical features in Phase: hydrogen bond acceptor, hydrogen bond donor, negative ionizable, positive ionizable, aromatic ring, and hydrophobic. The Glide XP energies for each pharmacophore site in the hypothesis are summed and ranked based on their total energetic contributions. A van der Waals scaling of 0.5 was used for receptor-based excluded volumes in each hypothesis to account for the shape of the active site. For each hypothesis, the 15 most favorable pharmacophore sites were kept, which is more than typical hypotheses. For TcTS, we were attempting to discover new regions of the active site that could lead to novel inhibitors, so analyzing more sites increases our likelihood of doing so.

3.2.4 Pharmacophore Screening

We searched the Clean Leads library of 4,230,832 compounds in order to find novel inhibitor scaffolds of TcTS. This library represents much of chemical space, which is ideal for this type of screening. Molecules were required to match a minimum of 3 or 4

sites on each hypothesis to be considered a “hit compound”, but preference was given to any molecules that matched more than the required sites. The chemical features of the required pharmacophore sites were used to filter the Leads library with Schrödinger’s Maestro v9.3 interface for only compounds that contained those features. Filtering the initial database saves time, but should have no effect on the screening results since the excluded compounds would never be chosen by Phase as hit compounds anyway. During the screening process, conformers were generated for each ligand using a rapid conformational sampling procedure. Ten conformers were created for each rotatable bond of each ligand, with a maximum of 100 conformers generated per structure (both defaults), which was a reasonable value given that the lead compounds are all relatively small and typically contained less than 10 rotatable bonds. Excluded volumes, required matches, and an intersite distance matching tolerance of 2.0 Å for all sites were applied during the screening process. Phase attempts to maximize overlap between the ligand conformations and the locations of the pharmacophore sites, which means the protein structure is not explicitly included during the virtual screening process. Only one structure was returned for each compound that matched the hypothesis. Finally, default settings were used for all coefficients when calculating the Phase fitness score and ranking the hit compounds.

3.2.5 Molecular Dynamics and Free Energy Calculations

To refine the ligand structures from Phase, the GPU accelerated *pmemd* code²⁰² of Amber 12²⁰³ was used to perform MD on each hit compound in complex with the catalytic domain of TcTS. The C-terminal lectin-like domain and the α -helix linker between the C- and N-terminal domains were excluded during MD simulations because the crystal structures implied these domains were unlikely to be directly involved with

the catalysis, and any allosteric binding contributions would not apply in the binding event of small molecule inhibitors. Because the e-pharmacophore screening process does not include the protein structure, several ligands had atoms that overlapped with protein atoms, so the atom positions were manually moved using VMD²⁰⁴ to remove any overlap prior to MD. The *antechamber* module of Amber was used to calculate the AM1-BCC charges for all ligands. The *gaff*¹⁸² and *ff99SB*¹⁸⁰ force fields were used to define all remaining parameters for the hit compounds and TcTS, respectively. All complexes were neutralized with counter ions and solvated with TIP3P water¹⁷⁷ in a truncated octahedron with a minimum of 12.0 Å between the solute and the edge of the unit cell. All initial structures underwent a seven-step minimization procedure using positional restraints on all solute heavy atoms, and slowly reducing the restraint weight from 10.0 to 0.0 kcal/mol/Å². After minimization, the protein was restrained with a positional restraint of 10.0 kcal/mol/Å² while the system was linearly heated from 10 K to 300 K over 2.0 ns. A seven-step equilibration period over 3.5 ns was used to slowly remove the positional restraints on the protein from 10.0 to 0.0 kcal/mol/Å². Finally, 10.0 ns of unrestrained MD was performed on all solvated protein-ligand complexes with coordinates saved every 1.0 ps, which was used for all analyses. The SHAKE algorithm¹⁷² was applied to fix all covalent bond distances involving hydrogen, allowing a time step of 2 fs for dynamics. All simulations were performed using periodic boundary conditions under constant pressure and temperature using an Andersen thermostat.¹⁷³

Hydrogen bonding analysis was performed using the *cpptraj* module of AmberTools using default hydrogen bond definitions. Single trajectory free energy of binding calculations were performed on 200 evenly spaced snapshots from the

unrestrained MD simulation with the *MMPBSA.py* program¹⁸⁹ released alongside AmberTools using the Generalized-Born (GB) model²⁰⁵ to calculate solvation free energies. GB is a good approximation of solvation energies compared to more rigorous free energy perturbation calculations when computing relative free energies of binding.²⁰⁶ We assume the solute entropy for all systems remains the same, so all entropic contributions to binding were neglected during free energy calculations.¹⁹⁶ It is important to note that the calculated free energies are not expected to be true binding free energies, and are used only to compare relative to one another.

3.3 Results and Discussion

In this section, we will describe the details of all three pharmacophore models, the rankings of the hit compounds discovered, and details of the two most promising inhibitors of TcTS obtained from pharmacophore screening.

3.3.1 Pharmacophore Hypotheses

3.3.1.1 1S0I pharmacophore model

This pharmacophore model was generated using the TcTS conformation from the 1S0I PDB X-ray crystal structure. Table 3-1 provides a breakdown of all 15 pharmacophore sites generated by Phase, including the corresponding interaction with TcTS each site represents. Several of the pharmacophore sites found by Phase were expected. For example, it is well known that the arginine triad (Arg35, Arg245, and Arg314) is important for stabilizing the substrate carboxylic acid during ligand binding for both sialidases and TcTS,⁴⁶ so it was not surprising that feature N10 was the most favorable site. Furthermore, the acceptor binding site between residues Tyr119 and Trp312 at the mouth of the binding site was previously shown to be important for substrate positioning in the active site,⁵⁹ explaining the presence of feature R20 in Table

3-1. The occurrence of these two features (N10 and R20) is a testament to the validity of the method at selecting reasonable pharmacophore sites.

Clearly, no single drug molecule would possibly form all 15 interactions described by the pharmacophore model. Thus, when screening for new inhibitors, we decided to require a compound match four of these features to be considered a “hit compound” because we found using more than four features resulted in no hit compounds.

Table 3-1. Pharmacophore sites of the 1S0I model. Sites marked with an asterisk (*) were required matches during screening. Feature labels were given by Phase, with the letter indicating the type of interaction (N = negative ionizable, A = hydrogen acceptor, D = hydrogen donor, H = hydrophobic, and R = aromatic ring). The scores are Glide XP summed energy values in units of kcal/mol.

Rank	Feature Label	Score	Interaction with TcTS
1	N10*	-13.39	Arginine triad (Arg35, Arg245, Arg314)
2	N109	-4.19	Arg35 and Arg53
3	A18	-3.14	Asp59 backbone –NH and side chain –OH of Tyr119
4	H84	-2.52	Hydrophobic pocket near Val95 and Leu176
5	D69	-2.20	Side chain carbonyl oxygen of Asn60
6	A3*	-1.94	Arg93
7	N17*	-1.67	Arg53 and Arg35
8	A22	-1.60	Ser229
9	D43	-1.60	Side chain oxygen of Tyr342
10	D60	-1.60	Side chain carbonyl oxygen of Gln195
11	D82	-1.37	Gln362
12	R20*	-1.25	Acceptor binding site between Tyr119 and Trp312
13	D45	-0.80	Asp96
14	D56	-0.72	Asp247
15	A16	-0.70	Arg314 and Arg35

Phase provides the ability to not specify the four mandated features, allowing any four features to be matched. However, this is very computationally demanding using a pharmacophore model with 15 features and seems unnecessary since we have knowledge of a few important features *a priori*. We decided to select the four required features based on our understanding of the protein structure and catalysis. Initially, we chose both N10 and R20 because of their previously mentioned importance for substrate binding. Our third choice was N17 (negative ionizable and hydrogen acceptors were both considered matches) because Pierdominici-Sottile *et al.* recently found Arg53 to be the most stabilizing non-catalytic residue of the transition state complex for TcTS.⁵⁶

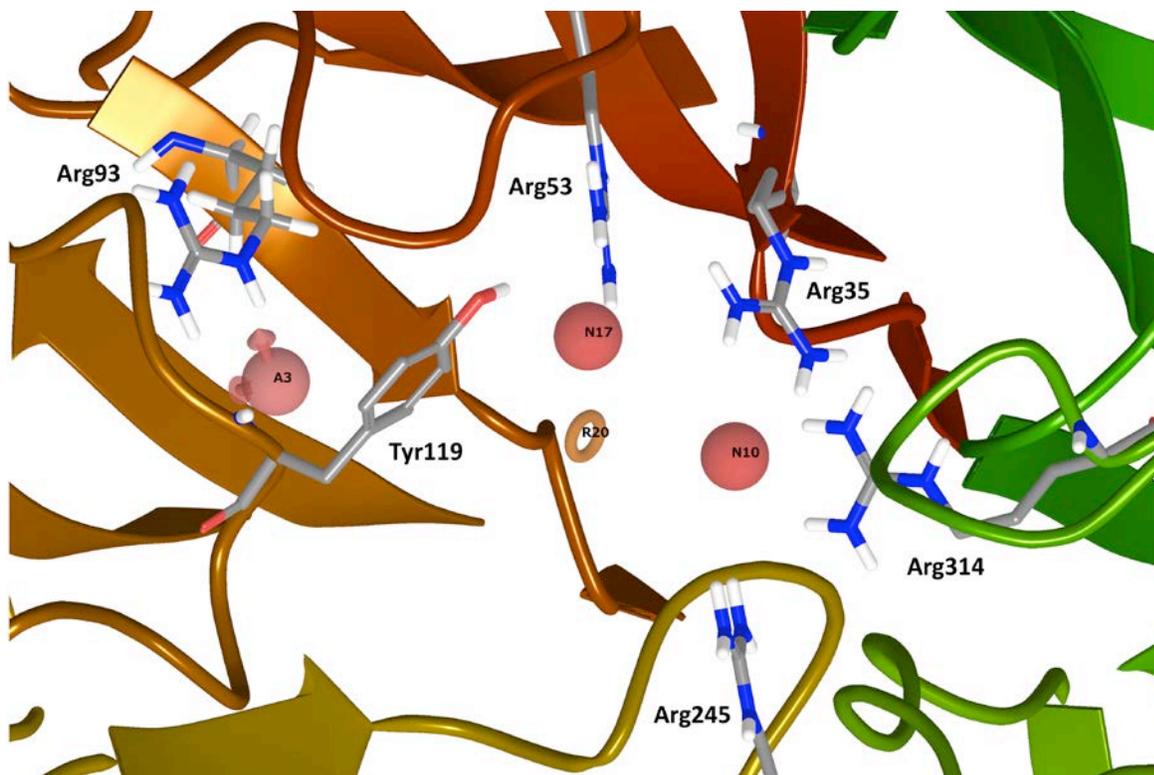


Figure 3-1. The required matches for the 1S0I pharmacophore used for screening lead compounds. Feature N10 interacts with the arginine triad (Arg35, Arg314, and Arg245), while N17 and A3 represent favorable interactions with the side chains of Arg53 and Arg93, respectively. The aromatic feature R20 is located in the acceptor binding site near Tyr119.

We wanted the final required feature to be a new site that has not been previously targeted by inhibitor design on TcTS. For this feature, we wanted to avoid parts of the active site that were solvent exposed and near the edge of the active site to avoid non-specific interactions and competition with the solvent. Feature A3 was chosen because it was in a back pocket of the active site and contained a key target residue—Arg93. This pocket is mostly hydrophobic except for the Arg93 residue, and is known to bind the *N*-acetyl group of the substrate.⁴⁶ From our understanding, Arg93 has never been targeted for drug design and thus has the ability to aid in the discovery of novel inhibitors against TcTS. It would have been easy to choose the four features with the highest score from Phase, but we believe that the four sites chosen here (N10, R20, N17, and A3—Figure 3-1) are logical choices based on the available information on TcTS.

3.3.1.2 Holo pharmacophore model

The holo pharmacophore model was generated using a TcTS structure from clustering of an MD simulation with TcTS bound to sialyl-lactose. The pharmacophore sites found by Phase for this model are described in Table 3-2. There are many similarities between this model and the 1S0I hypothesis described above. Once again, the top-ranking site was a negative ionizable feature (N39) interacting with the arginine triad. Other sites from the holo pharmacophore model that were very similar to the 1S0I model include N38, D41, A34, D55, and R137. Given that the 1S0I and holo pharmacophore models were generated using different conformations of TcTS, differences in the features were anticipated. The major differences included the presence of features interacting with Trp120, Arg145 and Ser57 only in the holo pharmacophore, and the presence of sites interacting with Ala59, Asn60, Ser229,

Gln195, Gln362, and Asp247 only in the 1S0I hypothesis. However, several of these differences are considered irrelevant because the TcTS residues involved are mobile and/or solvent exposed, such as Ser57, Asn60, and Gln362.

When selecting four sites as required matches when screening possible inhibitors, we wanted to choose sites that were different from the 1S0I model, but still maintained at least two features known as important for substrate binding to TcTS.

Table 3-2. Pharmacophore sites of the TcTS holo model. Sites marked with an asterisk (*) were required matches during screening. Feature labels were given by Phase, with the letter indicating the type of interaction (N = negative ionizable, A = hydrogen acceptor, D = hydrogen donor, and R = aromatic ring). The scores are Glide XP summed energy values in units of kcal/mol.

Rank	Feature Label	Score	Interaction with TcTS
1	N39*	-10.43	Arginine triad (Arg35, Arg245, and Arg314)
2	N38*	-3.72	Arg53
3	A5	-3.27	Arg245
4	N117	-2.87	Backbone –NH of Trp120
5	A33	-1.77	Trp120 backbone –NH and side chain –OH of Thr121
6	D41	-1.60	Side chain oxygen of Tyr342
7	A8	-1.59	Side chain –OH of Tyr342
8	A17	-1.49	Side chain –NH of Trp120
9	R45*	-1.37	Hydrophobic pocket near Val95 and Leu176
10	A34	-1.31	Arg35
11	D80	-1.00	Backbone carbonyl oxygen of Ser57
12	D55	0.80	Asp96
13	D20*	-0.71	Asp96 and Glu230
14	R137	-0.48	Acceptor binding site between Tyr119 and Trp312
15	R136	-0.39	Tyr119

The differences were to ensure the same hit compounds were not discovered while screening compounds against the pharmacophore models. Thus, features N39 and N38 were chosen as the two sites that were consistent from the 1S0I model, while the acceptor binding site feature was excluded to increase hit compound diversity. As mentioned previously, the hydrophobic pocket near Val95 and Leu176 contains the *N*-acetyl group of the substrate, but Phase found the pocket large and favorable enough to accommodate an aromatic ring. This finding is contrary to most TcTS drug design efforts that typically try to place a similar *N*-acetyl group at this location, so we selected R45 to represent this region for screening.

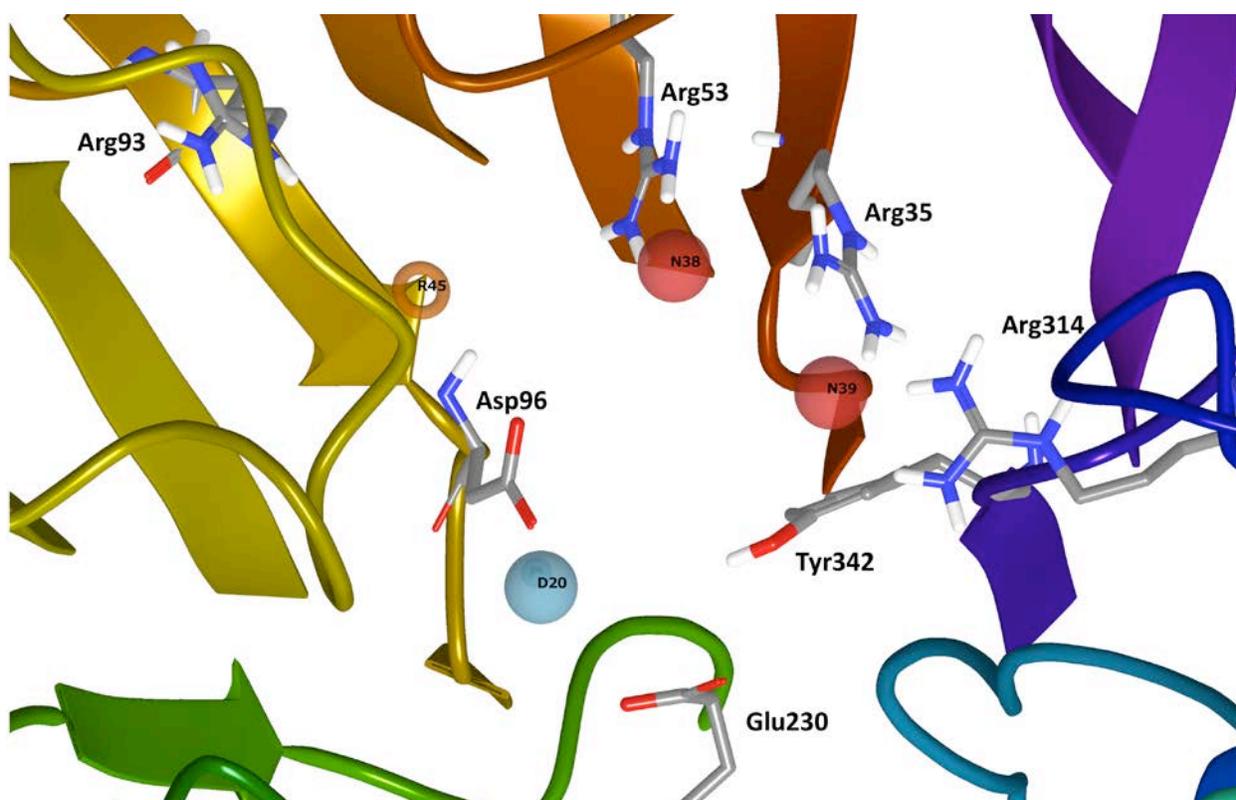


Figure 3-2. The required sites for the holo pharmacophore model used for screening lead compounds. Features N39 and N38 represent favorable interaction sites with Arg35/Arg314 and Arg53, respectively. The aromatic feature R45 is located in the hydrophobic region near Val95 and Leu176, while the hydrogen bond donor D20 feature is positioned to form interactions with the Asp96/Glu230 motif.

Finally, feature D20 was chosen as a new region of the active site to target.

Investigations of the TcTS structure near feature D20 discovered that this region is an ideal location for a hydrogen donor because of the presence of nearby negatively charged residues Asp96 and Glu230. Furthermore, Glu230 is important to the catalytic mechanism as a proton acceptor for the nucleophilic Tyr342⁵⁴ and stabilizing the substrate transition state,⁵⁶ which makes it an ideal drug design target residue. To the best of our knowledge, our screening approach is novel because no one has attempted placing an aromatic ring in the hydrophobic pocket near Val95/Leu176 or targeted the Asp96/Glu230 motif when designing drugs against TcTS, which makes these four features—N39, N38, R45, and D20 (Figure 3-2)—ideal for screening new inhibitors.

3.3.1.3 Apo pharmacophore model

The apo conformation of TcTS was the most unique of the three studied because of the flexibility of the unliganded enzyme.⁵⁹ This was the only pharmacophore model that did not include a negative ionizable site interacting with the arginine triad as the top-ranking feature (Table 3-3). Instead, a negative ionizable charge interacting with the side chain of Arg93 was the most favorable site (N136). The negative site interacting with the arginine triad (N132) unexpectedly ranked fifth among sites for the apo model. Asp59 and Glu227 are the only residues involved in an interaction with a pharmacophore site in the apo model that was not present in either of the other two models. Given the role of Asp59 as an acid/base catalyst for TcTS, it is interesting to note that the other pharmacophore models did not produce any sites interacting with the side chain of Asp59, while the apo model did. However, in the apo conformation, Asp59 is unlikely to be a residue of high interest to target for drug design because it is almost completely solvent exposed. Expectedly, Phase found no pharmacophore site in the

acceptor binding site between Tyr119 and Trp312 because the flexibility of the Trp312 loop precludes an aromatic ring from binding to this site in this conformation.⁵⁹ The re-ranking and scoring of several similar sites among the three pharmacophore models further enhances the idea that the pure rankings should not be the ultimate deciding factor when considering what pharmacophore sites to select as required matches during the screening process.

Table 3-3. Pharmacophore sites of the TcTS apo model. Sites marked with an asterisk (*) were required matches during screening. Feature labels were given by Phase, with the letter indicating the type of interaction (N = negative ionizable, P = positive ionizable, A = hydrogen acceptor, D = hydrogen donor, and R = aromatic ring). The scores are Glide XP summed energy values in units of kcal/mol.

Rank	Feature Label	Score	Interaction with TcTS
1	N136	-5.19	Arg93
2	P21*	-4.37	Asp96 and Glu230
3	A20	-2.60	Backbone –NH of Asp96
4	A3	-2.18	Backbone –NH of Trp120
5	N17*	-2.04	Arginine triad
6	N133	-1.89	Arg53 and Arg35
7	A15	-1.65	Arg245 and side chain –OH of Tyr342
8	D65	-1.60	Side chain -OH of Tyr342
9	D70	-1.60	Side chain oxygen of Tyr119
10	D77	-1.60	Side chain of protonated oxygen on Asp59
11	D82	-1.60	Backbone carbonyl oxygen of Asp247
12	D83	-1.60	Backbone carbonyl oxygen of Asp247
13	A37	-1.55	Side chain –OH of Ser229
14	R22*	-1.42	Hydrophobic pocket near Val95 and Leu176
15	D64	-1.20	Glu230

The choice of required matches for screening became more complicated for the apo pharmacophore model. The negative ionizable site representing the ligand interaction with the arginine triad (N17) was the first required site chosen. The second feature chosen was P21 because it represented a positively charged group interacting with two negatively charged residues—Asp96 and Glu230—that seemed extremely favorable as mentioned for the holo pharmacophore. Originally, we also chose sites N136 and R22 as the last two required matches. Both were used in screening of previous pharmacophore models, but never at the same time. The two sites represent regions of space in the pocket originally filled by the *N*-acetyl group of sialyl-lactose.

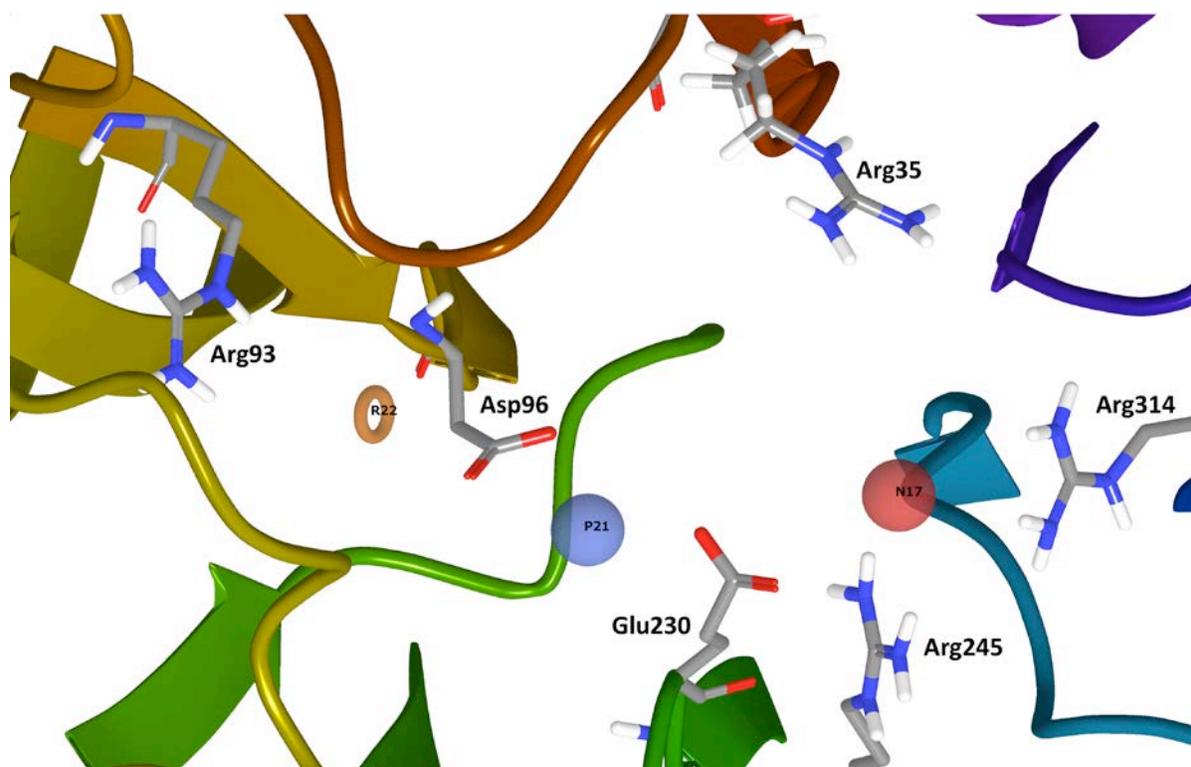


Figure 3-3. The required sites for the TcTS apo pharmacophore model used for screening lead compounds. N17 forms interactions with the arginine triad, while the positive charged feature P21 interacts with the Asp96/Glu230 motif. The aromatic feature R22 is positioned in the hydrophobic pocket near Val95 and Leu176.

Choosing both sites allowed us the ability to screen for compounds that were specifically suited for the sialic acid binding site and the *N*-acetyl binding region. Unfortunately, using these four sites—N17, P21, N136, and R22—resulted in no hit compounds during screening of the leads database. As Demir *et al.*⁵⁹ noted, the active site of apo TcTS is much larger and more open than the holo structure, and thus any hit compounds from screening would need to be larger to match the sites in this active site compared to the other pharmacophore models. Large molecules often do not make ideal drugs and the ZINC Leads database screened here primarily consists of smaller lead scaffolds, not large molecules. Consequently, we trimmed the required matches down from four to three and attempted using both sites as the third and final required site. The only combination that resulted in any hit compounds was using N17, P21, and R22 (Figure 3-3). Thus, these three sites were chosen as the required features for the screening results discussed below for the apo pharmacophore model.

3.3.2 Screening Results Using TcTS Pharmacophore Models

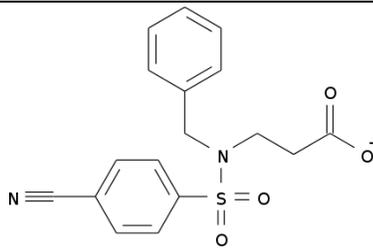
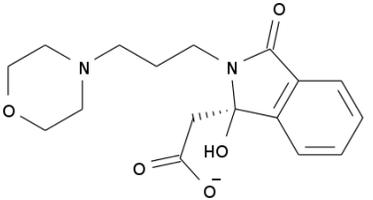
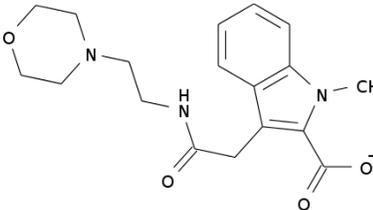
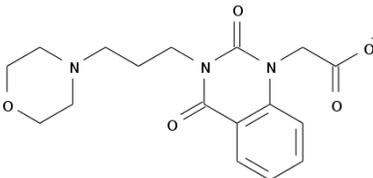
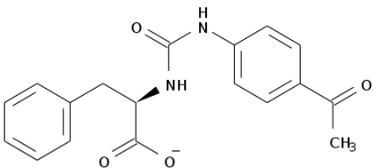
This section provides the virtual screening and binding free energy results from MD for the hit compounds found by Phase using the pharmacophore models described above.

3.3.2.1 Screening results for the 1S0I pharmacophore

Screening of the ZINC Leads database using the 1S0I pharmacophore resulted in 25 hit compounds. After performing MD and calculating the MM-GBSA binding free energy for each compound, the potential inhibitors were re-ranked and the best five compounds are shown in Table 3-4 (see Table A-1 for all compounds). For the Phase Fitness scores, the higher value the better the compound fits the shape and topology of the pharmacophore model. Conversely, the more negative the MM-GBSA binding free

energy the better the compound is expected to bind to TcTS. Aside from the compounds, Table 3-4 makes it clear why MD and MM-GBSA binding free energy calculations were necessary—there is no correlation between the Phase Fitness score and the MM-GBSA binding free energies.

Table 3-4. The best five compounds from screening results of the 1S0I pharmacophore according to MM-GBSA binding free energies. The Phase Fitness scores are unitless, while the MM-GBSA have units of kcal/mol. The uncertainties reported represent the standard error of the mean. See Table A-1 for the results of all 25 hit compounds. The binding mode of ZINC13359679 is discussed in more detail in Section 3.3.3.1.

Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
1	ZINC13359679		0.184	-76.3 ± 0.4
2	ZINC23187566		0.173	-58.0 ± 0.7
3	ZINC20107259		0.187	-56.4 ± 0.8
4	ZINC73658621		0.177	-53.5 ± 0.6
5	ZINC06614928		0.199	-49.2 ± 1.1

This observation is consistent for all hit compounds found by Phase from the 1S0I pharmacophore as well as the holo and apo (see Appendix A.1—Table A-2 and Table A-3). Because the pharmacophore screening includes no specific details of the receptor it would be unrealistic to expect the Phase Fitness to correspond with the MM-GBSA binding energies that are calculated as an average over 200 conformations of the compound noncovalently bound to the active site of TcTS. For drug design and discovery, Phase Fitness scores should be disregarded and replaced by a more sophisticated and rigorous binding affinity calculation that is known to accurately predict relative free energy differences, such as MM-GBSA.¹⁸⁷

As all of these compounds are derived from the 1S0I pharmacophore, they have some structural similarities because of the required features imposed during the screening process. Per the pharmacophore model, every hit compound contains a negative ionizable functional group (typically a carboxylic acid) that interacts with the arginine triad, a hydrogen acceptor or negative ionizable group that interacts with Arg53, a hydrogen acceptor interacting with Arg93, and an aromatic ring designed to fit into the acceptor binding site between Tyr119 and Trp312. Overall, the 1S0I pharmacophore model resulted in the best hit compounds of the three pharmacophore models according to the binding energies, with four compounds with binding free energies better than -50 kcal/mol, and ten compounds better than -40 kcal/mol (once again, these are not considered real binding free energies and are only used to compare relative to one another). The better TcTS inhibitors likely arise from the 1S0I model because of the pharmacophore sites chosen as requirements during the screening process. The 1S0I pharmacophore was the only hypothesis that required hit

compounds to have a presence in both the sialic acid and acceptor binding sites of TcTS, which is believed to yield more specific TcTS inhibitors than inhibitors that simply target the sialic acid binding site. The results here further support the idea that a strong TcTS inhibitor should target both of these sites.

The MM-GBSA binding free energies reported here are relative to one another and are used to compare directly to the calculated binding free energy of the TcTS substrate, sialyl-lactose. MM-GBSA calculations performed on the MD simulation of sialyl-lactose bound to TcTS by Demir *et al.*⁵⁹ determined that the average binding free energy was -54.8 ± 0.1 kcal/mol for sialyl-lactose using the same protocol reported here for the hit compounds. Thus, the three most promising hit compounds discovered using the 1S0I pharmacophore—ZINC13359679, ZINC23187566, and ZINC20107259—are projected to bind TcTS stronger than the natural substrate. ZINC13359679 is calculated to have a significantly stronger binding affinity (-76.3 kcal/mol) than the substrate (-54.8 kcal/mol), and thus the MD of this compound will be analyzed in more detail in Section 3.3.3.1. As is, ZINC23187566, ZINC20107259, and ZINC73658621 are likely to compete with sialyl lactose binding *in vivo*, but are not likely to be much better than the substrate. However, all of these compounds are good scaffolds for drug design efforts against TcTS because their structures are unique compared to any previously reported TcTS potential inhibitors. Modifications could be made to optimize each of the hit compounds in Table 3-4 to improve their binding affinity for TcTS.

It is important to note that unlike for the screening process, the binding energies are averages based on a dynamic process. Many of the hit compounds are initially in a single, ideal conformation within the TcTS binding pocket from screening that is not

maintained during the MD simulation. Instead, the binding free energy often gets worse during the simulation as the contacts between the protein and ligand relax until the ligand reaches a state of equilibrium with the protein that results in a binding free energy that oscillates around a single value. This natural worsening of the binding free energy during simulations is a result of the screening process only producing a single ligand conformation and that ligand conformation is not even directly based on the protein conformation.

3.3.2.2 Screening results for the holo pharmacophore

Screening of the ZINC Leads database using the holo pharmacophore resulted in 30 hit compounds. After performing MD and calculating the average MM-GBSA binding free energy for each compound, the potential inhibitors were re-ranked and the best five compounds are shown in Table 3-5 (see Table A-2 for all compounds). The hit compounds from the holo pharmacophore screening are all unique compared to the compounds discovered from the 1S0I pharmacophore, naturally because the selected required sites and locations for the two pharmacophores were different. Of course, this is ideal for the purposes of discovering new inhibitors as it broadens the search criteria for novel ligands.

These compounds all have a set of common features that can align to the required pharmacophore sites of the holo e-pharmacophore. Each compound has a negative ionizable group positioned to interact with the arginine triad, a hydrogen acceptor or negative ionizable group positioned to hydrogen bond with Arg53, an aromatic ring to fill the hydrophobic pocket near Val95 and Leu176, and a hydrogen donor group that can form interactions with the side chains of Asp96 and/or Glu230.

Table 3-5. The best five hit compounds from screening results of the holo pharmacophore according to MM-GBSA binding free energies. The Phase Fitness scores are unitless, while the MM-GBSA energies have units of kcal/mol. The uncertainties reported represent the standard error of the mean. See Table A-2 for the results of all 30 hit compounds. The binding mode of ZINC02576132 is discussed in more detail in Section 3.3.3.2.

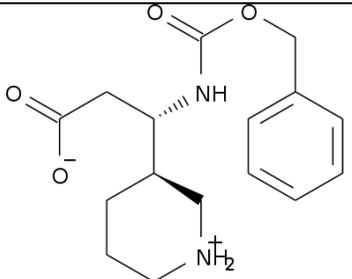
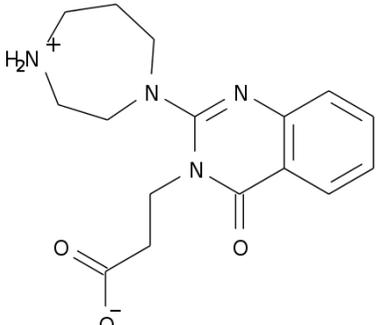
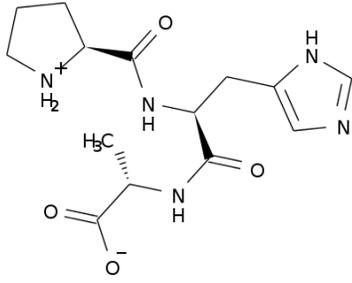
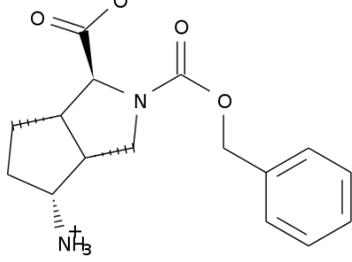
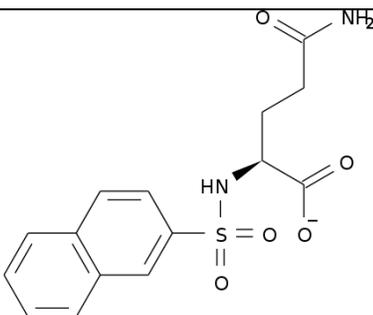
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
1	ZINC02576132		0.194	-73.2 ± 0.5
2	ZINC40351842		0.145	-55.1 ± 0.4
3	ZINC60181779		0.198	-52.2 ± 0.4
4	ZINC19795370		0.184	-48.5 ± 1.0

Table 3-5. Continued

Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
5	ZINC03406782		0.197	-45.0 ± 0.5

Overall, the hit compounds from the holo pharmacophore are not as strong of inhibitors compared to the compounds from the 1S0I pharmacophore screening. Only two compounds—ZINC02576132 and ZINC40351842—resulted in binding energies that were better than sialyl-lactose, while the other 28 hit compounds were worse. Similar to the results from the 1S0I pharmacophore screening, one compound, ZINC02576132, stands out as significantly better than sialyl-lactose and the other hit compounds with a binding free energy of -73.2 kcal/mol. For this reason, the MD and binding mode of ZINC02576132 will be analyzed in more detail in Section 3.3.3.2. The free energy of binding results suggest that ZINC40351842 and ZINC60181779 have a similar binding affinity as the substrate, and could compete for binding *in vivo*, while the other hit compounds are unlikely to be anything but very weak inhibitors of TcTS. However, rational drug design efforts could use these compounds as new scaffolds for modification and optimization that improve their TcTS binding affinities.

3.3.2.3 Screening results for the apo pharmacophore

Screening of the ZINC Leads database using the apo pharmacophore resulted in 27 hit compounds. After performing MD and calculating the average MM-GBSA binding

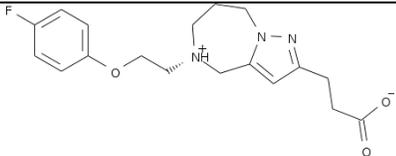
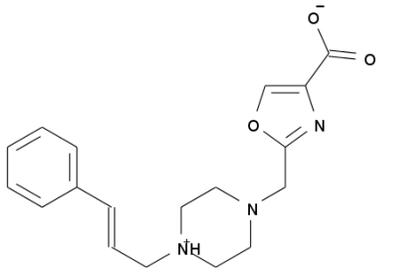
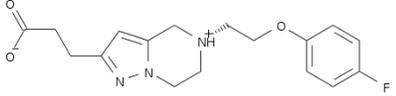
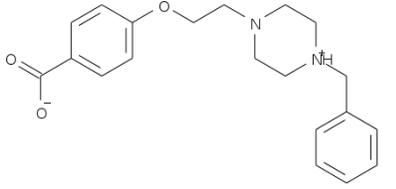
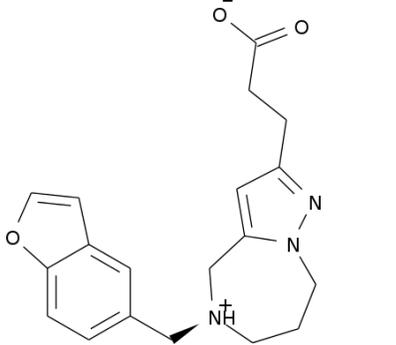
free energy for each compound, the potential inhibitors were re-ranked and the best five compounds are shown in Table 3-6 (see Table A-3 for all compounds).

The hit compounds from the apo pharmacophore screening were expected to be more chemically diverse compared to the 1S0I and holo pharmacophore results because only three pharmacophore sites were required for hit compounds instead of four like the 1S0I and holo models. Each hit compound was required to have a negative ionizable charge that could interact with the arginine triad, an aromatic ring for the hydrophobic pocket near Val95 and Leu176, and a positive ionizable group for hydrogen bonds with the side chains of Asp96 and Glu230. Unexpectedly, many of the hit compounds had very similar chemical structures, such as ZINC72447098, ZINC72432649, ZINC72477708, and ZINC72407055 (See Table A-3).

The likely cause of this result is the shape and topology of the apo pharmacophore required sites. The TcTS active site is known to be larger and more open in the apo conformation, which reduces the number of likely hit compounds because any potential hit compound must be large enough to accommodate the bigger active site. Furthermore, one of the required pharmacophore features included a positive ionizable group, which significantly reduces the number of compounds that could match this pharmacophore model.

Unlike the screening results from the 1S0I and holo pharmacophores, no hit compound from the apo pharmacophore screening was discovered with a binding free energy significantly better than sialyl-lactose. In fact, the best hit compound, ZINC72447098, had an average binding free energy identical to sialyl-lactose (-54.8 kcal/mol).

Table 3-6. The best five hit compounds from screening of the TcTS apo pharmacophore according to MM-GBSA binding free energies. The Phase Fitness scores are unitless, while the MM-GBSA energies are in units of kcal/mol. The uncertainties reported represent the standard error of the mean. See Table A-3 for the results of all 27 hit compounds.

Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
1	ZINC72447098		0.155	-54.8 ± 0.3
2	ZINC67946325		0.186	-50.1 ± 0.4
3	ZINC72145690		0.144	-44.0 ± 0.5
4	ZINC19850530		0.168	-42.0 ± 0.3
5	ZINC72432649		0.168	-41.8 ± 0.3

At best, ZINC72447098 would bind to TcTS with about the same affinity as sialyl-lactose, while all the other hit compounds would bind worse. Of course, these compounds should all be considered lead candidates for rational drug design against

TcTS, regardless of their relatively poor calculated binding free energies compared to the 1S0I and holo screening results. Specifically, the chemical scaffolds that are observed multiple times in Table A-3 should be considered for optimization given their prevalence as hit compounds discovered using the apo pharmacophore. The repeated occurrence of compounds similar to ZINC72447098 may indicate that this scaffold is viable as an inhibitor of TcTS with some optimization.

3.3.3 Analysis of the Most Promising TcTS Inhibitors

Phase found a total of 82 hit compounds from the ZINC Leads database using the three presented pharmacophore models, and MM-GBSA binding free energy calculations suggest that two of these compounds—ZINC13359679 and ZINC02576132—have significantly more affinity for TcTS than the natural substrate, sialyl-lactose. In this section we will analyze the MD simulations of these two compounds bound to TcTS.

3.3.3.1 ZINC13359679

ZINC13359679 was the top hit compound discovered using the 1S0I pharmacophore, with a MM-GBSA binding free energy of -76.3 kcal/mol (Table 3-4). The protein and ligand were found to be relatively stable during the 10 ns production trajectory, with an RMSD of less than 2.0 Å for most of the simulation (Figure 3-4). As the RMSD plot shows, neither the protein nor the ligand are undergoing any major conformational changes after the initial relaxation and equilibration steps, suggesting ZINC13359679 is stable within the active site of TcTS. The two spikes in RMSD near 1.5 ns and 4.3 ns are related to movements near the periphery of the enzyme and should not be associated with instability of the enzyme or active site.

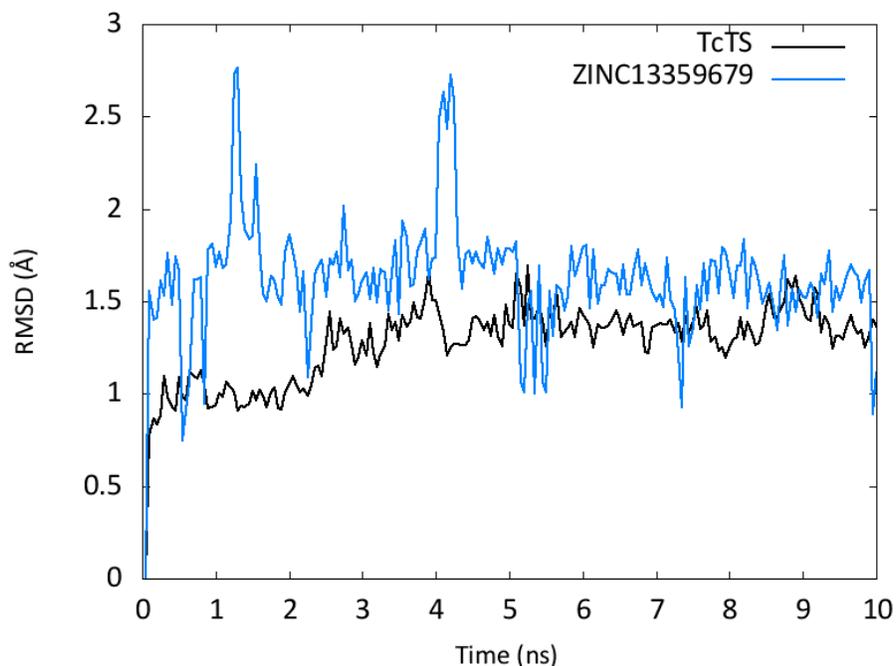


Figure 3-4. The RMSD of TcTS and ZINC13359679 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for ZINC13359679.

Since the binding free energy reported represents an average value of a dynamic property, it is also informative to examine the binding free energy as a function of time (Figure 3-5). The binding energy fluctuates as the simulation progresses, but the value oscillates around the average value and does not appear to be trending upward or downward. This plot indicates the stability of ZINC13359679 bound to TcTS in the conformation from pharmacophore screening. Another important observation is that the binding free energy is never worse than the average binding free energy of sialyl-lactose to TcTS (-54.8 kcal/mol) for the 200 frames analyzed, providing more optimism that ZINC13359679 would bind to TcTS better than the substrate *in vivo*.

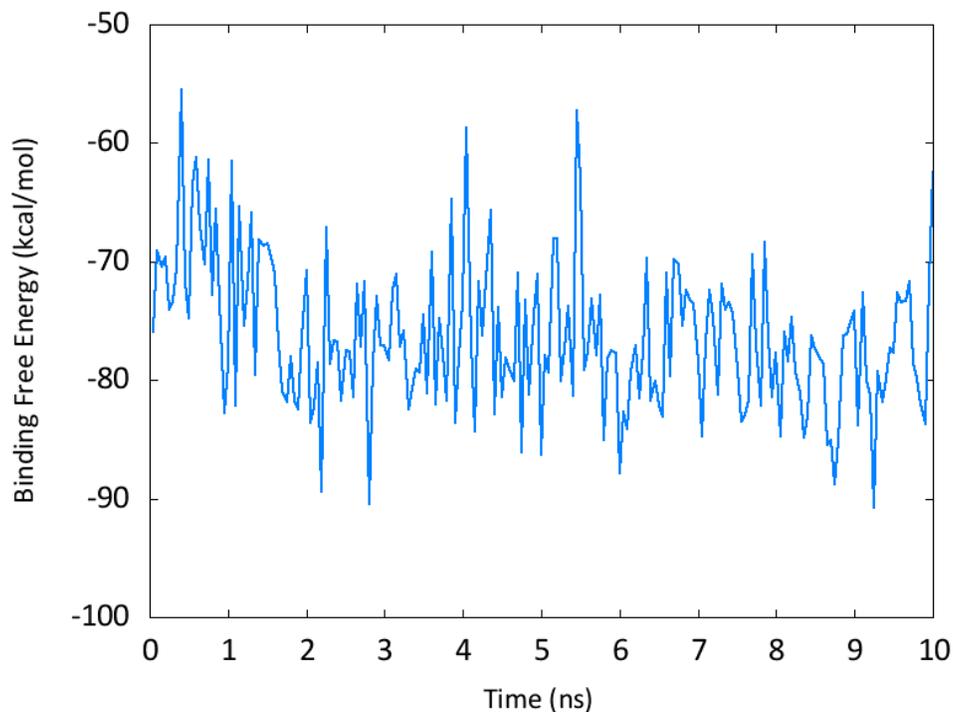


Figure 3-5. The MM-GBSA binding free energy of ZINC13359679 bound to TcTS as a function of time for a 10-ns MD simulation.

The contacts between TcTS and ZINC13359679 are a clear mixture of polar and nonpolar interactions (Figure 3-6). The strongest set of polar interactions naturally exists between the carboxylic acid and the arginine triad, which is also important for stabilizing the carboxylic acid of the substrate. In fact, hydrogen bonding analysis shows that hydrogen bonds exist between ZINC13359679 and the three residues that comprise the arginine triad (Arg35, Arg245, and Arg314) more than 80% of the simulation (Figure 3-6). Interactions with the arginine triad are a key component of any potential inhibitor of TcTS. The initial protein-ligand conformation from pharmacophore screening included the sulfate functional group forming hydrogen bonds with Arg53, and the cyano group hydrogen bonded to Arg93. However, relaxation of the ligand complexed with TcTS significantly reduced both of these interactions. The sulfate group is never close enough

to form a hydrogen bond with Arg53 during the simulation, but is likely close enough (~4.25 Å) to make favorable Coulombic interactions for the majority of the simulation. ZINC13359679 shifted slightly during relaxation to a conformation with the cyano group of the ligand hydrogen bonding to the side chain of Thr121 for approximately 13% of the simulation instead of Arg93.

The dominating nonpolar protein-ligand interactions involve π -stacking between aromatic rings on the ligand and three TcTS residues—Phe58, Tyr119, and Trp312 (Figure 3-6). The 1S0I pharmacophore contained a required feature involving an aromatic ring in the acceptor binding site between Tyr119 and Trp312, but the π -stacking interaction with Phe58 was unexpected. In the initial conformation, Phe58 is not involved in any protein-ligand interactions, but ~6.7 ns into the simulation, the side chain of Phe58 changes conformation and forms a T-shaped π -stacking interaction with a phenyl ring on ZINC13359679. In this conformation, Phe58 is positioned over the active site like a lid preventing the ligand from exiting. Trp312 is also making a T-shaped π -stacking interaction with the same phenyl ring of ZINC13359679 as Phe58, while Tyr119 forms a parallel π -stacking interaction with the other ligand aromatic ring. Phe58 has no known role in substrate binding, and is unlikely to form a similar interaction with the substrate given that the lactose in the acceptor binding site is always attached to another sugar, preventing Phe58 from “closing” the top of the active site. ZINC13359679 forces Tyr119 and Trp312 of the acceptor binding site to deviate from their traditional conformation with sialyl-lactose bound. Typically, Tyr119 and Trp312 ‘sandwich’ a lactose acceptor molecule, which means they are forming nonpolar interactions with the same sugar ring, but with ZINC13359679 they are π -stacking with

different aromatic rings. This is a reasonable observation given the known flexibility of the Trp312 loop on TcTS.⁵⁹ Although ZINC13359679 did not maintain all of the interactions originally described by the 1S0I pharmacophore model, it was able to form many favorable polar and nonpolar interactions with TcTS that make ZINC13359679 a strong candidate to inhibit TcTS.

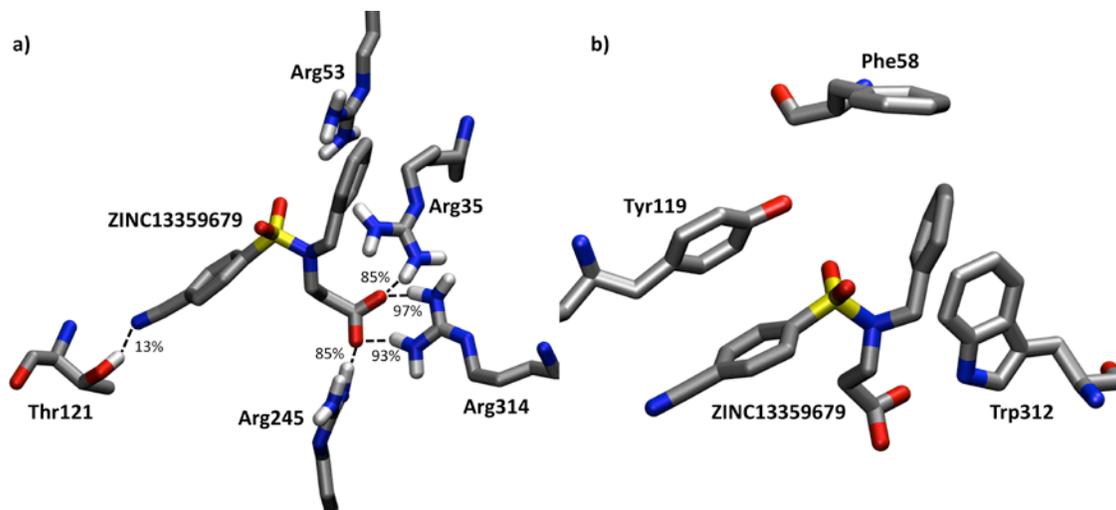


Figure 3-6. The protein-ligand interactions between TcTS and ZINC13359679. Specifically, the a) polar and b) nonpolar interactions. Percentages correspond to the time a hydrogen bond was present during the MD simulation. For clarity, all hydrogen atoms have been removed except for those involved in hydrogen bonding.

3.3.3.2 ZINC02576132

ZINC02576132 was the top hit compound discovered using the holo pharmacophore, with a MM-GBSA binding free energy of -73.2 kcal/mol (Table 3-5). During dynamics both TcTS and ZINC02576132 were very stable, producing RMSD values less than 1.0 Å (Figure 3-7), suggesting that neither the protein nor the ligand undergo any major conformational changes during the 10-ns simulation. The complex appears much more stable compared to TcTS in complex with ZINC13359679 (Figure 3-4), which had RMSD values closer to 2.5 Å. The lower structural fluctuations of

ZINC02576132 compared to ZINC13359679 is an indication that ZINC02576132 is more stable in the TcTS active site.

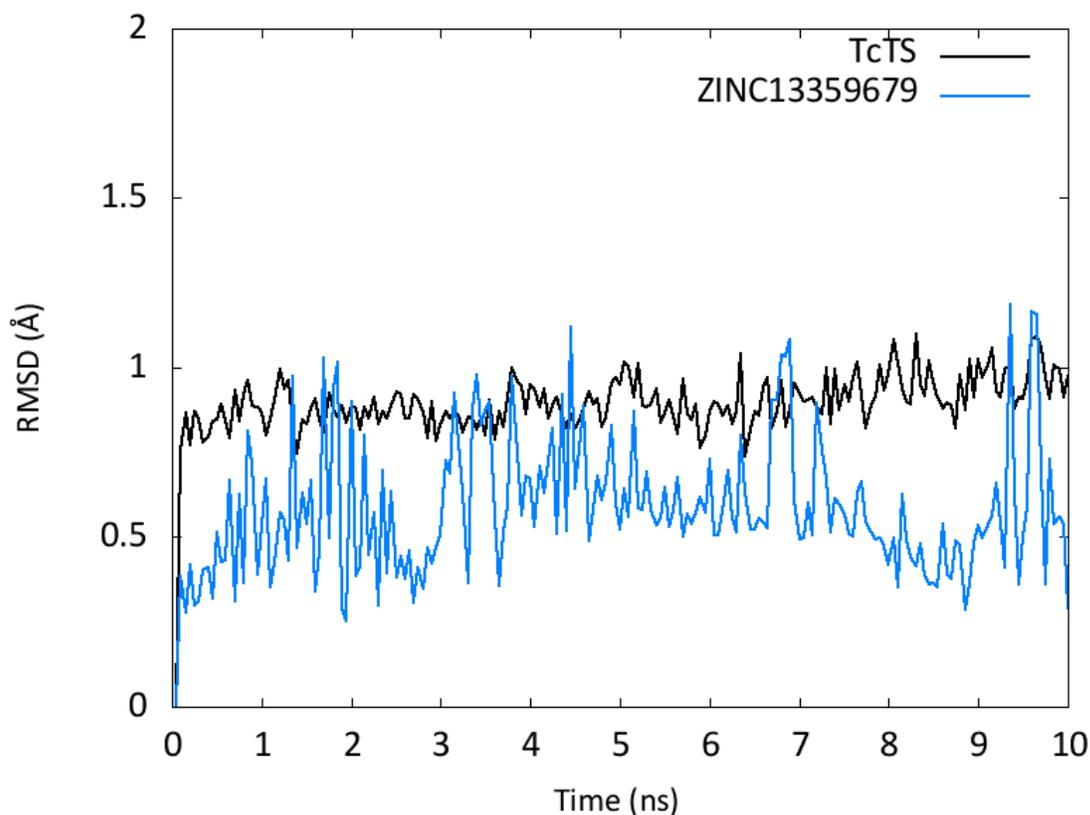


Figure 3-7. The RMSD of TcTS and ZINC02576132 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for ZINC02576132.

We also investigated the MM-GBSA binding free energy as a function of time to check for any upward or downward trends in the binding affinity (Figure 3-8). The binding energy fluctuates approximately the same amount as for ZINC13359679, and there is no large movement upwards or downwards that suggests the binding energy is trending to a different average energy than the one reported. Furthermore, once again the binding free energy is consistently better than sialyl-lactose (-54.8 kcal/mol), which is desirable for any potential inhibitor of TcTS.

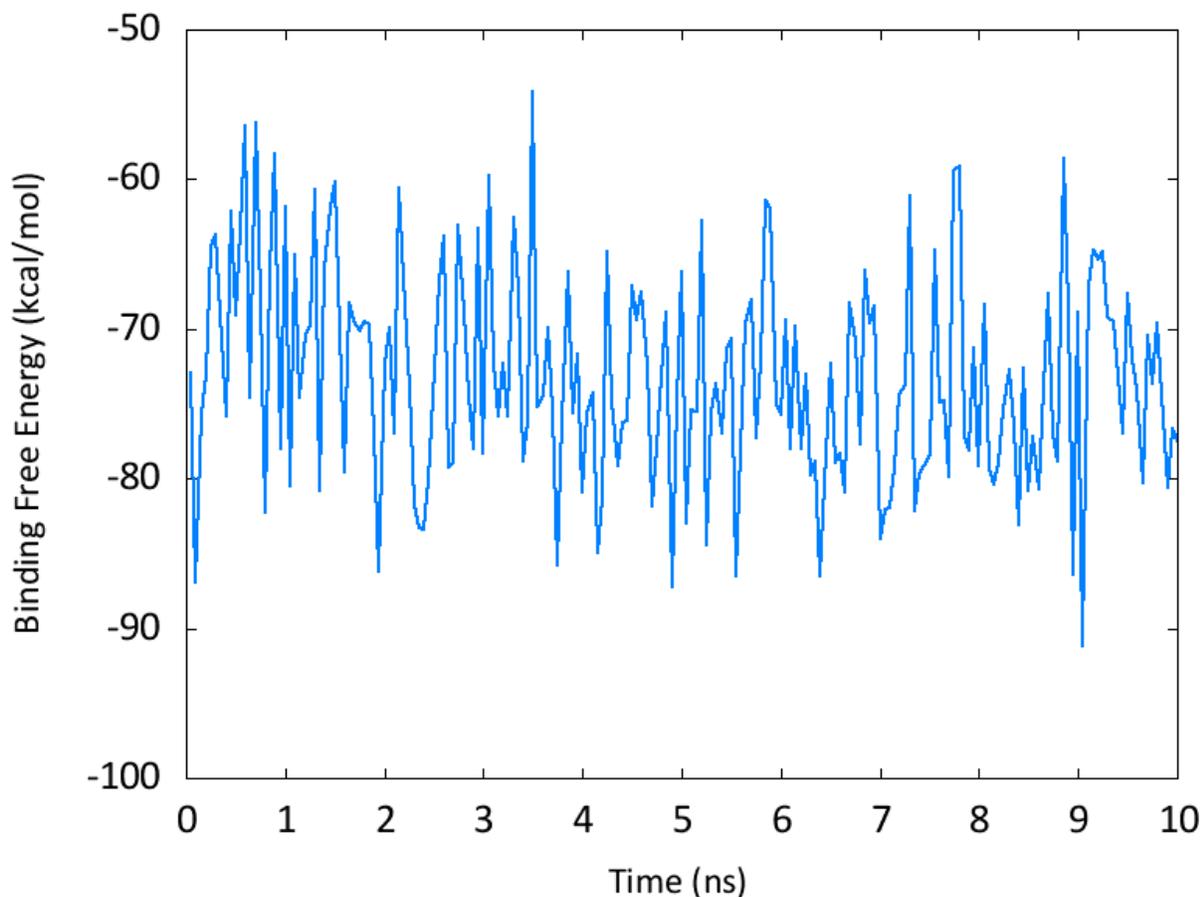


Figure 3-8. The binding free energy of ZINC02576132 bound to TcTS as a function of time for a 10-ns MD simulation.

Unlike ZINC13359679, the interface between ZINC02576132 and TcTS appears to be dominated by polar interactions (Figure 3-9). The nonpolar interactions present are simple van der Waals contacts with no real possibilities for protein-ligand π -stacking since the ZINC02576132 phenyl group is located in the hydrophobic pocket near Val95 and Leu176, not in the acceptor binding site between Tyr119 and Trp312. The hydrogen bonds between the ligand carboxylate and the arginine triad do not have as high occupancy as ZINC13359679 (all more than 80% for ZINC13359679; none more than 54% occupied for ZINC02576132), but this is compensated for by increased hydrogen bonds with other protein residues. For example, Arg53 is ideally positioned to form two

hydrogen bonds (66% and 44% occupied) with the carbonyl oxygen on ZINC02576132. Likely more important is the -NH_2^+ group of the ligand that forms a total of three hydrogen bonds with the Asp96/Glu230 motif, including a hydrogen bond with an occupancy over 90% with Asp96 (the most sustained TcTS-ZINC02576132 hydrogen bond).

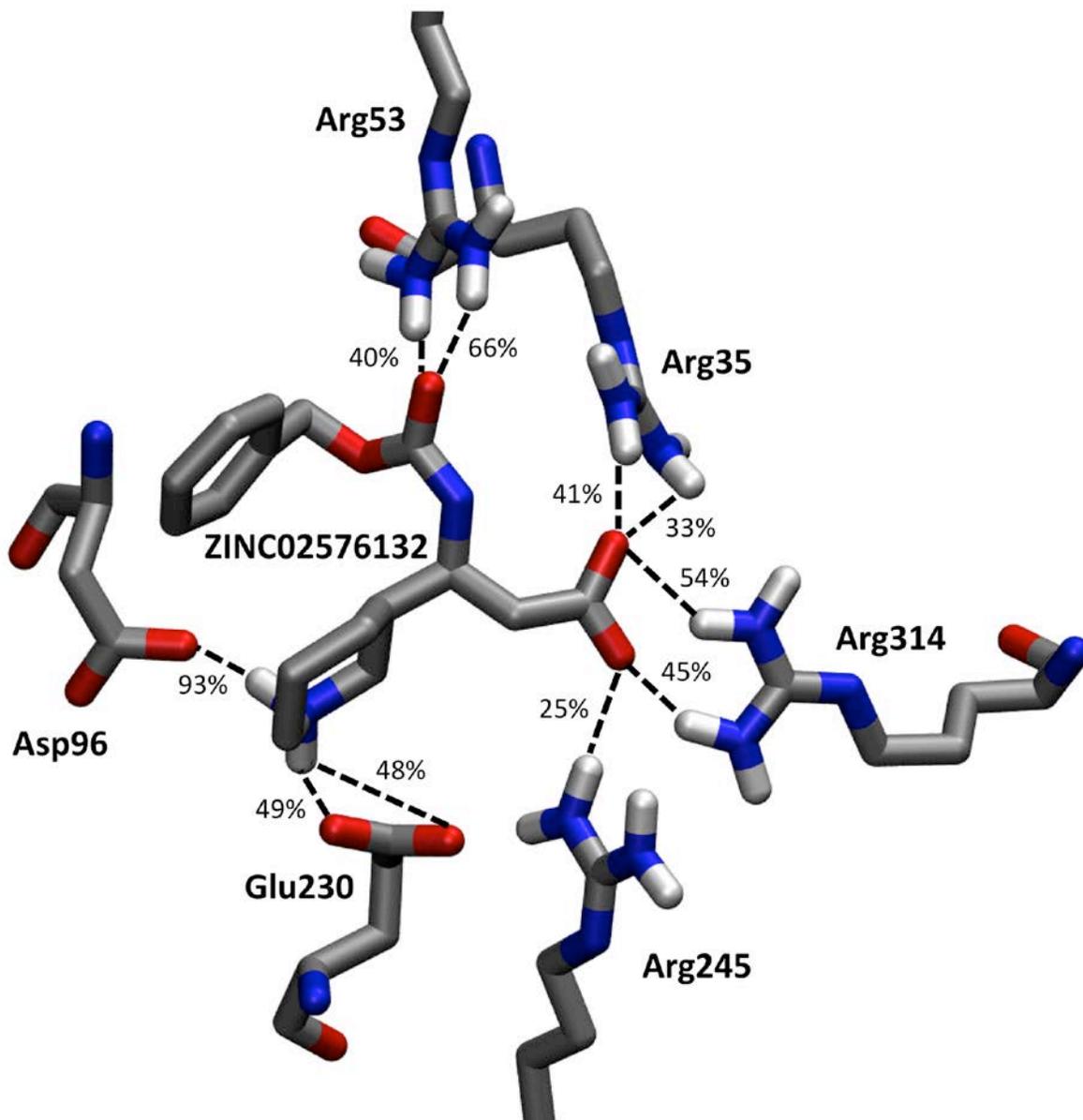


Figure 3-9. The main interactions between TcTS and ZINC02576132 observed during a 10-ns MD simulation. The percentages represent the time a hydrogen bond was present during the dynamics.

The holo pharmacophore specifically required a hydrogen donor be present at this location for any hit compounds with the intent of interacting with this motif. The important role that Glu230 plays in acid/base catalysis combined with the close proximity of Asp96 makes this area of the active site a key target for drug design, and ideally suited to accommodate a positive ionizable group from an inhibitor (such as the -NH_2^+ group of ZINC02576132). The strong binding free energy observed for ZINC02576132 provides further evidence that the Asp96/Glu230 motif could be a key to designing low nM inhibitors of TcTS. Although ZINC02576132 is stable in the active site and has a strong binding affinity for TcTS, key optimization could still make it a better inhibitor. Specifically, a clever addition of a phenyl group near the carboxylate could allow ZINC02576132 to also form π -stacking interactions with Tyr119 and Trp312 in the important acceptor binding site.

3.4 Concluding Remarks

In this work, we wanted to create pharmacophore models detailing the important residues for ligand binding to TcTS and utilize that information to screen for novel inhibitors. To our knowledge, no pharmacophore models of TcTS have been created previously due to a lack of known active molecules. But the method recently proposed by Loving *et al.*¹⁶² allows for the generation of pharmacophore models without *a priori* knowledge of actives by using the energetic results of docked fragments to the protein. We created three e-pharmacophore hypotheses using three different conformations of TcTS (PDB 1S0I, holo, and apo) each with 15 unique features. To our knowledge, these are the first pharmacophore models ever reported for TcTS and provide a detailed description of locations important for drug inhibition.

Different combinations of three or four sites from each pharmacophore were rationally chosen as required features while screening ZINC's Leads database for novel inhibitor scaffolds. The required sites included TcTS residues or motifs that were deemed important to TcTS function and substrate binding based on the pharmacophore models and available literature: Arg53, Arg93, Asp96, Tyr119, Glu230, Trp312, the hydrophobic pocket near Val95/Leu176, and the arginine triad (Arg35, Arg245, and Arg314). Screening of over 4 million lead compounds resulted in 82 hit compounds, which were all simulated using MD and analyzed using MM-GBSA free energy calculations to directly compare the relative binding affinities of each with sialyl-lactose. Although many of the hit compounds are considered novel lead scaffolds, two compounds—ZINC13359679 and ZINC02576132—had significantly better average binding free energies than sialyl-lactose. The MD simulations were analyzed for both compounds to determine the binding modes and important protein-ligand interactions. These molecules are the newest compound scaffolds proposed to inhibit TcTS in the search for suitable drugs to combat Chagas' disease.

Future work will focus on using the pharmacophore models developed here to screen other databases containing drug-like compounds to discover more unique hit compounds as potential TcTS inhibitors. Finally, we plan to obtain these hit compounds from vendors and perform kinetic assays to calculate *in vitro* inhibition constants against TcTS.

CHAPTER 4
COMBINED MOLECULAR DOCKING, MOLECULAR DYNAMICS AND BINDING FREE
ENERGY CALCULATIONS TO DESIGN AND DISCOVER NOVEL INHIBITORS FOR
TRYPANOSOMA cruzi TRANS-SIALIDASE

4.1 Preliminaries

Virtual screening (VS), the process of testing large numbers of compounds for inhibitory activity *in silico*, has been useful for the discovery of new drugs. Specifically, active inhibitors have been discovered using VS on enzymatic targets such as farnesyltransferases,²⁰⁷ carbonic anhydrase,²⁰⁸ cysteine proteases,²⁰⁹ and trihydroxynaphthalene reductase.²¹⁰ VS typically involves molecular docking to screen a database of compounds. Because molecular databases can contain millions of compounds, VS methods must be quick at determining if a molecule has the potential to be active against an enzyme. As a result, VS methods often suffer from a lack of accuracy at predicting and ranking binding affinities for a set of compounds. For these reasons, VS is usually combined with other more rigorous methods, such as docking with more sophisticated search and scoring algorithms, *de novo* design, molecular dynamics, and free energy of binding calculations.

Schrödinger's Glide molecular docking program contains three different search and scoring algorithms—HTVS, SP, and XP—that vary in their speed and accuracy. Glide HTVS is the computationally cheapest and least robust docking method available in Schrödinger and is ideal for VS studies intended to remove unrealistic compounds as inhibitors. In most cases, the more promising compounds from HTVS docking should be considered for Glide SP, which is significantly more accurate than HTVS at distinguishing active compounds from inactive ones.¹⁵² After SP docking, Glide XP can be used to re-dock and re-score the best compounds in order to help reproduce

experimental protein-ligand binding modes and affinities.¹⁵¹ However, Glide XP is not required in this process if more sophisticated techniques will be employed, such as MD.

Enzymes are not rigid structures *in vivo*, so it would be unrealistic to expect a single protein-ligand conformation from docking to fully describe all the binding modes of an inhibitor to its receptor. A trajectory generated from MD creates an ensemble of conformations that provides much more detail about the binding mode of an inhibitor than a single docked structure. Furthermore, these snapshots can be used to calculate the free energies of binding for multiple inhibitors using more rigorous end-state calculated, such as Molecular Mechanics/Generalized Born Surface Area (MM-GBSA).¹⁸⁸ The combination of molecular docking and molecular dynamics has become a powerful tool for computational drug design and discovery.¹⁹⁷

In this chapter, we use a VS method that combines molecular docking, *de novo* design, molecular dynamics, and free energy of binding calculations to discover new potential inhibitors of *Trypanosoma cruzi* trans-sialidase (TcTS). Millions of compounds have been docked using Schrödinger's Glide program, with MD performed on the most promising drug candidates using Amber. *De novo* design techniques have also been applied to several of these molecules in an attempt to optimize their binding to the TcTS active site. MM-GBSA free energies of binding of each potential inhibitor have been calculated using snapshots from the corresponding MD simulations. The binding modes of the top inhibitor candidates from MD are reported in more detail.

4.2 Methods

In general, the methodology described in this chapter is similar to the workflow depicted in Figure 2-5. The details of each step are laid out in more detail in the following sections.

4.2.1 Acquisition of Compounds

4.2.1.1 Molecular libraries

Over 35 million structures were obtained from the ZINC¹⁹⁸ website (<http://zinc.docking.org>) from 25 different libraries. A list of the molecular databases downloaded and used in this study along with their corresponding molecule counts can be found in Table 4-1.

Table 4-1. List of ZINC molecular libraries and their corresponding number of compounds used for virtual screening.

Molecular Library	Number of Compounds
ChemBridge	889,605
ChemDiv	1,020,993
Chemical Block	125,361
Clean Drug-like	10,384,763
Clean Fragments	504,074
Clean Leads	5,135,179
DrugBank Experimental	5,046
DrugBank Nutraceuticals	124
DrugBank Small Molecules	1,408
Enamine	1,698,842
Innovapharm	329,080
Natural Products	93,478
National Cancer Institute (NCI)	329,394
NCI Diversity Set II	41,674
Otava	193,226
PBMR Labs	480,244
Pharmeks	259,523
Princeton Biomolecular Research	798,402
PubChem	10,080,211
Scientific Exchange	47,827
Sigma Aldrich	77,614
Specs	206,615
TimTec	184,675
UORSY	1,190,916
VitasM	1,180,514
Total	35,258,788

Libraries obtained from other sources were also screened for potential inhibitors, including cHEMBL (3,027,553 compounds), Asinex Gold (177,463 compounds), and a

collection of compounds maintained at the University of Florida by Dr. Alan Katritzky (31,754 compounds). Structure similarity searches on the PubChem website (<http://pubchem.ncbi.nlm.nih.gov/>) were performed to help identify available derivatives of two promising compounds—ZINC16247520 and ZINC19328706.

4.2.1.2 *De novo* design

A few compounds were selected for optimization using *de novo* design methods with either LigBuilder¹⁴⁰ or AutoGrow.¹³⁸ These compounds included DANA, oseltamivir, sialic acid, ZINC16247520, and ZINC20472820. DANA and oseltamivir were chosen because of their ability to strongly inhibit related neuraminidase enzymes, with the idea that a *de novo* design technique might improve their binding affinity for TcTS.

ZINC16247520 and ZINC20472820 were chosen because they provided unique protein-ligand interactions, but had the potential to make other contacts that could improve their binding affinities. LigBuilder was also used to combine (i.e. link) multiple docked fragments with decent docking scores into a single compound for MD analysis.

4.2.1.3 Rational drug design

Two compounds—ZINC02044476 and a DANA derivative—were modified using rational modifications based on their respective structures from molecular docking results.

4.2.2 Docking Potential Ligands with Glide

4.2.2.1 Receptor grid generation

The snapshots generated from MD simulations of TcTS apo and holo structures from a previous study⁵⁹ were clustered as detailed in Chapter 3. The apo and holo grid files used in Chapter 3 are identical to the ones used for the docking calculations here.

4.2.2.2 Molecular docking

The Glide HTVS docking scheme with default settings was applied to all molecular libraries introduced in Section 4.2.1.1. The compounds with binding free energies better than -7.00 kcal/mol (41,187 compounds) were chosen for further analysis using Glide SP. Those compounds were docked with the default setting for Glide SP using the docked conformations from HTVS. Finally, the best 3,000 compounds from SP were selected for Glide XP docking with default settings.

4.2.3 *De novo* Drug Design

The small fragment library released with AutoGrow¹³⁸ was used to modify ZINC16247520 in an attempt to increase the favorable protein-ligand interactions without making the compound too large. All possible modification sites on the ligand were chosen as possible substitution locations. LigBuilder 2.0¹⁴⁰ was used to modify DANA, oseltamivir, and ZINC20472820. Similar to AutoGrow with ZINC16247520, all possible sites (non-polar hydrogen atoms) were used as potential locations for functional group substitutions. LigBuilder was preferred over AutoGrow for most automated design because AutoGrow tended to run significantly longer without noticeably better results.

Unlike AutoGrow, LigBuilder does not explicitly dock the compounds it designs. Instead, the input ligand structure (i.e. anchor) is docked into position previously, and the *de novo* algorithm simply makes modifications to the ligand in its docked conformation and re-calculates the new binding score. Because this is a very crude method for determining the binding free energies, all compounds created by LigBuilder were re-docked using Glide SP followed by Glide XP both with default settings again.

Re-docking with Glide results in a more reliable docked conformation and binding free energy compared to the LigBuilder results.

4.2.4 Molecular Dynamics

MD was performed using Amber²⁰³ on the most promising compounds from Glide docking results. The MD procedure was exactly as described in the Methods section of Chapter 3. Ten nanoseconds of production MD was performed for each compound, except in cases where additional MD was completed to gain a better understanding of the inhibitor binding mode.

4.2.5 Free Energies of Binding

The binding free energies were calculated using the Generalized Born solvation model (MM-GBSA) with *MMPBSA.py*¹⁸⁹ as previously described in the Methods section of Chapter 3.

4.3 Results

The twelve compounds with the best free energies of binding from MM-GBSA calculations are listed in Table 4-2. The twelve compounds listed in Table 4-2 are the inhibitors analyzed that resulted in free energies of binding that were stronger than those calculated for DANA, which is known to be a poor inhibitor of TcTS. For a relative comparison, free energies of binding were calculated using MD trajectories of the substrate (sialyl-lactose) and DANA previously performed in our group⁵⁹ that corresponded to -54.8 ± 0.1 kcal/mol and -53.8 ± 0.3 kcal/mol, respectively. The difference between the calculated binding free energies is consistent with the experimental relative binding affinities.^{63,64} The MM-GBSA binding affinities reported in Table 4-2 are averaged over 10-ns MD simulations, except for DANA Result 034 and PubChem 54353125 that were run for 50 ns and 110 ns, respectively.

Table 4-2. The best potential inhibitors of TcTTS from virtual screening and *de novo* drug methods. MM-GBSA binding free energies are in units of kcal/mol and are averaged over 10 ns of MD, except DANA Result 034 and PubChem 54353125 that were ran for 50 ns and 110 ns, respectively. The MM-GBSA errors are reported as standard errors of the mean.

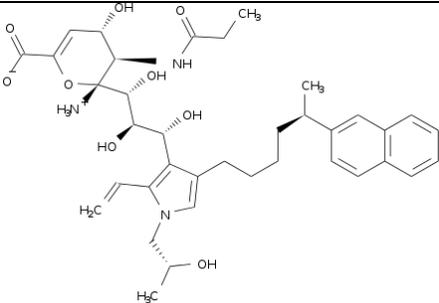
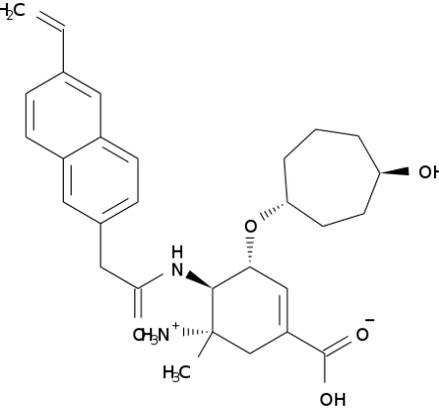
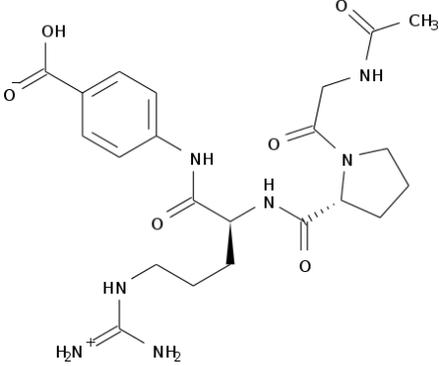
Rank	Ligand Name	Structure	Source	MM-GBSA (kcal/mol)
1	DANA Result 034		LigBuilder	-91.2 ± 0.1
2	Oseltamivir Result 70		LigBuilder	-72.5 ± 0.2
3	PubChem 54353125		PubChem	-72.1 ± 0.2

Table 4-2. Continued

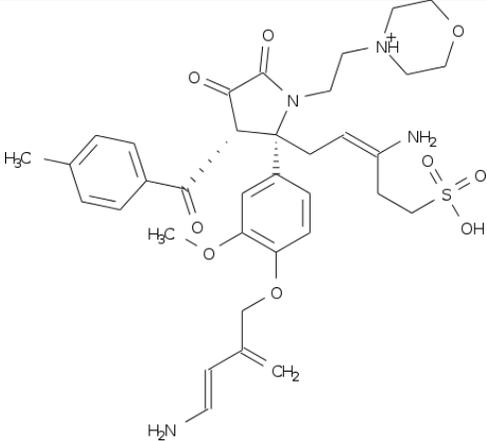
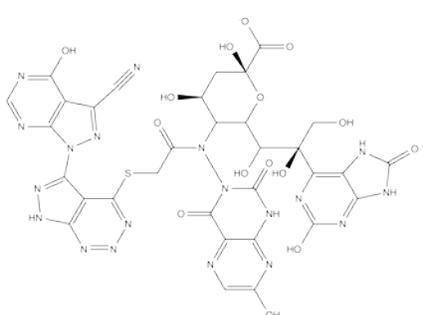
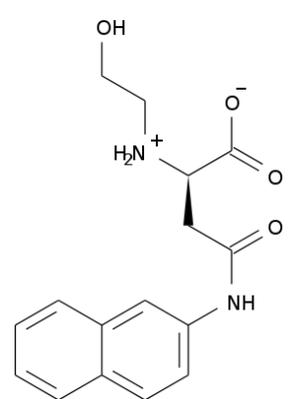
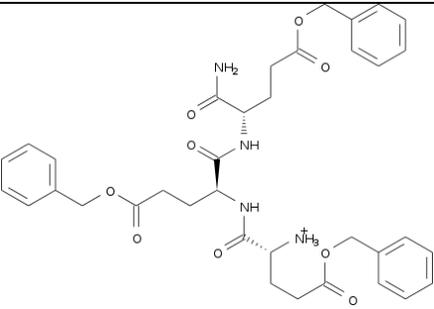
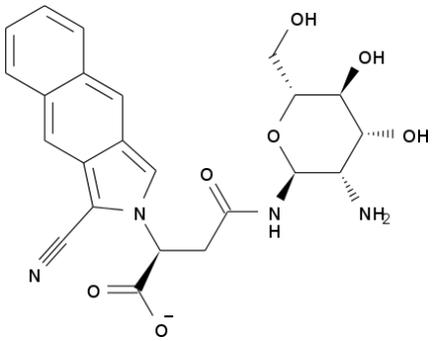
Rank	Ligand Name	Structure	Source	MM-GBSA (kcal/mol)
4	ZINC16247520 Rank 10		AutoGrow	-62.0 ± 0.2
5	Sialic Acid Rank 1		AutoGrow	-61.5 ± 0.3
6	ZINC04939766		ZINC Leads	-58.5 ± 0.2

Table 4-2. Continued

Rank	Ligand Name	Structure	Source	MM-GBSA (kcal/mol)
7	Zanamivir phosphonate- Rank 62		Literature	-57.9 ± 0.3
8	Zanamivir phosphonate Rank 1		Literature	-55.8 ± 0.3
9	ZINC34334192		ZINC Leads	-54.7 ± 0.2
10	ZINC04280204		ZINC Leads	-54.6 ± 0.2

Table 4-2. Continued

Rank	Ligand Name	Structure	Source	MM-GBSA (kcal/mol)
11	Asinex 380341		Asinex	-54.3 ± 0.4
12	PubChem 12000964		PubChem	-54.1 ± 0.2

Since MM-GBSA calculations are more rigorous and involve the analysis of many protein-ligand conformations from an MD trajectory, the relative rankings of potential inhibitors are based on these results instead of the Glide docking free energies. Furthermore, MM-GBSA calculations have previously been shown as an improvement over docking results.¹⁹⁷ This is especially important given the lack of correlation between Glide and MM-GBSA binding free energies, even using Glide XP, the most sophisticated scoring function Schrödinger offers (Figure 4-1).

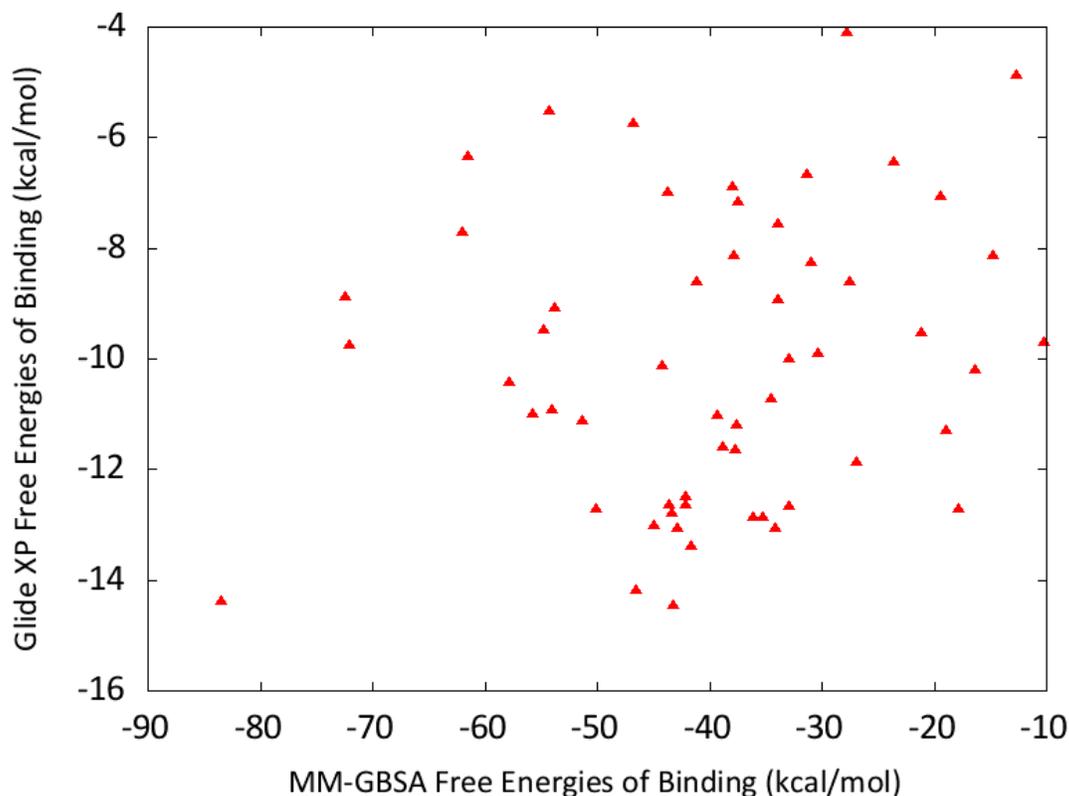


Figure 4-1. A plot of MM-GBSA binding free energies compared to Glide XP docking free energies. The results suggest there is no clear correlation between the two methods.

4.4 Discussion

The most promising potential inhibitors of TcTS were analyzed in great detail, but a detailed list of all potential inhibitors simulated with MD after docking to TcTS can be found in the Appendix (Table A-4). In this section, we will examine the binding modes of the best eight potential inhibitors discovered from virtual screening and *de novo* design. Furthermore, we will detail potential structural modifications that were made attempting to improve the binding free energies where applicable. Finally, we will examine a molecule created by combining fragments with strong free energies of binding from docking calculations.

4.4.1 Best Potential Inhibitors

4.4.1.1 DANA Result 034

DANA, known as a strong inhibitor of related neuraminidases, was found to be a poor inhibitor of TcTS.^{63,64} Nevertheless, medicinal chemists attempting to find strong inhibitors of TcTS have attempted to modify DANA to make it specific to TcTS, but it is our understanding that *de novo* drug design has never been performed on DANA. Thus, a Glide XP docked conformation of DANA to the holo TcTS grid was used as input for LigBuilder 2.0¹⁴⁰ to generate many derivatives of DANA. The top 200 derivatives from LigBuilder were subsequently docked to the holo TcTS grid using Glide SP followed by Glide XP docking methods. The 34th best derivative according to LigBuilder resulted in the best Glide XP score (-14.38 kcal/mol) of all 200 compounds, and 50 ns of production MD on DANA Result 034 bound to TcTS resulted in a MM-GBSA free energy of binding of -91.2 ± 0.1 kcal/mol (Table 4-2). The stability of the enzyme and ligand was measured from the initial conformation using RMSD, as shown in Figure 4-2. The RMSD for the backbone atoms of TcTS remains consistently below 1.5 Å during the 50-ns trajectory, while the ligand, DANA Result 034 is very mobile with the RMSD changing rapidly between 1.5 Å and 3 Å throughout the simulation. The flexibility of the ligand is localized in the flexible region containing the naphthalene functional group, shown in green in Figure 4-3. The regions of DANA Result 034 that bind to the sialic acid binding site of TcTS remain mostly rigid throughout the simulation, while the naphthalene group moves freely. AutoGrow likely added the naphthalene group during *de novo* design to provide the ligand the ability to π -stack with Tyr119 or Trp312, that are known to be important during TcTS catalysis.

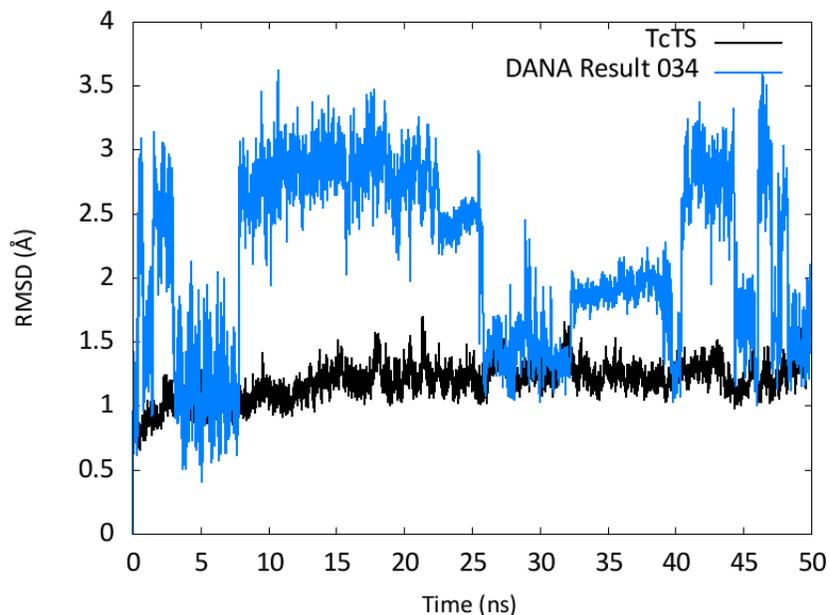


Figure 4-2. The RMSD of TcTS and DANA Result 034 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for DANA Result 034.

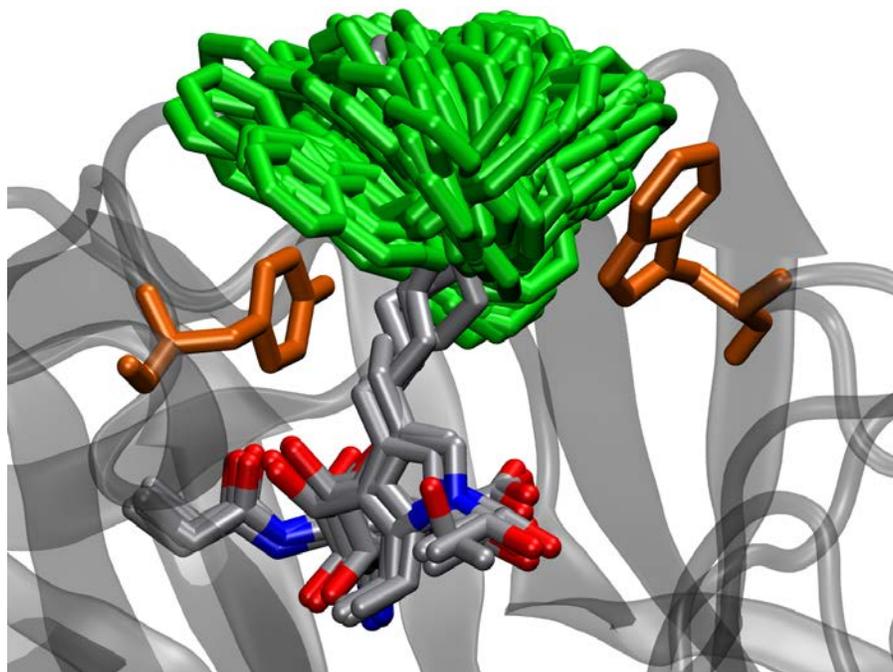


Figure 4-3. Depiction of the flexibility of the naphthalene region (green) of DANA Result 034 compared to the rest of the ligand. Enzyme residues immediately flanking the naphthalene, Tyr119 (left) and Trp312 (right) are shown in orange. For clarity, hydrogen atoms have been excluded.

However, the linker region between naphthalene and the rest of the ligand is flexible, and the length of the linker forces the naphthalene group to extend partially into the solvent, which yields an extremely flexible region of the ligand. Despite the variability of the ligand, the free energy of binding of DANA Result 034 bound to TcTS remains relatively consistent during the simulation and does not appear to be getting better or worse as the simulation progresses (Figure 4-4).

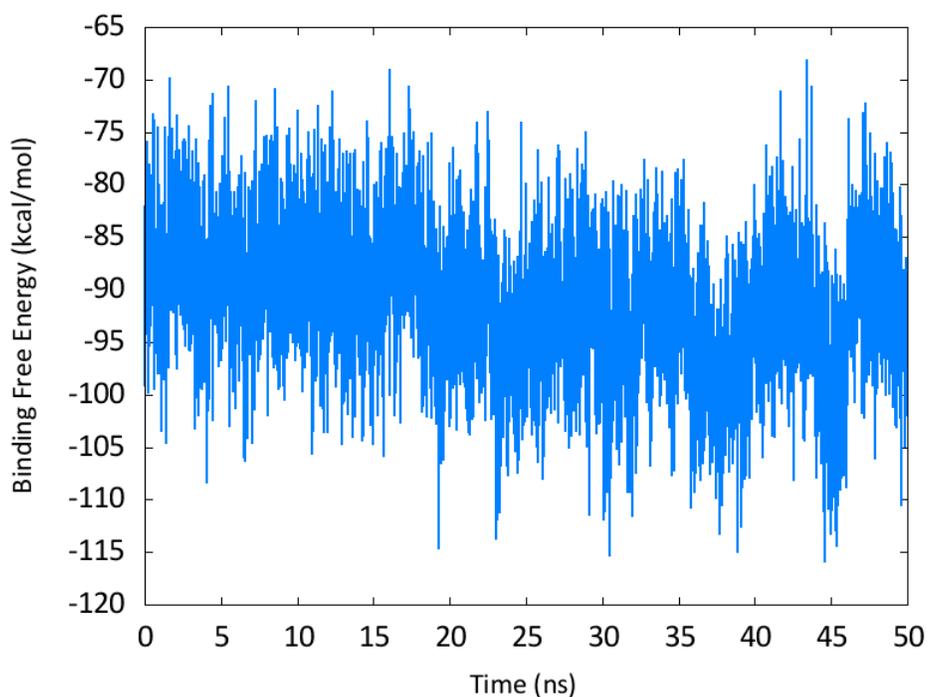


Figure 4-4. The MM-GBSA binding free energy of DANA Result 034 bound to TcTS as a function of time for a 50-ns MD simulation.

Importantly, the binding free energy remains no worse than 15 kcal/mol *better* than the MM-GBSA binding free energy of the substrate (sialyl-lactose, -54.8 kcal/mol). It should be noted once again that the calculated MM-GBSA binding free energies are not considered real binding free energies, should not be considered quantitative, and should only be used to compare the relative affinities of ligands for the same system. Regardless, the strong free energy of binding from MM-GBSA calculations is

encouraging and suggests that DANA Result 034 has the potential to bind much tighter to the enzyme than the substrate *in vivo*, although the binding mode would not make it a match for any of the e-pharmacophore models presented in Chapter 3. Unfortunately, the variability in the ligand structure at the mouth of the TcTS active site means there likely are still structural improvements that can be made to improve the binding affinity.

We used rational drug design ideas to propose modifications to DANA Result 034 that could potentially improve the overall binding free energy of the ligand, with a special emphasis on alterations of the linker arm and naphthalene ring (Figure 4-5).

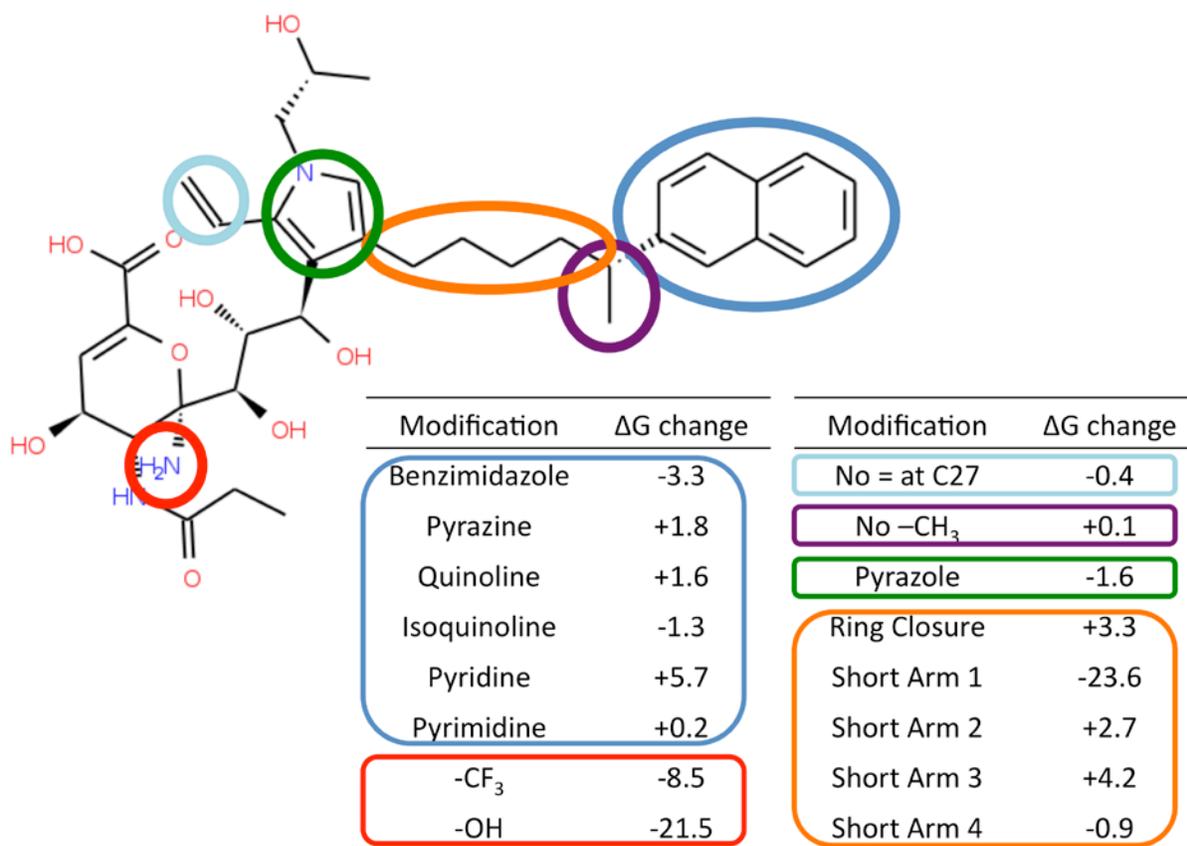


Figure 4-5. Modifications made to DANA Result 034. The structural changes are color-coded in the table alongside the corresponding change in free energy of binding compared to the original DANA Result 034 inhibitor. Free energies are in units of kcal/mol.

Because DANA was the core structure, only functional groups that were added by AutoGrow were used as potential mutation sites. Structural modifications were performed manually using Avogadro²¹¹ without changing the remaining coordinates. This allowed multiple modifications to the structure without re-docking the ligand, which could potentially result in an undesirable change in the overall binding mode. Since the binding modes of all DANA Result 034 derivatives considered were consistent, the MD and free energies of binding could be directly compared relative to one another. Two of the substitutions—Short Arm 1 and –OH—resulted in significantly better inhibitors of TcTS with improvements in the binding free energy of -23.6 kcal/mol and -21.5 kcal/mol, respectively. The Short Arm 1 modification involved the removal of a –CH₂– group from the linker region (circled in orange in Figure 4-5) that allows the naphthalene group to orient itself between Tyr119 and Trp312 forming a favorable π -stacking network. Meanwhile, the –OH group replaced the –NH₃⁺ group (circled in red in Figure 4-5) that interacts with the Asp96/Glu230 motif deep in the active site. Unfortunately, the structural substitutions were not found to be additive, though, because a DANA Result 034 derivative that combined multiple promising substitutions was actually worse than the original compound by 6.1 kcal/mol from MM-GBSA calculations.

In general, DANA Result 034 is predicted to be a strong inhibitor of TcTS using molecular docking, MD, and MM-GBSA calculations. Proposed modifications suggest that structural improvements could make the ligand a better inhibitor of TcTS. However, DANA Result 034 has several shortcomings as a drug option to treat Chagas' disease. First, DANA Result 034 is large and flexible, while an ideal drug would be small and rigid in comparison. Second, although DANA is readily available, the process of making

such a large, complex molecule with many chiral centers could create several technical difficulties during synthesis. Third, the specificity of DANA Result 034 to bind to TcTS over other similar enzymes (specifically neuraminidases) is unknown. Fourth, it has not been specifically tested if this compound with so many potential hydrogen bonding partners would have any unintended interactions during drug delivery in a patient.

4.4.1.2 Oseltamivir Result 70

Similar to DANA Result 034, Oseltamivir Result 70 was designed using LigBuilder based on a structure of Oseltamivir (more commonly known as the antiviral drug Tamiflu) docked to TcTS using Glide XP. Oseltamivir is a strong inhibitor of neuraminidase, which is similar in structure and sequence to TcTS. After LigBuilder generated 200 different Oseltamivir variations, the derivatives were docked to the holo TcTS grid using Glide. The best ligand from Glide XP was Result 70, which was found to be the second best inhibitor of TcTS after 10 ns of MD and MM-GBSA calculations with a free energy of binding of -72.5 ± 0.2 kcal/mol (Table 4-2).

TcTS does not undergo any large conformational changes during the simulation with an RMSD consistently below 1.5 Å (Figure 4-6). Oseltamivir Result 70, on the other hand, undergoes a conformational change corresponding to an RMSD change of more than 1.0 Å near 4 ns into the simulation. This structural change is subtle and involves a 180° rotation of the plane of the naphthalene functional group in the acceptor binding region of the active site (Figure 4-7). Since this rotation results in a structure similar to the original conformation, it is not surprising that the binding free energy remains fairly constant over this part of the simulation without any significant changes other than anticipated fluctuations caused by the normal dynamics of the system (Figure 4-8).

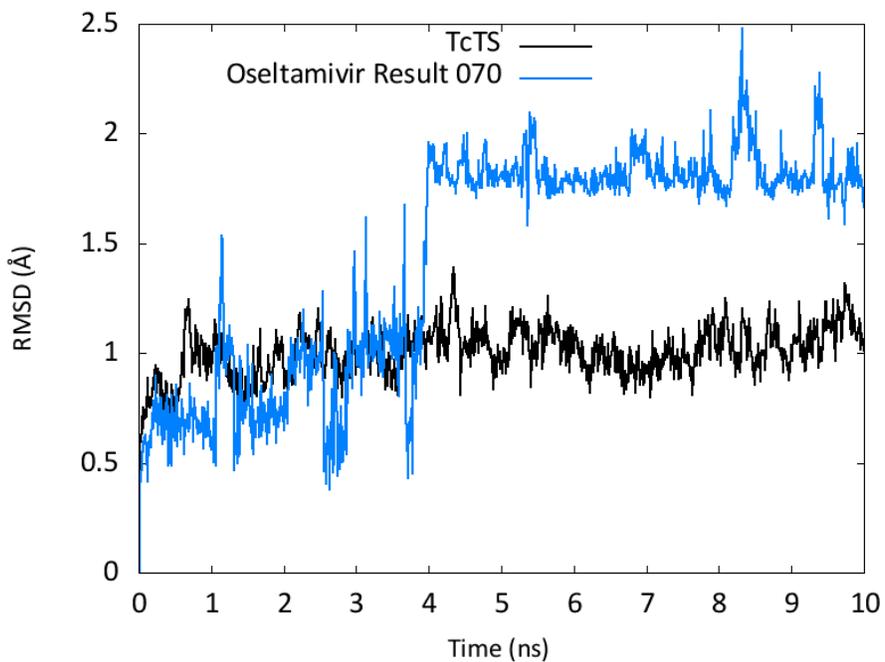


Figure 4-6. The RMSD of TcTS and Oseltamivir Result 70 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for Oseltamivir Result 70.

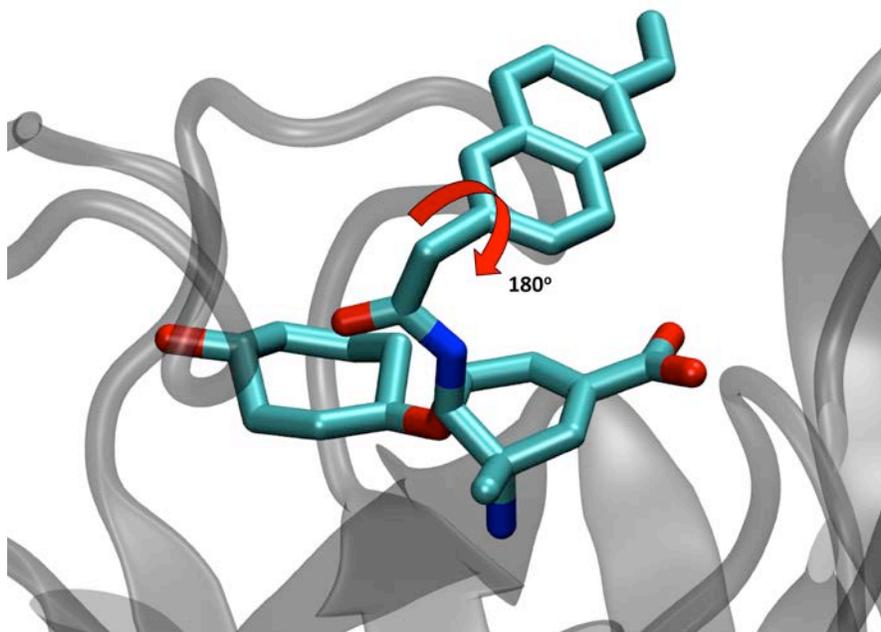


Figure 4-7. Depiction of the rotation conformational change observed for Oseltamivir Result 070 during MD around 4 ns. For clarity, all hydrogen atoms have been removed.

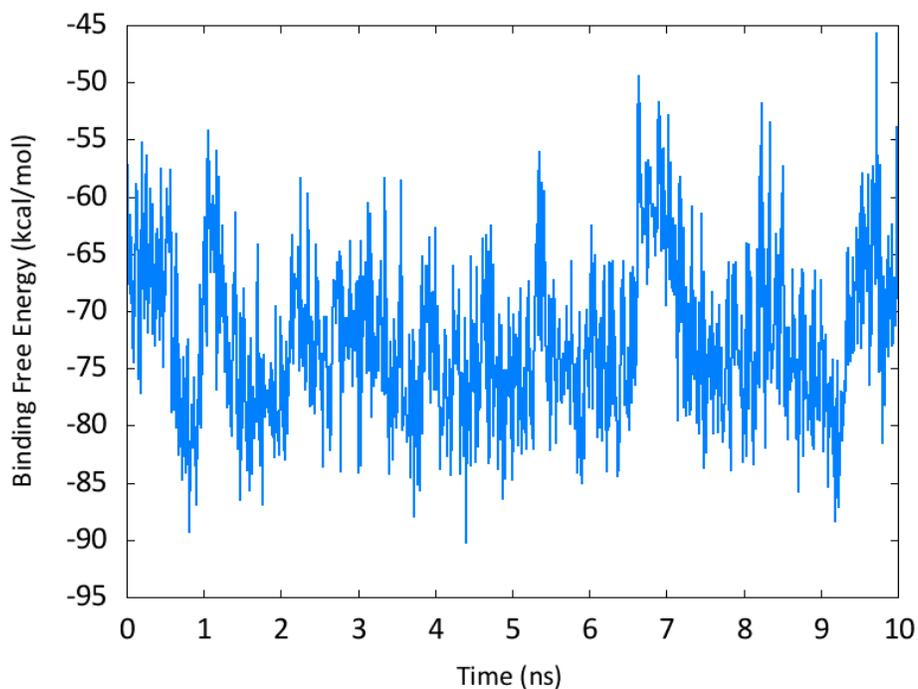


Figure 4-8. The MM-GBSA binding free energy of Oseltamivir Result 70 bound to TcTS as a function of time for a 10-ns MD simulation.

Oseltamivir is a strong inhibitor of neuraminidase and binds directly to the sialic acid binding pocket in TcTS. LigBuilder designed an oseltamivir derivative (Oseltamivir Result 70) that interacts with new regions of the TcTS. This new potential inhibitor of TcTS interacts deeper in the hydrophobic pocket near Val95/Leu176, in addition to maintaining a presence in the acceptor binding site near Tyr119 and Trp312. The formation of these new protein-ligand interactions makes this oseltamivir derivative a more specific binder to TcTS than the original compound. However, Oseltamivir Result 70 still suffers from the same disadvantages as DANA Result 034, mainly being too large, flexible, and likely difficult to synthesize. Similar to DANA Result 034, Oseltamivir Result 70 would not be a fit for any of the e-pharmacophore models described in Chapter 3.

4.4.1.3 PubChem 54353125

chEMBL 198898 (Figure 4-9) had the best free energy of binding from docking of the metabolite molecular library chEMBL with Glide HTVS to the holo grid. The protein-ligand interactions from docking suggested this was a promising lead candidate to inhibit TcTS, but the MM-GBSA binding free energy after 10 ns of MD was poor (-18.1 ± 0.2 kcal/mol) because the compound was not stable in the active site. A structure similarity search was performed on the PubChem website (<http://pubchem.ncbi.nlm.nih.gov/>) to obtain compounds that shared at least 80% structure identity to chEMBL 198898. The PubChem search resulted in 224 unique structures that were subjected to Glide SP and Glide XP docking. The molecule with the best binding free energy from docking was PubChem 54353125 (Figure 4-9), which had a Glide XP binding free energy of -9.74 kcal/mol. Unlike its chEMBL counterpart, PubChem 54353125 is stable and resulted in a free energy of binding of -72.1 ± 0.2 kcal/mol after 110 ns of unrestrained MD from MM-GBSA calculations (Table 4-2). This compound remains the best potential inhibitor of TcTS without being modified from a molecular library using either *de novo* or rational drug design methods.

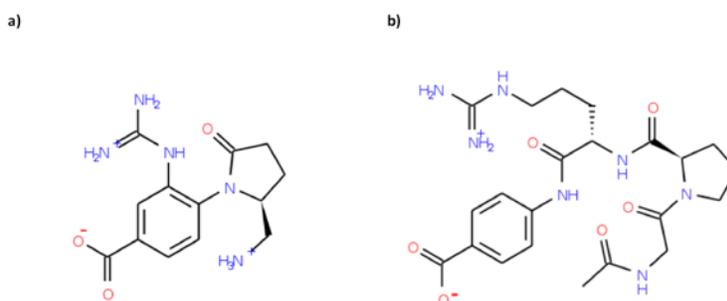


Figure 4-9. The 2D structure of chEMBL 198898 and PubChem 54353125. PubChem 54353125 (b) was obtained from performing a structure similarity search of chEMBL 198898 (a).

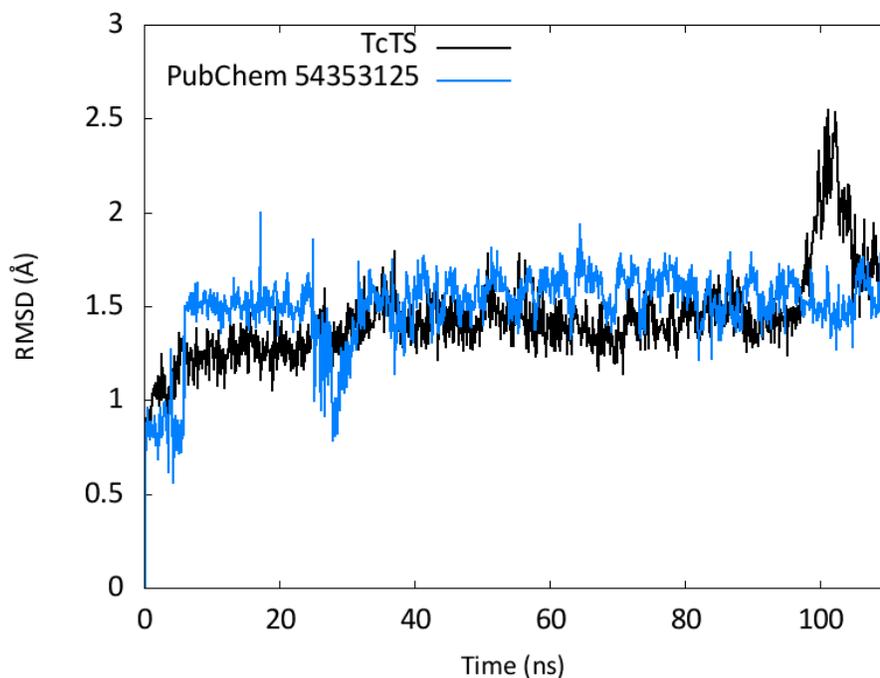


Figure 4-10. The RMSD of TcTS and PubChem 54353125 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for PubChem 54353125.

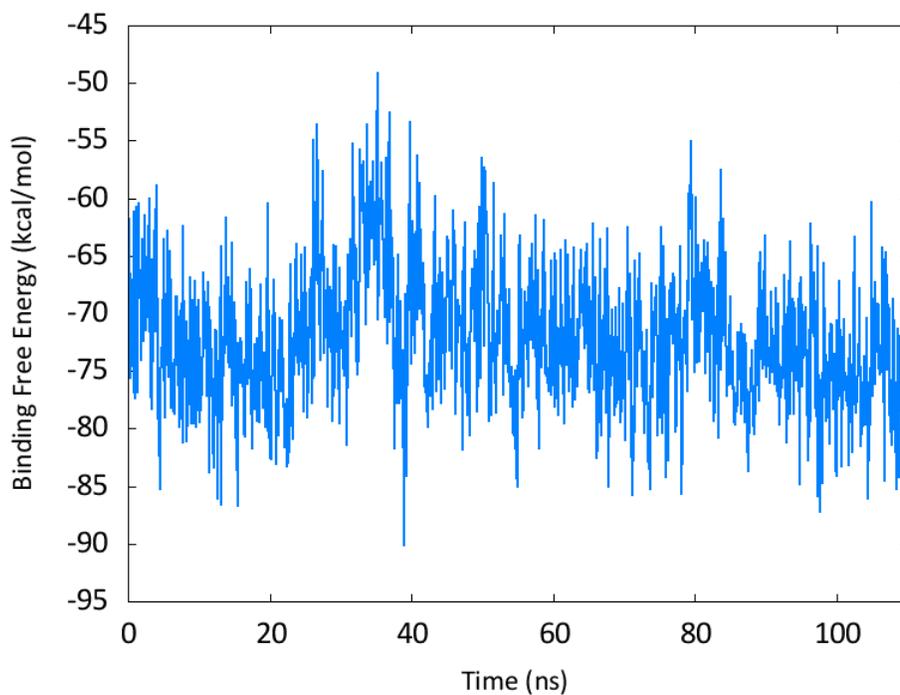


Figure 4-11. The MM-GBSA binding free energy of PubChem 54353125 bound to TcTS as a function of time for a 110-ns MD simulation.

The plots of RMSD (Figure 4-10) and free energy of binding (Figure 4-11) for PubChem 54353125 further show the stability of the ligand in the TcTS active site. The RMSD for both the enzyme and ligand remain under 2.0 Å for the entire 110-ns simulation, excluding a sharp increase in RMSD for TcTS near the end of the simulation corresponding to movement of the N-terminal tail in solution away from the active site. The binding free energy also remains relatively stable, and maintains a binding free energy better than sialyl-lactose (-54.8 kcal/mol) during the simulation except for a brief stretch near 37 ns.

The strong binding free energy of PubChem 54353125 from MD is a direct result of the binding mode and the resulting protein-ligand interactions (Figure 4-12).

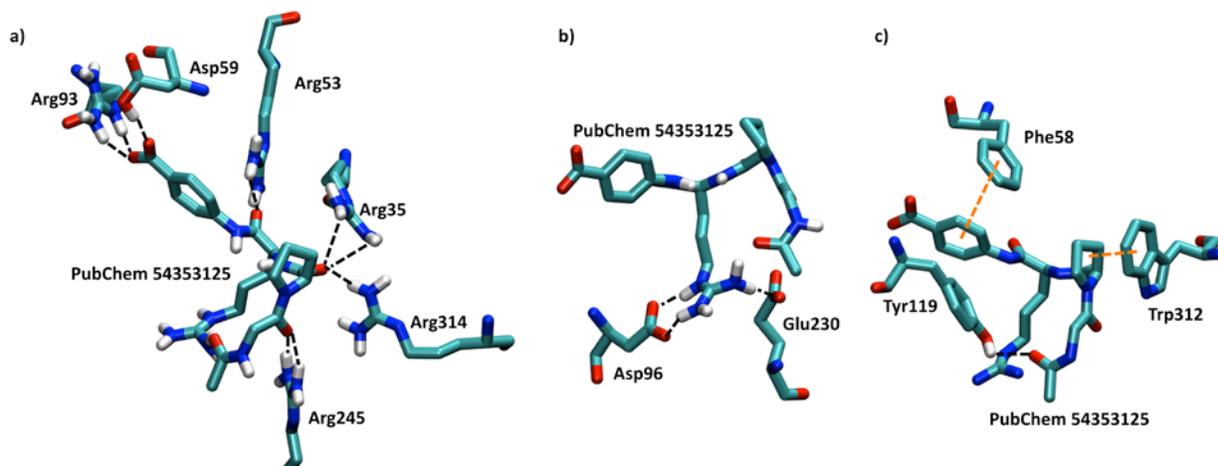


Figure 4-12. Protein-ligand interactions between TcTS and PubChem 54353125. Polar interactions in the sialic acid binding site are shown using two different orientations a) and b) to show all hydrogen bonds. Interactions in the acceptor binding site c) are also shown. Hydrogen bonds are represented using black dotted lines while π -stacking interactions are depicted using orange dotted lines. For clarity, only polar hydrogen bonds are shown.

Throughout the MD simulation, PubChem 54353125 forms hydrogen bond interactions with nine different residues—Arg93, Asp59, Arg53, Arg35, Arg314, Arg245, Asp96, Glu230, and Tyr119—and π -stacking interactions with two residues—Tyr312 and

Phe58. As anticipated for any strong inhibitor of TcTS, PubChem 54353125 forms hydrogen bonds with the arginine triad (Arg35, Arg314, and Arg245), but unexpectedly the binding mode predicted by Glide XP involves two neutral carbonyl oxygen atoms on the ligand hydrogen bonding to the triad. The predicted binding mode was a surprise because the ligand has a carboxylic acid, which is the more traditional hydrogen bonding partner with the arginine triad, that Glide XP positioned to interact with Arg93 instead. The discovery of a molecule using carbonyl oxygen atoms instead of a carboxylic acid interacting with the arginine triad could lead to the discovery of a completely new class of TcTS inhibitors. Although, it is important to note that the lack of a carboxylic acid interacting with the arginine triad makes PubChem 54353125 not a match for any of the e-pharmacophore models presented in Chapter 3. Another strong protein-ligand interaction is between the guanidinium group on the ligand and the Asp96/Glu230 motif. The positive charge of the guanidinium group is intended to mimic the carbocation of the transition state during catalysis.

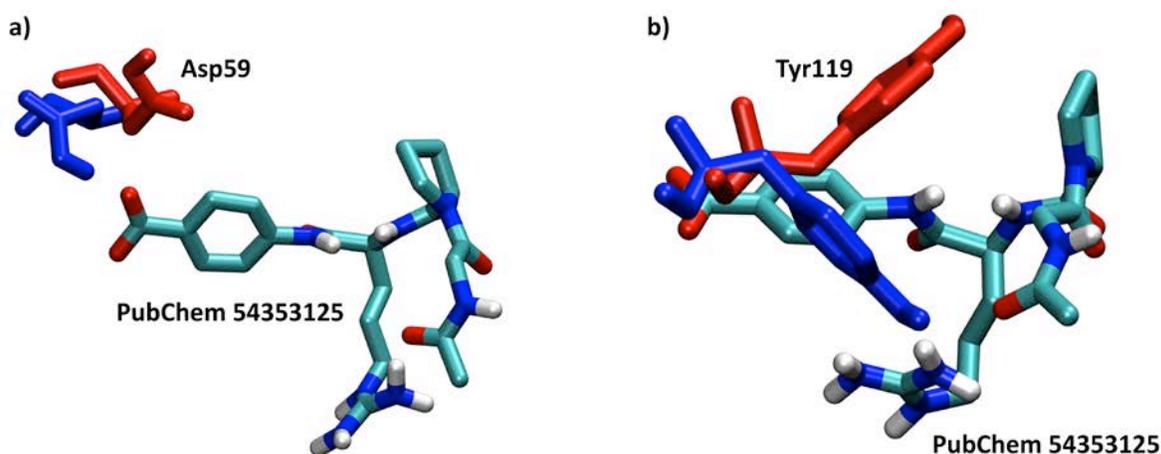


Figure 4-13. Representations of the conformational changes for two TcTS residues during MD of the TcTS-PubChem 54353125 complex. For a) Asp59 and b) Tyr119, the original Glide XP docked conformation is shown in red, while the conformation after 110 ns of MD is shown in blue.

Comparing the Glide docked conformation with the structures from MD help demonstrate the importance of performing simulations after docking. A docked structure from Glide is created using a fixed receptor structure, and MD allows the enzyme and ligand to adapt to one another, sometimes creating significantly different conformations of the protein-ligand interface. For example, the Glide XP docked conformation of TcTS-PubChem 54353125 has Asp59 and Tyr119 in their expected location with the substrate in the active site. However, during the MD simulation, it is observed that these two residues actually change conformations and rotate down to form specific hydrogen bonds with the ligand (Figure 4-13). Although the MD simulation was important for obtaining the relaxed binding mode of PubChem 54353125 in complex with TcTS, it is worth noting that despite these two residues significantly changing conformations during MD, neither had an overwhelming impact on the overall free energy of binding for the potential inhibitor.

PubChem 54353125 had the strongest free energy of binding of any compound docked in this study that was not derived from *de novo* drug design. Although the ligand is relatively large and fairly flexible, the presence of PubChem 54353125 in the PubChem database indicates that a synthesis route exists to produce the molecule. And if the compound can be created, there is an opportunity to obtain the compound and test it for inhibitory activity against TcTS *in vitro*. Based on the docking and MD results presented here, PubChem 54353125 has the potential to be a strong inhibitor of TcTS.

4.4.1.4 ZINC16247520 Rank 10

ZINC16247520 (Figure 4-14), one of the better potential inhibitors of TcTS from the ZINC ChemBridge database, had a MM-GBSA free energy of binding of -46.6 ± 0.3 kcal/mol after 146 ns of unrestrained MD. The Glide XP binding energy (-14.19

kcal/mol) for ZINC16247520 was the second best of any ligand docked during this study, only slightly behind DANA Result 034 (-14.38 kcal/mol). This ligand was unique in its binding mode because it was the first potential strong inhibitor of TcTS analyzed that did not have a carboxylic acid interacting with the arginine triad.

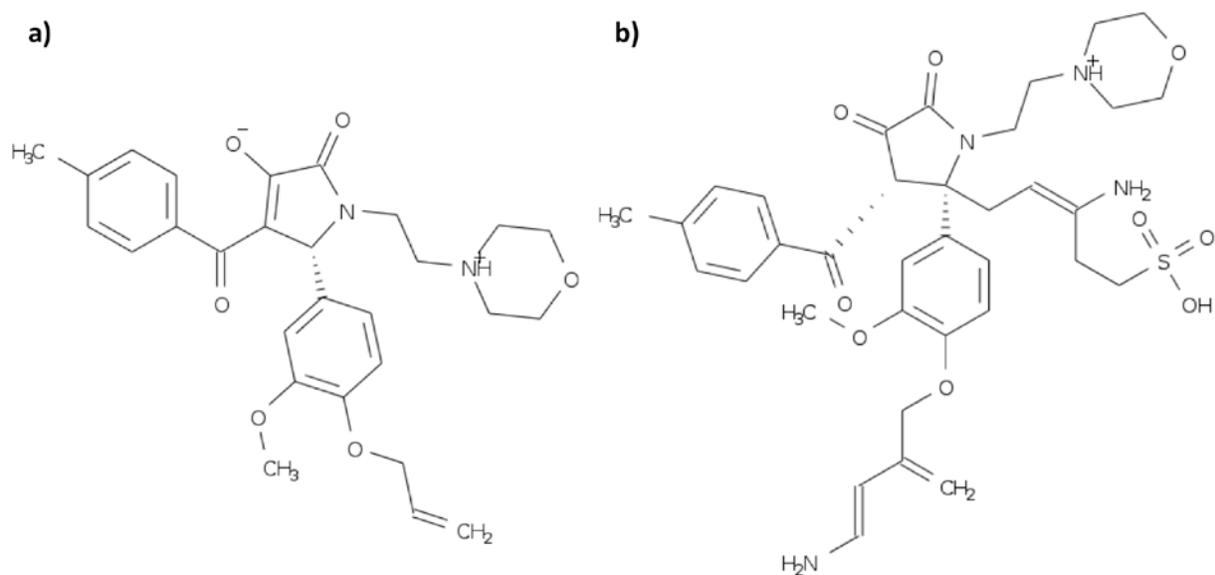


Figure 4-14. The 2D structures of ZINC16247520 and a derivative. The structure of a) ZINC16247520 was used as input for AutoGrow, which created a new compound, b) ZINC16247520 Rank 10, using *de novo* design.

Instead, ZINC16247520 has oxygen atoms on adjacent carbon atoms—one oxygen atom that is double-bonded to a carbon while the other oxygen is negatively charged and single-bonded to its carbon—that interact with the arginine triad (Figure 4-15). Placing the oxygen atoms on adjacent carbon atoms instead of a single carbon atom (such as with carboxylic acid) allows the oxygen atoms to be separated more spatially and able to interact more strongly with all three residues in the arginine triad instead of two, which was observed during MD of several potential inhibitors with carboxylic acids at this location. Furthermore, ZINC16247520 forms several other key interactions within the active site of TcTS, including a toluene group in the acceptor binding site between

the aromatic rings of Trp312 and Tyr119. Another strong TcTS-ZINC16247520 interaction involves a hydrogen bond between the positively charged nitrogen on the ligand and the side chain of Asp96. Unfortunately, despite the promising protein-ligand interactions the MM-GBSA binding free energy of ZINC16247520 was still 7.2 kcal/mol worse than DANA, a known poor inhibitor of TcTS. Considering ZINC16247520 as a strong lead scaffold inhibitor, we used its structure as input for AutoGrow to design new compounds by making key substitutions at positions on the ligand using the small molecule library provided with the AutoGrow source code. The top four unique compounds with the proper binding pose (corresponding to AutoGrow results Rank 1, 2, 5, and 10) were re-docked using Glide SP and Glide XP to confirm the binding mode for each. After docking, each compound was subjected to the MD procedure as described in the Methods section, followed by MM-GBSA free energy of binding calculations.

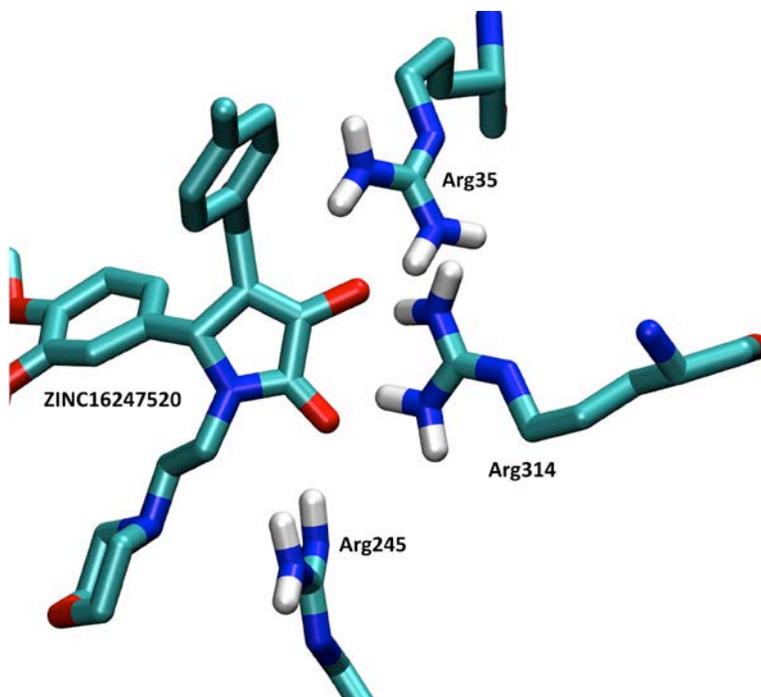


Figure 4-15. Depiction of ZINC16247520 interacting with the arginine triad of TcTS.

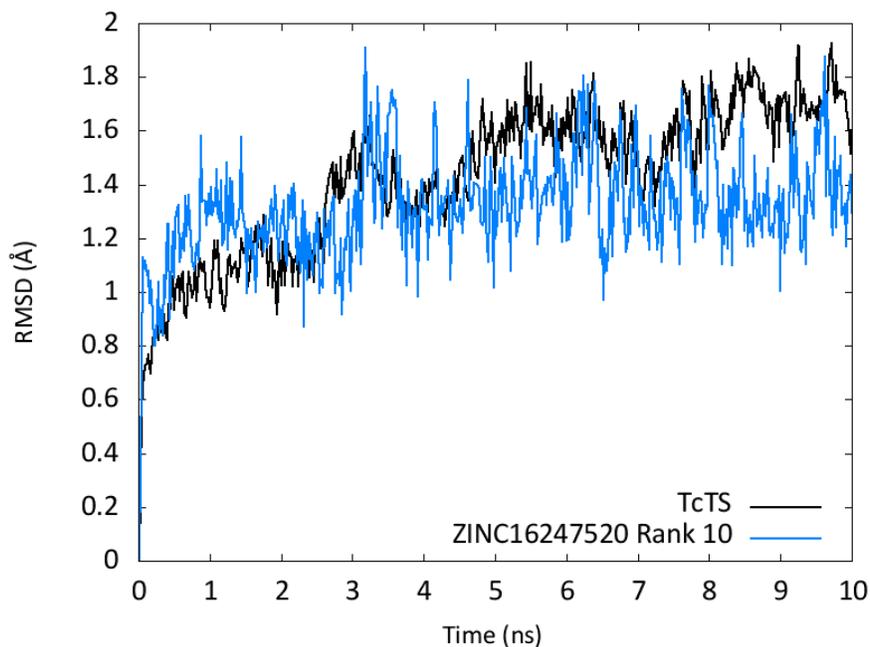


Figure 4-16. The RMSD of TcTS and ZINC16247520 Rank 10 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for ZINC16247520 Rank 10.

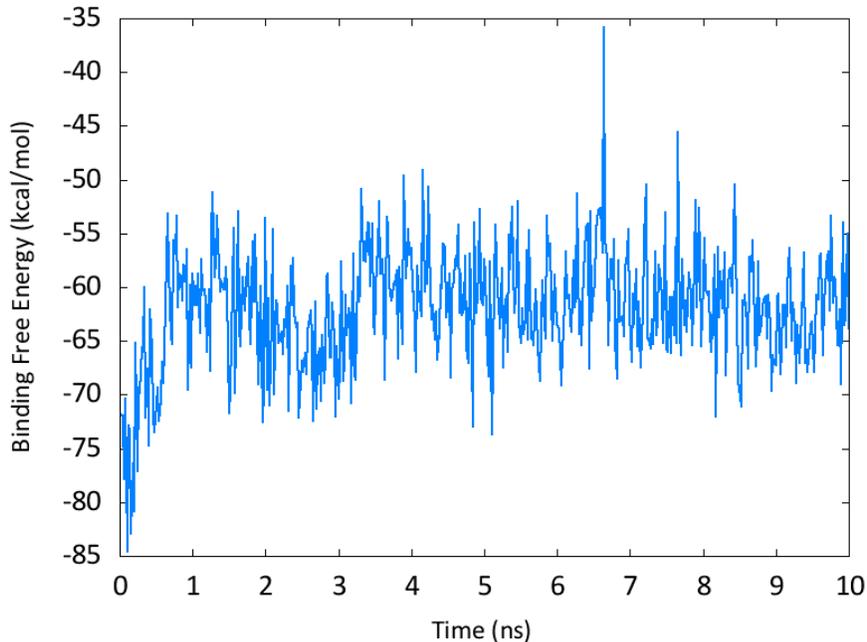


Figure 4-17. The MM-GBSA binding free energy of ZINC16247520 Rank 10 bound to TcTS as a function of time for a 10-ns MD simulation.

The best AutoGrow-derived modification of ZINC16247520 was Rank 10 (Figure 4-14), which had Glide XP and MM-GBSA binding free energies of -7.71 kcal/mol and -62.0 ± 0.2 kcal/mol, respectively. The RMSD for both the enzyme and ligand remain under 2.0 Å over the duration of the 10 ns simulation (Figure 4-16), suggesting a certain stability of the ligand in the active site. Furthermore, the MM-GBSA free energy of binding sharply increases in the first nanosecond, followed by an oscillation around the average for the final 9 ns of unrestrained MD (Figure 4-17).

The structural modifications by AutoGrow actually change the binding mode of the ligand slightly. In particular, the new sulfate group pushes the ligand oxygen atoms that were interacting with the arginine triad (Figure 4-15) away from Arg314 and Arg245, leaving the sulfate to interact with Arg314 and Arg245 while one of the two oxygen atoms remains interacting with Arg35. The other significant difference between ZINC16247520 and the Rank 10 AutoGrow derivative is the addition of a terminal $-NH_2$ that forms multiple hydrogen bonds deep within the TcTS active site with the side chains of Arg93, Asp59, and the backbone of Ser118 (Figure 4-18). The new binding mode and modifications made by LigBuilder actually make ZINC16247520 Rank 10 a match for both the holo and apo e-pharmacophore models described in Chapter 3, although the aromatic ring does not overlap perfectly with the aromatic pharmacophore feature of both models. ZINC16247520 Rank 10 also matches three of the four features of the 1S0I e-pharmacophore, while the only feature with no corresponding match on this inhibitor is R20 because there is no aromatic ring in the acceptor binding site between Tyr119 and Trp312.

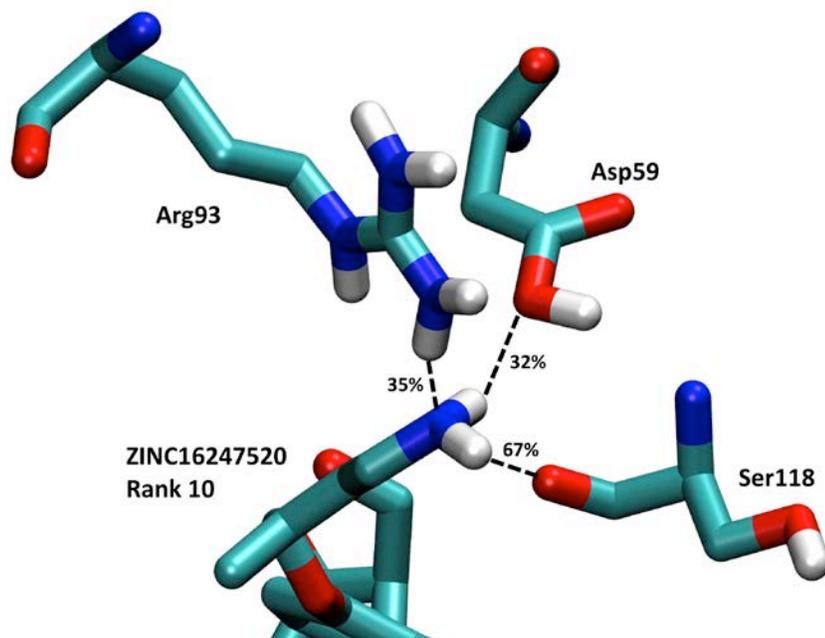


Figure 4-18. The hydrogen bonding network between ZINC16247520 Rank10 and TcTS residues Arg93, Asp59, and Ser118.

This hydrogen bonding network is not present in ZINC16247520 because the terminal amino group was an addition by AutoGrow to the original compound. Thus, the favorable protein-ligand interactions formed help explain the improved binding free energy of the Rank 10 compound compared to ZINC16247520.

Based on our calculations, ZINC16247520 Rank 10 appears to be a stronger TcTS inhibitor than ZINC16247520 with a 15.4 kcal/mol improvement in the MM-GBSA free energy of binding. The addition of sulfate and a terminal amino group to the ligand appear to create new favorable protein-ligand interactions not found in the original compound. However, as with all *de novo* designed inhibitors, it is unclear if this derivative can be easily synthesized from the scaffold compound. Furthermore, questions exist regarding the efficacy of such a large, flexible inhibitor when tested *in vivo*, but if these issues can be addressed ZINC16247520 has the potential to be a strong inhibitor of TcTS.

4.4.1.5 Sialic Acid Rank 1

AutoGrow was used to modify and optimize the structure of sialic acid within the TcTS active site using the large fragment library released with the AutoGrow source code. The best sialic acid derivative according to AutoGrow, which utilizes the AutoDock Vina²¹² scoring function, was simulated bound to TcTS for 10 ns followed by MM-GBSA calculations. This derivative, Sialic Acid Rank 1, had a MM-GBSA free energy of binding of -61.5 ± 0.3 kcal/mol, which corresponds to the fifth best inhibitor discovered in this study (Table 4-2). From the 2D structure shown in Table 4-2, it is clear that AutoGrow significantly increases the size of the ligand, starting from sialic acid with 37 atoms and ending with Sialic Acid Rank 1 containing 91 atoms. Despite the increase in potential ligand flexibility, the RMSD of Sialic Acid Rank 1 remains low (below 1.0 Å) throughout the simulation (Figure 4-19). In fact, as Figure 4-20 shows, Sialic Acid Rank 1 remains fairly rigid in its bound state within the TcTS active site.

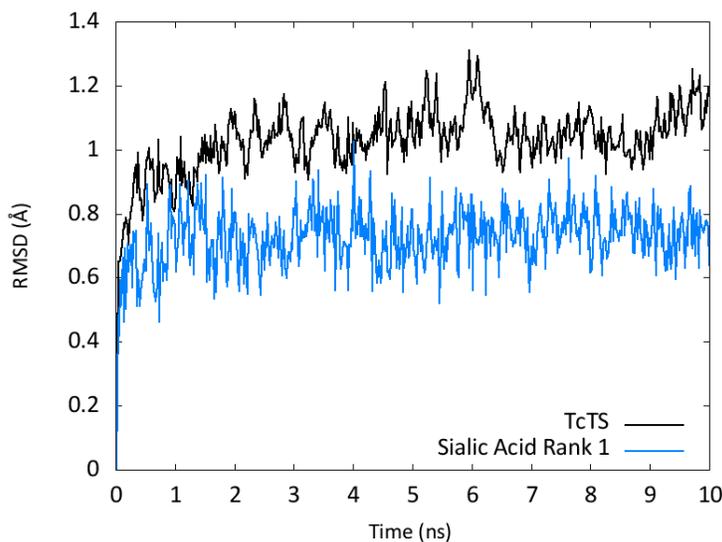


Figure 4-19. The RMSD of TcTS and Sialic Acid Rank 1 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for Sialic Acid Rank 1.

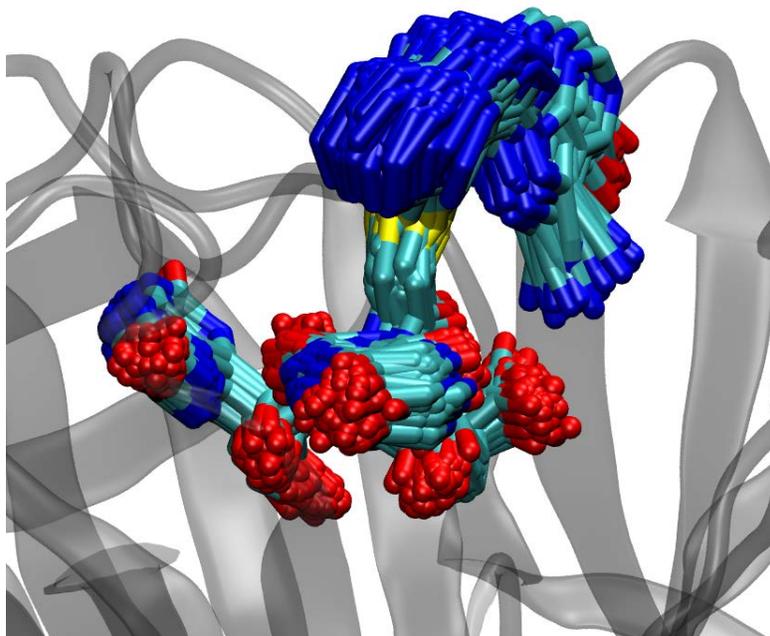


Figure 4-20. Overlay of 250 evenly spaced conformations of Sialic Acid Rank 10 bound to TcTS during a 10-ns MD simulations.

Sialic Acid Rank 1 is involved in many protein-ligand interactions with TcTS, although no hydrogen bond between the ligand and Arg53 prevents this molecule from matching any of the e-pharmacophore models in Chapter 3. Since the core of the compound is sialic acid, the molecule maintains those same interactions, including hydrogen bonds with the arginine triad and Asp96. However, the modifications by AutoGrow created additional interactions not present in the original compound. Specifically, new hydrogen bonds were formed with Arg93 (33% of the simulation), Trp120 (backbone 80%, side chain 29%), and Glu362 (100%). Additionally, several new π -stacking interactions arose because of the AutoGrow substitutions, including ring stacking between the ligand and several TcTS residues—Tyr113, Tyr119, Trp120, and Trp312. Although Tyr119 and Trp312 are residues previously shown to be important for substrate binding, the stacking interactions with Tyr113 and Trp120 represent

previously unidentified interactions with the potential to be important for the future design of novel strong TcTS inhibitors.

The binding mode of Sialic Acid Rank 1 is well defined, the ligand is stable within the TcTS active site, and the binding free energy is better than sialyl-lactose, but there are still several problems with this compound as a drug molecule. First, the size and flexibility of Sialic Acid Rank 1 would make its use as a drug molecule difficult because drugs are preferably smaller and rigid. Second, synthesizing this compound is likely very difficult if not impossible. Third, the core scaffold of this compound is sialic acid, which would likely be cleaved by the active enzyme, preventing the desired inhibition. Finally, the free energy of binding calculations performed in this study neglect any entropic contributions to binding based partly on the assumption that the unbound conformation of the ligand is similar to the bound conformation without many other available microstates. Unfortunately, this compound would be substantially more flexible in solution compared to the bound state. Thus, it is hypothesized that if entropic contributions to binding were included in these calculations, the entropic cost for this ligand would likely be significantly more compared to the other ligands, which would lower its ranking among the potential inhibitors analyzed in this study.

Despite its deficiencies as a legitimate drug candidate, Sialic Acid Rank 1 displays many potential protein-ligand interactions that could be useful for the rational drug design of other compounds as inhibitors of TcTS.

4.4.1.6 ZINC04939766

ZINC04939766 was the best unmodified TcTS inhibitor from a ZINC database (ZINC Leads) and sixth best overall (Table 4-2) with a MM-GBSA free energy of binding of -58.5 ± 0.2 kcal/mol from 10 ns of unrestrained MD. This is the smallest potential

inhibitor analyzed so far with only 40 atoms, but it was still able to have a binding free energy better than the substrate of TcTS (sialyl-lactose, -54.8 kcal/mol). ZINC04939766 was chosen for analysis with MD and MM-GBSA calculations because it scored well during an initial screening of the ZINC Leads database with Glide HTVS. The docked conformation of ZINC04939766 would match the required features of the holo e-pharmacophore model described in Chapter 3, but would not match the 1S0I or apo models. During the MD the ligand did not maintain its original binding mode derived from Glide, but instead slightly changed conformations around 2.5 ns into the simulation, as shown in the RMSD plot (Figure 4-21). This change in binding mode also simultaneously increased the MM-GBSA free energy of binding by 15 kcal/mol (Figure 4-22).

Visualization of the MD trajectory shows that the increase in binding free energy corresponds with a loss of protein-ligand interactions between ZINC04939766 and the arginine triad of TcTS, especially with Arg35 and Arg314 (Figure 4-23). The hydrogen bond distances between the ZINC04939766 carboxylic acid and the closest hydrogen atoms on the arginine residues increase from by over 2 Å for Arg35 and Arg314. However, the hydrogen bond distance did not change significantly for Arg245, as ZINC04939766 remains hydrogen bonded to it for 48% of the simulation, compared to just 18% and 12% for Arg314 and Arg35, respectively. The loss of this important hydrogen bonding network between the ligand and receptor is the main culprit in the loss of 15 kcal/mol of binding free energy after the conformational change occurs around 2.5 ns.

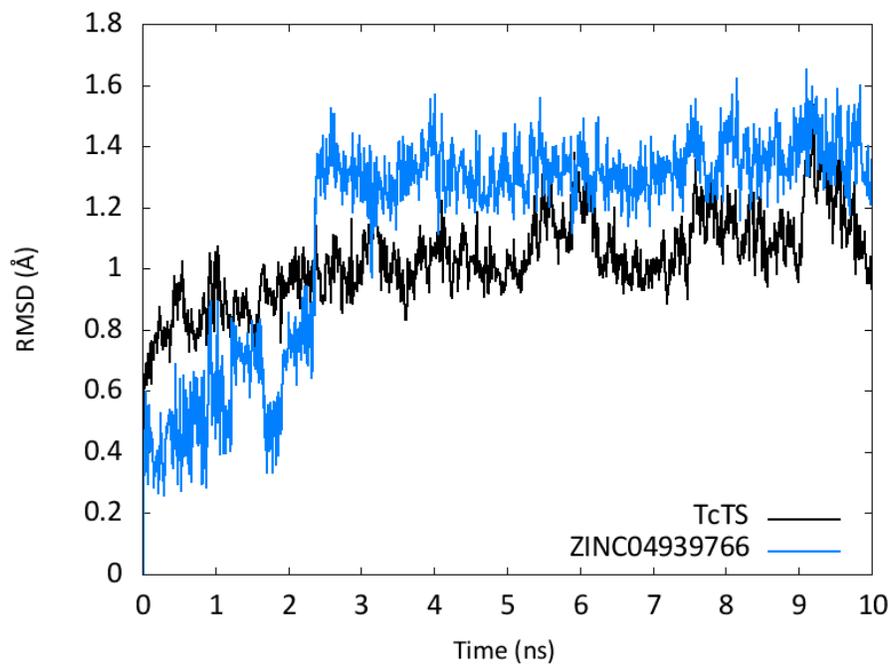


Figure 4-21. The RMSD of TcTS and ZINC04939766 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for ZINC04939766.

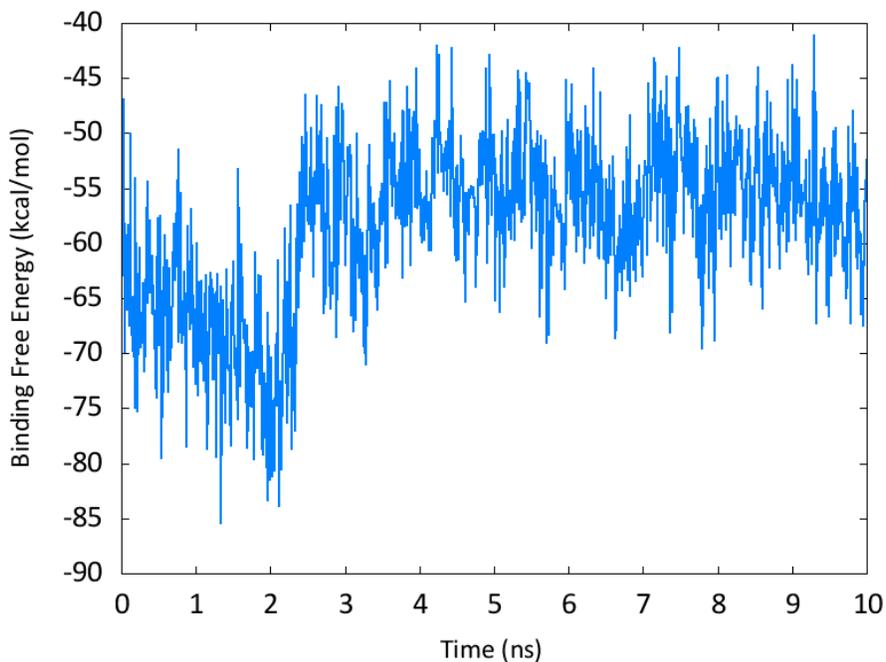


Figure 4-22. The MM-GBSA binding free energy of ZINC04939766 bound to TcTS as a function of time for a 10-ns MD simulation.

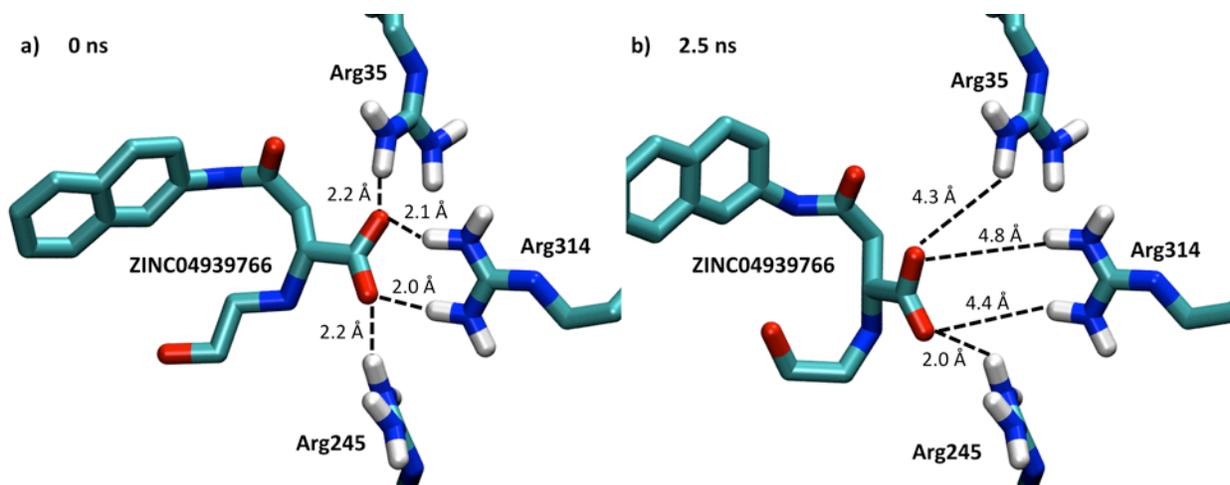


Figure 4-23. The binding mode of ZINC04939766 with the arginine triad of TcTS. The conformations shown represent the distances between the carboxylic acid of ZINC04939766 and the closest hydrogen atoms for the arginine residues at a) 0 ns and b) 2.5 ns, after the ligand undergoes a conformational change in the active site.

The change in binding mode at 2.5 ns affected ligand interactions with the arginine triad, but the other protein-ligand interactions remained relatively stable for the entirety of the simulation. In particular, there are π -stacking interactions that occur between the naphthalene group of ZINC04939766 and both Tyr113 and Tyr119, which was also observed with Sialic Acid Rank 1. However, the prominent polar interactions occur between the ligand and the Asp96/Glu230 motif (Figure 4-24). In fact, the four most persistent TcTS-ZINC04939766 hydrogen bonds occurred in this area of the bound complex at the bottom of the TcTS active site. The formal positive charge on the ligand nitrogen atom is perfectly positioned to make sustained, favorable interactions with the negative charges of Asp96 and Glu230. The strength of these hydrogen bonds suggests they play a key role in making ZINC04939766 a strong potential inhibitor of TcTS.

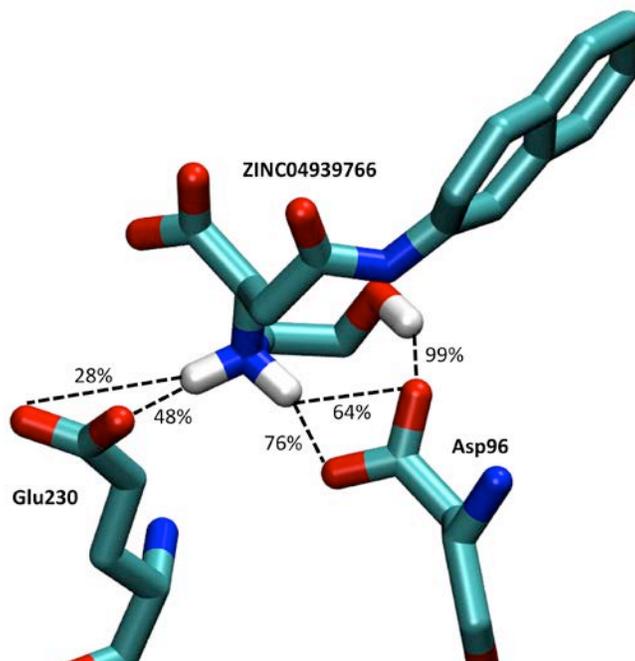


Figure 4-24. The hydrogen bonding network between ZINC04939766 and the Asp96/Glu230 motif on TcTS.

Unlike many of the potential inhibitors previously proposed, ZINC04939766 is commercially available and could be easily obtained for *in vitro* kinetic assays. Computationally, the original docked pose of ZINC04939766 had a MM-GBSA free energy of binding near between -65 and -75 kcal/mol, but a shift in the binding mode of the ligand around 2.5 ns into the simulations resulted in a sharp increase to near -55 kcal/mol. ZINC04939766 consistently forms favorable interactions with residues involved in π -stacking—Tyr113 and Tyr119—and hydrogen bonding—Asp96 and Glu230. However, the ligand is incapable of maintaining strong hydrogen bonds with the arginine triad, a fundamental region for enzyme inhibition because of its importance in substrate binding. Possible structural modifications should be addressed to improve the interactions of ZINC04939766 with the arginine triad, such as lengthening the connection between the carboxylic acid and the rest of the ligand. Ideally, any structural

changes would allow the carboxylic acid to hydrogen bond with the arginine triad without breaking the strong hydrogen bonding network between ZINC04939766 and the Asp96/Glu230 motif.

4.4.1.7 Zanamivir phosphonate

Zanamivir phosphonate, proposed as a potent inhibitor of the neuraminidases for both avian and human influenza viruses by Shie *et al.*,²¹³ was considered here for its potential to inhibit TcTS given the high structure and sequence similarity between neuraminidases and *trans*-sialidase. Experimental details of the protonation states of the phosphonate and guanidinium groups were not discussed in the literature, thus multiple protonation states of Zanamivir phosphonate were created using Avogadro and those compounds were docked to the holo grid of TcTS using Glide SP and Glide XP. The various states all had similar Glide XP binding free energies (-11.0 ± 0.2 kcal/mol) so MD and MM-GBSA calculations were ran on a protonated and deprotonated state. Furthermore, the deprotonated state resulted in three different relevant binding modes from Glide XP; so all three of these modes were also simulated. Zanamivir phosphonate deprotonated Rank 62 (ZP62) and Zanamivir phosphonate protonated Rank 1 (ZP1), the best two of those four structures, had MM-GBSA free energies of binding better than sialyl-lactose. Specifically, the MM-GBSA binding free energy of ZP62 was -57.9 ± 0.3 kcal/mol, while the free energy of binding for ZP1 was -55.8 ± 0.3 kcal/mol (Table 4-2). The binding modes for these Zanamivir phosphonate compounds would not have been discovered by screening the e-pharmacophore models in Chapter 3 because the aromatic ring is not positioned properly to match the aromatic feature of the pharmacophore hypotheses.

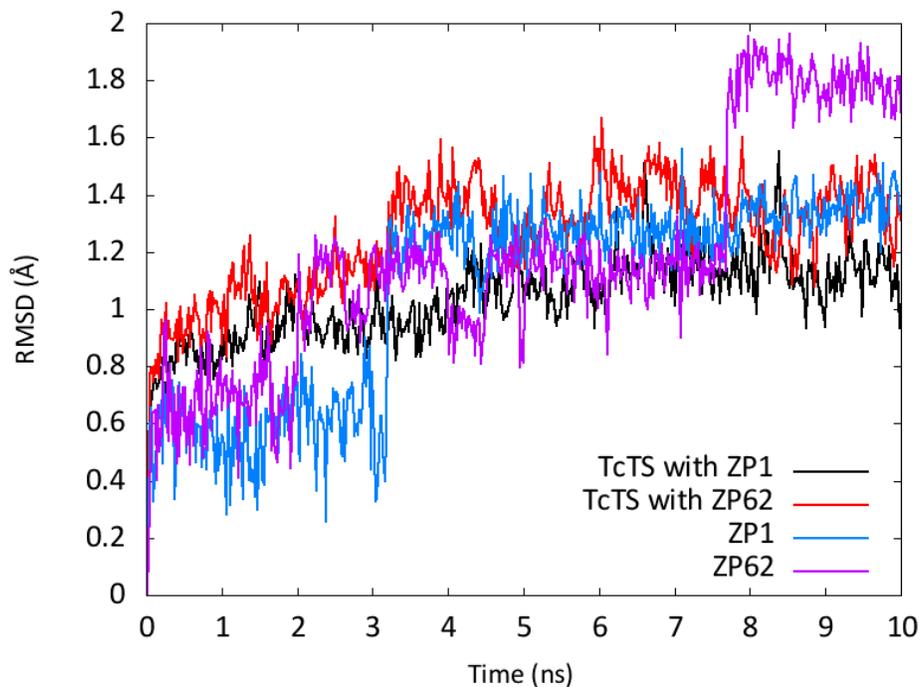


Figure 4-25. The RMSD plots from TcTS-ZP1 and TcTS-ZP62 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for the TcTS structures, and all atoms for the two ligands.

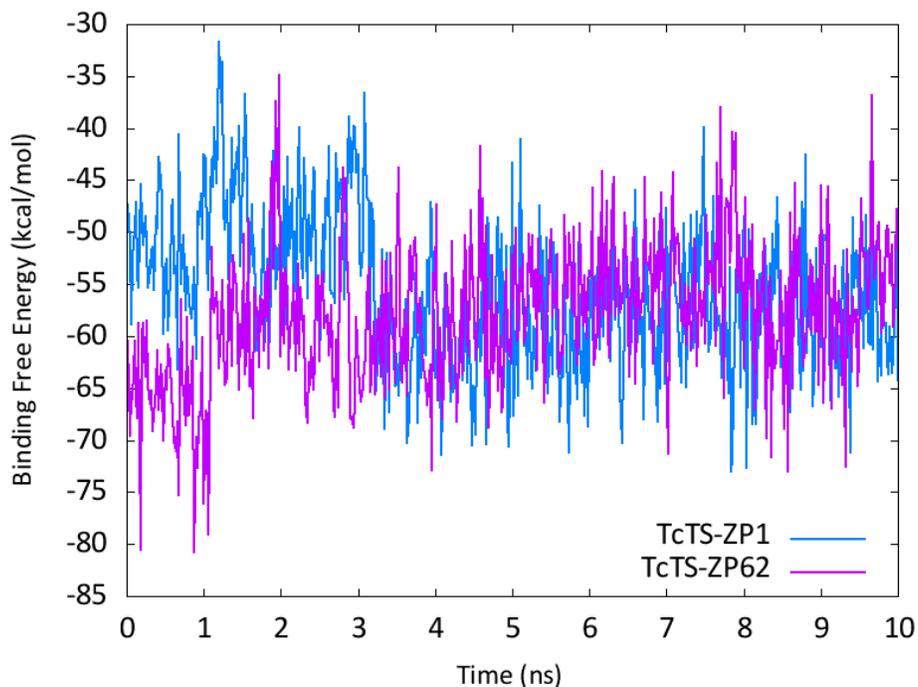


Figure 4-26. The MM-GBSA binding free energy of ZP1 (blue) and ZP62 (purple) bound to TcTS as a function of time for a 10-ns MD simulation.

The RMSD plots showing the stability of the TcTS-ZP1 and TcTS-ZP62 structures during MD are shown in Figure 4-25. Overall, neither complex deviates too far from the initial docked structure since all RMSD values remain below 2.0 Å. The enzyme structures appear to relax successfully into a local minimum and remain there for at least the last 5 ns. The ligands, however, both undergo a clear conformational change (or two) during the simulation. For ZP1, the RMSD increase from 0.6 to 1.3 Å at 3.2 ns corresponds to a rotation of the terminal hydroxide on the glycerol chain that results in the formation of a hydrogen bond between the hydroxyl group and the side chain of Arg93. This change in RMSD at 3.2 ns also coincides with an improvement in the free energy of binding for TcTS-ZP1 resulting from this new hydrogen bond with Arg93 (Figure 4-26).

ZP62 appears to undergo two different conformational changes during the MD simulation—one at 2.0 ns and the other at 7.8 ns. The increase in RMSD for ZP62 at 2.0 ns appears to correspond to a rotation of the terminal hydroxyl on the glycerol chain and a rotation of the guanidinium group on the ligand. This rotation places the hydroxyl group in position to hydrogen bond with the side chain of Arg53, but it just replaces one of the other hydroxyl groups that was interacting with Arg53, resulting in a nearly zero net change in the binding free energy (Figure 4-26). Likewise, the rotation of the guanidinium group on ZP62 breaks three hydrogen bonds, but reforms three different hydrogen bonds similar in strength and atomic makeup, so there is still little overall change in the MM-GBSA binding free energy. The sharp increase (and subsequent drop) in binding energy of TcTS-ZP62 at 2.0 ns is likely a result of the ligand reorienting itself before the new hydrogen bonds are formed. The RMSD change in ZP62 from 1.2

to 1.8 Å at 7.8 ns corresponding to another rotation of the flexible glycerol chain also incurred no net change in the binding free energy because the number of hydrogen bonds broken and formed were again the same.

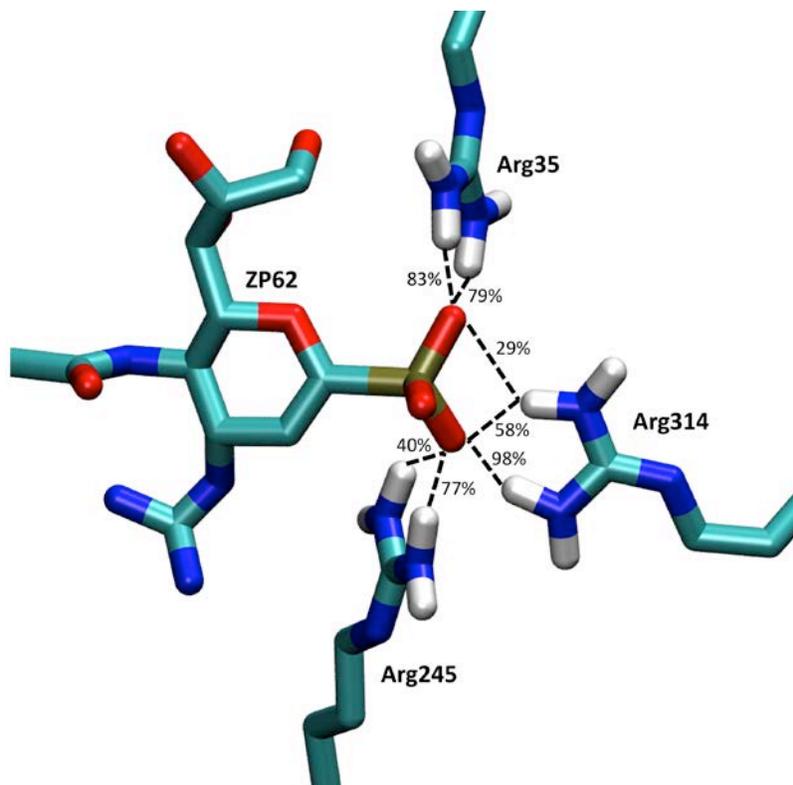


Figure 4-27. The hydrogen bonding network between the phosphonate group of ZP62 and the arginine triad of TcTS. The percentages represent the fraction of time the hydrogen bond was present during a 10 ns MD simulation.

The substitution of the carboxylic acid for a phosphonate on Zanamivir, which is unable to inhibit TcTS at 1 mM concentration,²¹⁴ is designed to improve the protein-ligand interactions with the arginine triad. The MD simulations suggest this phosphonate-arginine triad interaction is strongly favorable, with a stable hydrogen bonding network as shown for TcTS-ZP62 in Figure 4-27. The phosphonate oxygen atoms make seven significant hydrogen bonds with the arginine triad, which is more than any other potential ligand analyzed in this study. Given the arginine triad's role in

substrate binding, it is likely important for an inhibitor to make strong interactions with this motif, as the substitution of a phosphonate group on Zanamivir would suggest here.

Although simulations were performed on both the protonated and deprotonated guanidinium group of Zanamivir phosphonate, it is still unclear if a preference exists for one of the protonation states because both states obtained similar MM-GBSA free energies of binding. For the deprotonated state, Glide XP found three dominant binding modes that were all simulated. The MM-GBSA results suggest that the conformation ranked 62nd by Glide XP was actually the better inhibitor of TcTS, which shows the importance of performing MD and MM-GBSA calculations to re-rank potential inhibitors. Based on visual inspection prior to MD, conformation 62 was anticipated to give more reliable results because the binding mode was similar to other related compounds, such as DANA.

Zanamivir phosphonate is a more potent inhibitor of neuraminidase than Zanamivir, and based on the MM-GBSA results in this study Zanamivir phosphonate is expected to be an improved inhibitor of TcTS. Zanamivir does not inhibit TcTS, while Zanamivir phosphonate is calculated to have a stronger binding affinity than the natural enzyme substrate, sialyl-lactose. The improved affinity for TcTS would be a direct result of the substitution of a carboxylic acid for a phosphonate group that interacts with the arginine triad. A synthetic route to make Zanamivir phosphonate has already been discussed in the literature,²¹³ thus it could be synthesized and tested *in vitro*. However, further structural modifications should be considered to make the potential inhibitor more effective, including the addition of an aromatic group designed to interact with Tyr119 and Trp312 in the acceptor binding site.

4.4.2 *De novo* Design Candidate Combining ZINC Fragments

As mentioned in Chapter 3, we docked the Clean Fragments library of ZINC to the holo grid of TcTS using Glide XP. We selected the best non-overlapping fragments that interacted with four distinct motifs of the TcTS active site—arginine triad, Asp96/Glu230, acceptor binding site between Tyr119 and Trp312, and the hydrophobic pocket flanked by Val95 and Leu176. These four sites were all determined to be important for inhibitor binding by the e-pharmacophores described in Chapter 3. The four best fragments were combined together using LigBuilder's *link* algorithm to make a single molecule we will call 'Super fragment.' The Super fragment contained 60 total atoms and had a free energy of binding of -39.4 ± 0.3 kcal/mol after 72 ns of unrestrained MD at constant temperature (300 K) and pressure (1 atm). The RMSD of both TcTS and the ligand are shown in Figure 4-28. The Super fragment seems stable (RMSD < 1.5 Å) for nearly 40 ns, but then ligand clearly changes conformations illustrated by the elevated RMSD. Although more subtle, the increase in RMSD of the Super fragment also coincides with a noticeable raise in RMSD of TcTS by 0.5 Å.

Shortly after the noticeable RMSD change in the ligand, the free energy of binding also begins to increase and is about 40 kcal/mol worse at 72 ns than it was at 40 ns (Figure 4-29). The MM-GBSA binding energy begins the simulation around -90 kcal/mol, but progressively worsens to a final value of +8.7 kcal/mol. The positive binding energy at the end of the simulation suggests the ligand likely dissociated from the enzyme.

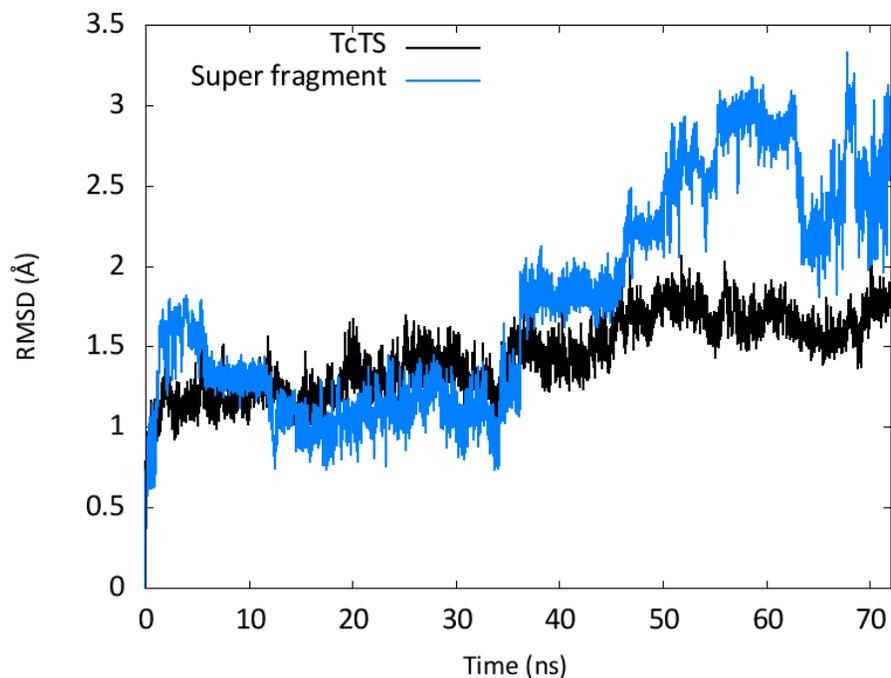


Figure 4-28. The RMSD of TcTS and the Super fragment using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for the Super fragment.

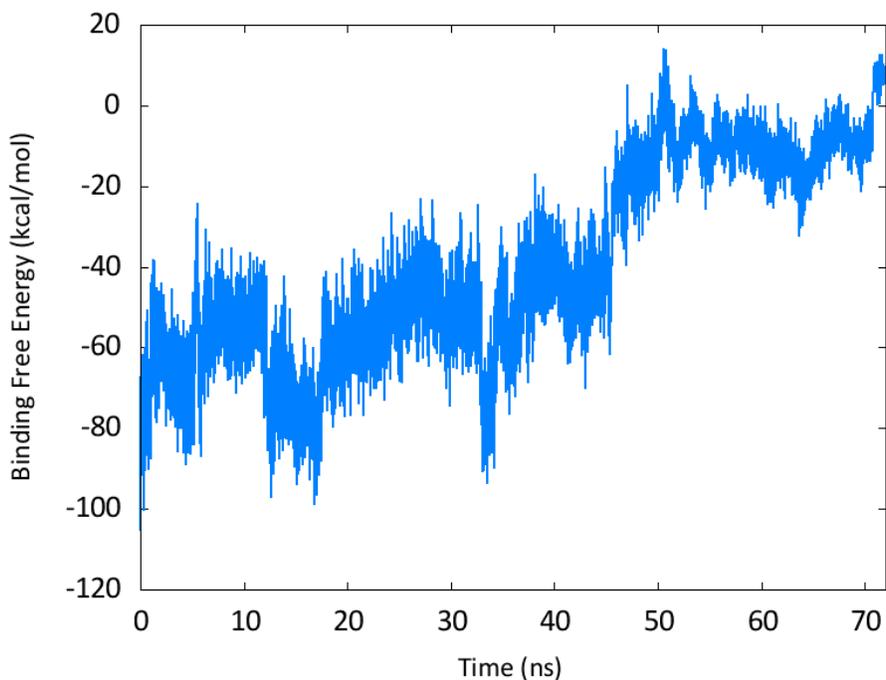


Figure 4-29. The MM-GBSA binding free energy of the Super fragment bound to TcTS as a function of time for a 72-ns MD simulation.

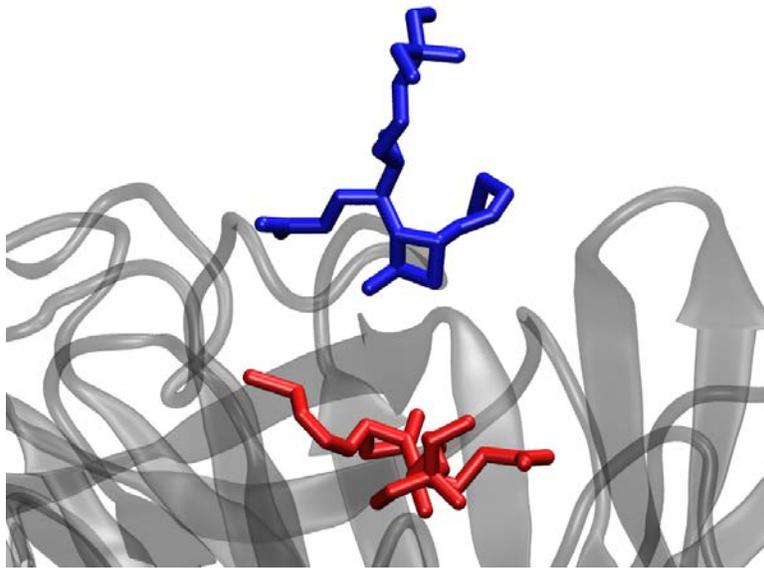


Figure 4-30. The position of the Super fragment relative to the TcTS active site at 0 ns (red) and 72 ns (blue) showing the dissociation of the ligand during MD.

The trajectory of TcTS and the Super fragment confirmed that the ligand dissociated from the enzyme. As Figure 4-30 shows, the Super fragment begins the simulation deep in the active site, but after 72 ns of unrestrained MD the ligand is positioned at the edge of the active site forming only minimal contact with TcTS. Dissociation begins around 48 ns, which corresponds to the sharp increase in binding free energy in Figure 4-29. Overall, the strong initial MM-GBSA binding free energy suggests the Super fragment has the potential to inhibit TcTS, but the lack of stability in the active site is concerning, and likely means that structural modifications of the Super fragment would be necessary to make it a viable drug candidate against Chagas' disease.

4.5 Concluding Remarks

In this study, we have used a combination of molecular docking, MD, and MM-GBSA binding free energy calculations to discover new potential inhibitor of TcTS.

Specifically, we docked over 35 million structures and ran MD and calculated binding free energies for over 100 different compounds. Four of the best five drug possibilities—DANA Result 034, Oseltamivir Result 70, ZINC16247520 Rank 10, and Sialic Acid Rank 1—were designed using *de novo* software. Although these compounds appear to have very strong inhibitory activity of TcTS, there are serious concerns about their viability as drug treatments because they are large, potentially difficult to synthesize, and flexible. The best candidate for experimental testing is PubChem 54353125, which had a free energy of binding of -72.1 ± 0.2 kcal/mol, because it could be synthesized or purchased. However, PubChem 54353125 is still large and flexible and may have difficulty working as an effective inhibitor of TcTS *in vivo*. Finally, the sixth best potential inhibitor—ZINC04939766—had a MM-GBSA binding free energy of -58.1 ± 0.2 kcal/mol (nearly 4 kcal/mol better than the substrate, sialyl-lactose) and is relatively small compared to some of the other top candidates and is also available from commercial vendors. Unfortunately, the calculated binding free energy of ZINC04939766 is significantly worse, decreasing the chance of it being a strong inhibitor of TcTS *in vitro*. Overall, several molecules have been proposed (Table 4-2) as novel inhibitors of TcTS that have the potential to develop into drugs in the treatment of Chagas' disease.

CHAPTER 5
MMPBSA.py: AN EFFICIENT PROGRAM FOR END-STATE FREE ENERGY
CALCULATIONS

5.1 Preliminaries

Free energy calculations have proven useful for a number of topics in computational biology, such as drug design and protein structure determination.^{175,215} Several methods are available to calculate free energies, such as Free Energy Perturbation,²¹⁶ Replica Exchange Free Energy Perturbation,²¹⁷ and Thermodynamic Integration.²¹⁸ These methods, though theoretically rigorous, are computationally demanding and become prohibitively expensive as system size increases.

These methods converge poorly for complex systems, so the reaction coordinate is often divided into intermediate states. End-state free energy methods, on the other hand, reduce computational cost by eliminating the need for simulating these intermediate states. Modeling the solvent implicitly further reduces the computational cost by eliminating the noise caused by explicit solvent molecules. As a result, end-state methods have been used extensively in computational studies.^{219–221} We will not discuss here the advantages and disadvantages of applying end-state methods—for a recent review, see Refs. [175] and [222].

This chapter will focus on the implementation of end-state methods in *MMPBSA.py*, a program released with the open source AmberTools package.²⁰³ *MMPBSA.py* has already been applied to several systems, including calculating the binding free energies of DNA-binding molecules²²³ and HIV protease inhibitors.²²⁴

Portions of this chapter are excerpted and adapted with permission from Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *J. Chem. Theory Comput.* 2012, 8, 3314–3321. Copyright 2012 American Chemical Society.

5.2 Theory and Methods

Before describing the capabilities of *MMPBSA.py*, we will briefly review the theory and calculations of the end-state methods that *MMPBSA.py* is capable of performing.

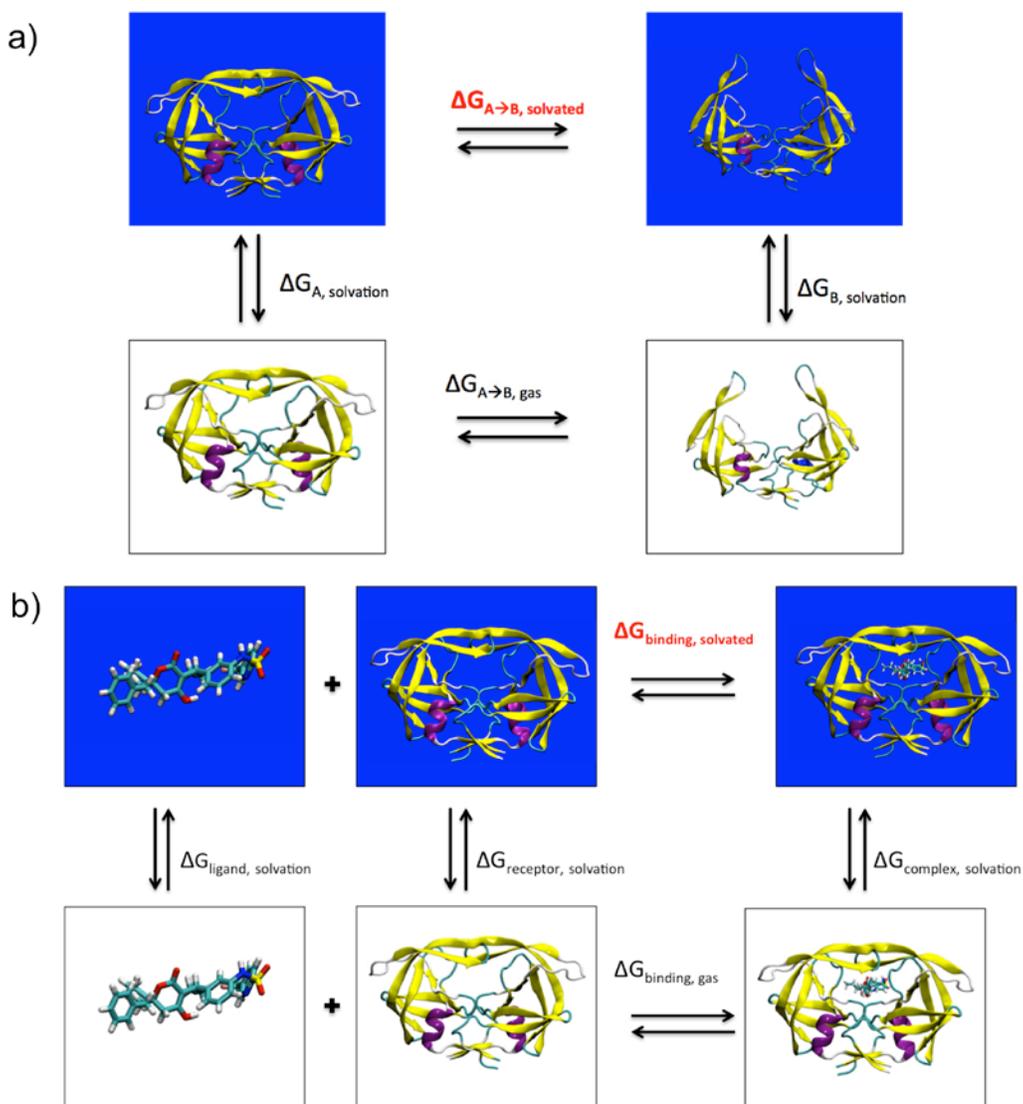


Figure 5-1. Thermodynamic cycles for common free energy calculations. Example cycles are shown for a) stability free energy calculations for conformation A and B, and b) binding free energy calculations for a protein-ligand complex. Solvated systems are shown in blue boxes, while systems in the gas phase are in white boxes. Free energies that are directly calculated are shown in black, while the free energy of interest is shown in red.

5.2.1 Stability and Binding Free Energy Calculations

End-state calculations are frequently used for two types of analyses—calculating the relative stability of multiple conformations of a system and calculating the binding free energy in a non-covalently bound, receptor-ligand complex,¹⁷⁵ shown as thermodynamic cycles in Figure 5-1. Stability calculations compare the free energies of multiple conformations to determine their relative stability. If we consider the process of a biomolecule changing conformations from state A to state B, then the free energy associated with that conformational change is calculated according to Equation 5-1.

$$\Delta G_{A \rightarrow B, \text{solvated}} = \Delta G_{B, \text{solvated}} - \Delta G_{A, \text{solvated}} \quad (5-1)$$

Similarly, binding free energies are calculated by subtracting the free energies of the unbound receptor and ligand from the free energy of the bound complex, shown in Equation 5-2.

$$\Delta G_{\text{binding, solvated}} = \Delta G_{\text{complex, solvated}} - [\Delta G_{\text{receptor, solvated}} + \Delta G_{\text{ligand, solvated}}] \quad (5-2)$$

The free energy change associated with each term on the right-hand side of Equation 5-1 and 5-2 is estimated according to Equation 5-3.

$$\Delta G_{\text{solvated}} = E_{\text{gas}} + \Delta G_{\text{solvation}} - TS_{\text{solute}} \quad (5-3)$$

In Equation 5-3, the $\Delta G_{\text{solvation}}$ represents a true free energy, since the solvent has been averaged due to the use of an implicit solvent model. We then use $\Delta G_{\text{solvated}}$ in Equation 5-1 and 5-2 for consistency, but it is important to note that at this point, we only have one structure for the solute. The gas-phase energies (E_{gas}) are often the molecular mechanical (MM) energies from the force field, while the solvation free energies ($\Delta G_{\text{solvation}}$) are calculated using an implicit solvent model, and the entropic contribution (S) is estimated using known approximations. The free energy of solvation

is further decomposed as the sum of electrostatic and nonpolar contributions. Several implicit solvent models are available to calculate the solvation free energies, including Generalized Born (GB),¹⁷⁵ Poisson-Boltzmann (PB),¹⁷⁵ and Reference Interaction Site Model (RISM).¹⁷⁶

The energies described in the equations above are single point energies of the system. However, in practice, end-state calculations estimate these energies according to averages from an ensemble of representative structures. For example, expressing Equation 5-3 in terms of averages yields Equation 5-4,

$$\Delta G_{solvated} \cong \langle E_{gas} \rangle + \langle \Delta G_{solvation} \rangle - T \langle S_{solute} \rangle = \frac{1}{N} \sum_{i=1}^N E_{i,gas} + \frac{1}{N} \sum_{i=1}^N \Delta G_{i,solvation} - \frac{T}{N} \sum_{i=1}^N S_{i,solute} \quad (5-4)$$

where i is the index of a particular frame and N is the total number of frames analyzed.

There are two approaches to generating the necessary ensembles for the bound and unbound state of binding energy calculations—all ensembles can be extracted from a single molecular dynamics (MD) or Monte Carlo (MC) trajectory of the bound complex, or trajectories can be generated for each state using separate simulations.²²⁵ These approaches are called the *single trajectory protocol* (STP) and *multiple trajectory protocol* (MTP), respectively, and each approach has distinct advantages and disadvantages.

The STP is less computationally expensive than the MTP because only a single trajectory is required to generate all three ensembles. Furthermore, the internal potential terms (e.g., bonds, angles, and dihedrals) cancel exactly in the STP because the conformations in the bound and unbound ensembles are the same, leading to lower fluctuations and easier convergence in the binding free energy. The STP is appropriate if the receptor and ligand ensembles are comparable in the bound and unbound states.

However, the conformations populating the unbound ensembles typically adopt strained configurations when extracted from the bound state ensemble, thereby over-stabilizing the binding compared to the MTP.

5.2.2 Free Energy Decomposition

Amber²⁰³ provides several schemes to decompose calculated free energies into specific residue contributions using either the GB or PB implicit solvent models,²²⁶ following the work of Gohlke et al.^{219,226} Interactions can be decomposed for each residue by including only those interactions in which one of the residue's atoms is involved—a scheme called *per-residue* decomposition. Alternatively, interactions can be decomposed by specific residue pairs by including only those interactions in which one atom from each of the analyzed residues is participating—a scheme called *pairwise* decomposition. These decomposition schemes can provide useful insights into important interactions in free energy calculations.^{219,226}

It is important to note, though, that solvation free energies using GB and PB are not strictly pairwise decomposable since the dielectric boundary defined between the protein and the bulk solvent is inherently non-local and depends on the arrangement of all atoms in space. Thus, care must be taken when interpreting free energy decomposition results.

An alternative way of decomposing free energies is to introduce specific mutations in the protein sequence and analyze how binding free energies or stabilities are affected.¹⁹⁶ *Alanine scanning*, a technique in which an amino acid in the system is mutated to alanine, can highlight the importance of the electrostatic and steric nature of the original sidechain.²²⁷ Assuming the mutation will have a negligible effect on protein conformation, we can incorporate the mutation directly into each member of the original

ensemble. This avoids the need to perform an additional MD or MC simulation to generate an ensemble for the mutant.

5.2.3 Entropy Calculations

The implicit solvent models used to calculate relative stability and binding free energies in end-state calculations often neglect some contributions to the solute entropy. If we assume that biological systems obey a rigid rotor model, we can calculate the translational and rotational entropies using standard statistical mechanical formulae,²²⁸ and we can approximate the vibrational entropy contribution using one of two methods. First, the vibrational frequencies of normal modes can be calculated at various local minima of the potential energy surface—a method we will refer to as *nmode*.²²⁸ Alternatively, the eigenvalues of the mass-weighted covariance matrix constructed from every member of the ensemble can be approximated as frequencies of global, orthogonal motions—a technique we will refer to as the *quasi-harmonic* approximation.²²⁹ Using either the *nmode* or quasi-harmonic frequencies, we can sum the vibrational entropies of each mode calculated from standard formulae.²²⁸

Nmode calculations are typically computationally demanding for large systems because they require minimizing every frame, building the Hessian matrix, and diagonalizing it to obtain the vibrational frequencies (eigenvalues). Because of the Hessian diagonalization, normal mode calculations scale as roughly $(3N)^3$, where N is the number of atoms in the system. While the quasi-harmonic approach is less computationally expensive, a large number of frames, or members of the ensemble, are typically needed to extrapolate the asymptotic limit of the total entropy for each ensemble, which increases the computational cost of the original simulation.²³⁰

Relative free energy calculations between related systems often assume the solute entropy will be the same for each system being compared, thereby removing the need to explicitly calculate them.¹⁹⁶

5.3 Results and Discussion

Here we will demonstrate many of the calculations that *MMPBSA.py* can perform. First we will describe the general workflow for using *MMPBSA.py* to perform end-state free energy calculations, followed by specific examples.

5.3.1 General Workflow

MMPBSA.py is a program written in Python and *nab*²³¹ that streamlines the procedure of preparing and calculating free energies for an ensemble generated by MD or MC simulations whose general workflow is shown in Figure 5-2. *MMPBSA.py* builds upon original ideas by Irina Massova and Peter Kollman as described in Ref. [196]. The process of calculating binding free energies can be a tedious procedure that *MMPBSA.py* aims to shorten and simplify.

Python is a useful programming language for performing tasks that are not numerically intensive, and because it is available on virtually every platform, Python programs are highly portable. Nucleic Acid Builder²³¹ (*nab*), a molecule-based programming language included with AmberTools, contains functionality pertinent to building, manipulating, and performing energy calculations on biological systems, such as proteins and nucleic acids.

End-state calculations often require multiple *topology files* that contain the parameters corresponding to the force field. We recommend users run simulations using explicit solvent, which would require the user to provide both solvated and unsolvated topology files to *MMPBSA.py*. It is necessary that all topology files have a

consistent set of parameters, especially for binding free energy calculations.

MMPBSA.py checks the user's topology files prior to binding free energy calculations to prevent erroneous results due to inconsistencies that may not be immediately obvious.

The Python utility *ante-MMPBSA.py* (released alongside *MMPBSA.py*) allows a user to easily create topology files with a consistent set of parameters, including changing the intrinsic implicit solvent radius set to fit the desired solvent model.

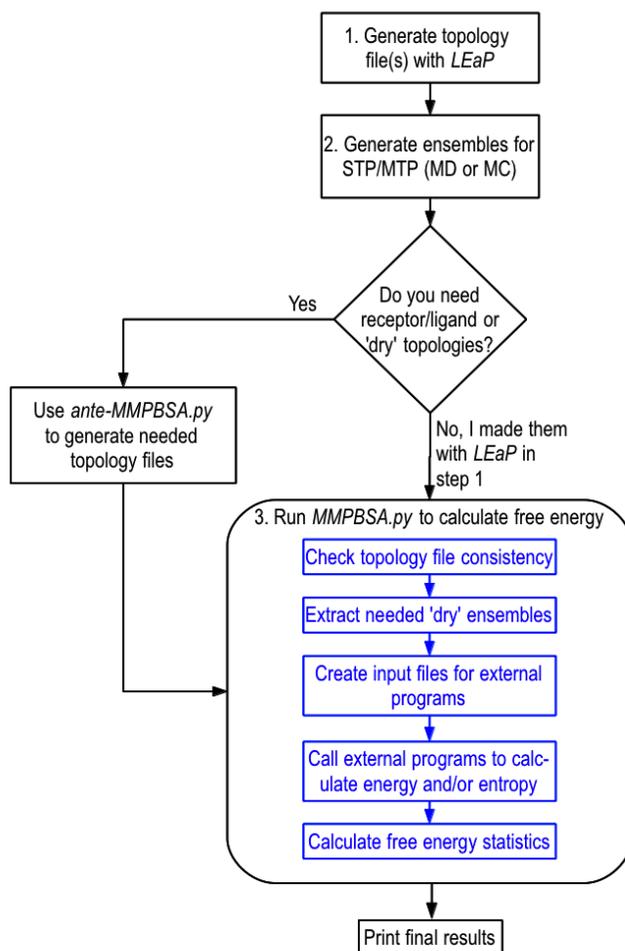


Figure 5-2. General workflow for performing end-state calculations with *MMPBSA.py*. *LEaP* is a program in Amber used to create topology files for dynamics. The workflow shown in step 3 is the series of steps that *MMPBSA.py* automates. 'Dry' topologies and ensembles are systems without explicit solvent that are subsequently treated using an implicit solvent model. 'External programs' refers to the executables that perform the energy calculations (e.g. *sander*).

The use of *MMPBSA.py* is similar to that of Amber's MD engines *sander* and *pmemd*. The command-line flags common to both *MMPBSA.py* and the MD engines are identical, and input files contain input records separated with similar, Fortran-style namelists, indicated with an ampersand (&).

The *MMPBSA.py* input file contains a `&general` namelist for variables that control general behavior. For example, variables that control the subset of frames analyzed (`startframe`, `endframe`, and `interval`) and the amount of information printed in the output file (`verbose`) are specified here. An example of this section is shown below.

```
General MMPBSA.py input file
&general
    startframe=1, endframe=100, interval=2,
    keep_files=0, verbose=1, strip_mask=:WAT:Cl-:Na+,
/
```

Users should avoid analyzing correlated structures from their simulations. Because the autocorrelation time of the free energy is system dependent, the autocorrelation function of the free energy can be calculated to determine the ideal frequency to extract snapshots to avoid analyzing correlated structures (e.g. 5 ps).²²⁰ The `keep_files` variable controls which temporary files remain after the calculations conclude successfully allowing users to optionally inspect the intermediate steps or regenerate the final output file using a different `verbose` value. By default, *MMPBSA.py* removes all known ions from the original trajectory file via the `strip_mask` variable. This can be modified in the input file, however care must be taken if the user chooses to

retain particular ions or water molecules during the free energy calculations because the implicit solvent models utilized by *MMPBSA.py* are not parameterized to account for these species. Therefore, it is not recommended to include water or ions during the calculations.

The following sections introduce the other namelists allowed in the *MMPBSA.py* input file and their functionality.

5.3.2 Poisson-Boltzmann: MM-PBSA

Solvation free energies can be calculated using the Poisson-Boltzmann (PB) implicit solvent method with a non-polar solvation term based on the solvent accessible surface area (SASA)²²¹ by including the `&pb` namelist in the *MMPBSA.py* input file, demonstrated below.

```
MMPBSA.py input file for running PB
&general
  startframe=1, endframe=100, interval=2,
/
&pb
  istrng=0.1, exdi=80, indi=1.0,
  inp=2, cavity_surften=0.0378, cavity_offset=-0.5692,
  fillratio=4, scale=2.0,
  linit=1000, prbrad=1.4, radiopt=1,
/
```

Users can specify external thermodynamic variables based on experimental conditions to control the calculation, such as the ionic strength (`istrng`) and the external and internal dielectric constants (`exdi` and `indi`, respectively). Variables

specific to the PB calculation are shown on the final three lines of the `&pb` section of the example input file above. The default method for calculating non-polar solvation free energies (`inp`) and its associated variables (`cavity_surften` and `cavity_offset`) are recommended based on a previous optimization of these parameters.²³² The default `fillratio` value of 4.0 is recommended to prevent errors for small molecules lying outside the focusing finite-difference grid that occurs when `fillratio` is set too low. `Scale` is the number of PB grid points per angstrom, and may be adjusted to obtain coarser or finer resolution, although increasing the resolution comes at the cost of computational time. The default value of 2.0 for the `scale` control variable was determined as a compromise between speed and accuracy. The suggested value of 1000 for the number of iterations to perform of the linear PB equation (`limit`) is usually sufficient to reach convergence. A 1.4 Å probe radius (`prbrad`) is recommended for simulations performed with water as the solvent to model the size of a typical water molecule. The optimized radii set (`radiopt`) described previously²³³ shows improved accuracy over other sets, and thus is recommended for most users.

The PB solvation free energy can be calculated using two possible programs. The default solver is `pbsa`,²³⁴ whose routines can be called from either a `nab` program (default) or `sander`, although users can access the third-party *Adaptive Poisson-Boltzmann Solver* (`apbs`) solver²³⁵ by linking `apbs` with `sander` using `iAPBS`.²³⁶

5.3.3 Generalized Born: MM-GBSA

The Generalized Born (GB) implicit solvent method with a SASA term can be used to calculate the solvation free energy²¹⁹ by including the `&gb` namelist in the `MMPBSA.py` input file, demonstrated below.

```

MMPBSA.py input file for running GB

&general

    startframe=10, interval=2, endframe=2010,

/

&gb

    igb=5, saltcon=0.1,

/

```

The GB calculation has its own set of control variables, such as salt concentration (`saltcon`) and the specific GB model (`igb`). Amber provides five different GB models^{205,237–239} that can be used with *MMPBSA.py* depending on user preference. Similar to PB, the GB solvation free energies can be solved using either a *nab* program (default) or *sander*. For analysis of proteins, we recommend setting `igb` to 8 (*sander* only) based on recent results showing an increased stability for salt bridges with this GB model.²³⁹ However, a value of 2 for `igb` is preferred for many users analyzing proteins without access to *sander*.¹⁸⁸

MMPBSA.py can use *sander* to treat part of the system in a MM-GBSA calculation quantum mechanically using a hybrid quantum-mechanical/molecular-mechanical (QM/MM) Hamiltonian²⁴⁰ with the `ifqnt` control variable. The residues treated quantum mechanically (`qm_residues`) are defined in the `&gb` namelist of the *MMPBSA.py* input file, as demonstrated below.

```

MMPBSA.py input file for running QM/MMGBSA

&general

    startframe=7, interval=9, endframe=2011,

```

```

/
&gb
  igb=5, saltcon=0.1, ifqnt=1, qmcharge_com=0,
  qmcharge_lig=0, qm_residues=241,
  qm_theory='PM3',
/

```

The charge of the region treated quantum mechanically must be specified for both the complex (`qmcharge_com`) and ligand (`qmcharge_lig`). Several semi-empirical Hamiltonians are available in *MMPBSA.py* for the QM region,²⁴¹ including AM1,²⁴² MNDO,²⁴³ and PM3.²⁴⁴

5.3.4 Reference Interaction Site Model: MM/3D-RISM

MMPBSA.py can also use the Reference Interaction Site Model (RISM)¹⁷⁶ to calculate solvation free energies through a *nab* program. Specifically, three-dimensional RISM (3D-RISM) determines the equilibrium distribution of solvent around the solute to calculate thermodynamic properties. The MM/3D-RISM method has its own namelist in the *MMPBSA.py* input file with its own control variables. An example is shown below.

```

MMPBSA.py input file for running MM/3D-RISM

&general
  endframe=200,
/

&rism
  closure=kh, polardecomp=1,
/

```

Three different closure relations (`closure`) are available to approximate the RISM integrals: the hypernetted-chain approximation (HNC), Kovalenko-Hirata (KH), and the partial series expansion of order- n (PSE- n).²⁴⁵ While 3D-RISM can directly decompose the solvation free energy into the polar and non-polar (cavity and dispersion) terms (`polardecomp`), this decomposition nearly doubles the computational cost of the calculation, and 3D-RISM calculations already take substantially longer than PB or GB.

5.3.5 Free Energy Decomposition

Free energy decomposition can provide information about the local interactions of a system in addition to global free energies.²¹⁹ *MMPBSA.py* can calculate per-residue and pairwise energy decompositions with *sander*. Decomposition must be combined with GB and/or PB in the *MMPBSA.py* input file, and control variables for both per-residue and pairwise decomposition schemes are specified in the `&decomp` namelist, as shown below.

```
MMPBSA.py input file for running per-residue decomp
&general
  startframe=6, interval=26, endframe=2007,
/
&gb
  igb=5, saltcon=0.1,
/
&decomp
  idecomp=1, print_res='1-10,200-241',
/
```

Per-residue decomposition (`idecomp` options 1 and 2) estimates the contribution of each residue to the total free energy. Pairwise energy decomposition (`idecomp` options 3 and 4) estimates the interaction energy of specific residue pairs. By default, *MMPBSA.py* will print the decomposition results for all residues of the system, but a subset can be specified using `print_res` to reduce the amount of data printed. We recommend new users perform decomposition analysis with GB solvent models before attempting to use the PB models because of the difficulty and time-consuming nature of the PB decomposition analysis. Furthermore, we reemphasize that decomposition analysis is not pairwise decomposable since the dielectric boundary defined between the protein and the bulk solvent is inherently non-local and depends on the arrangement of all atoms in space.

5.3.6 Alanine Scanning

Alanine scanning is an alternative to decomposition analysis that involves mutating a single amino acid in the system to alanine. Alanine scanning is enabled by including the `&alanine_scanning` namelist in the *MMPBSA.py* input file and must be specified with at least one type of calculation to be performed on the mutated structure. An example of alanine scanning using GB is shown below.

```
MMPBSA.py input file for running alanine scanning
&general
  startframe=1, endframe=200, interval=5,
/
&gb
  igb=5, saltcon=0.1,
```

```
/  
&alanine_scanning  
  mutant_only=1,  
/  

```

The free energy can be calculated for the mutant trajectory only (`mutant_only=1`) or both the mutant and original trajectories (`mutant_only=0`). When alanine scanning is enabled, *MMPBSA.py* determines the mutated residue by comparing the original and mutant topology files. Then, *MMPBSA.py* creates a mutant trajectory by modifying the original atomic coordinates of the mutated residue to the atomic positions of alanine, discarding extraneous atoms and adjusting bond lengths to their equilibrium values. Once the mutated trajectory is created, the normal workflow is resumed.

5.3.7 Solute Entropy

Including solute entropy with the energies calculated in the sections above is necessary to obtain true free energies. *MMPBSA.py* assumes a rigid rotor model and calculates the rotational and translational solute entropies according to standard formulae. Vibrational entropies are calculated from the frequencies of global motions, which are obtained via normal mode analysis (`nmode`) or quasi-harmonic analysis.

5.3.7.1 Normal Mode Analysis

In `nmode`, *MMPBSA.py* uses a *nab* program to minimize frames to a local minimum, then constructs and diagonalizes the Hessian matrix to obtain the vibrational frequencies. Control variables for `nmode` calculations are specified in the `&nmode` namelist of the *MMPBSA.py* input file, as demonstrated below.

```
MMPBSA.py input file for running normal mode analysis
```

```

&general
  startframe=1, endframe=200, interval=5,
/
&nmode
  maxcyc=10000, drms=0.001,
  nmode_igb=1, nmode_istrng=0.1,
/

```

Each frame is minimized to a local minimum using the *xmin* minimizer in *nab* prior to building the Hessian matrix. Details of the minimization, such as the number of minimization cycles (*maxcyc*) and the convergence criteria (*drms*), can be adjusted in the input file, however the default values are sufficient for most systems that are able to be minimized. The normal modes can be calculated in the gas-phase or in implicit solvent using a GB model (*nmode_igb*). Nmode calculations can be time-consuming, and they require large amounts of computational resources. For this reason, we advise using only a small number of frames for nmode analysis compared to typical PB and GB analyses. *MMPBSA.py* also provides control variables (*nmstartframe*, *nmendframe*, and *nminterval*) in the *&nmode* section that allow users to define a subset of the frames analyzed by PB or GB for nmode calculations. This provides the users with the ability to combine the PB and/or GB results with the entropic contributions using a different number of frames with only a single call to *MMPBSA.py*.

5.3.7.2 Quasi-harmonic Approximation

Entropy can also be calculated using the quasi-harmonic approximation in which all analyzed frames are aligned to the average structure of the ensemble. The mass-

weighted covariance matrix of all atoms is calculated from these frames and then diagonalized to obtain pseudo-vibrational frequencies. Quasi-harmonic calculations are run when the control variable `entropy` is set to 1 in the `&general` namelist of the `MMPBSA.py` input file. Because minimizations are not required and only one matrix diagonalization is necessary, quasi-harmonic calculations are faster than normal mode calculations. However, obtaining converged results can be difficult because the method depends on the number of frames used to calculate the covariance matrix.

5.3.8 Running in Parallel

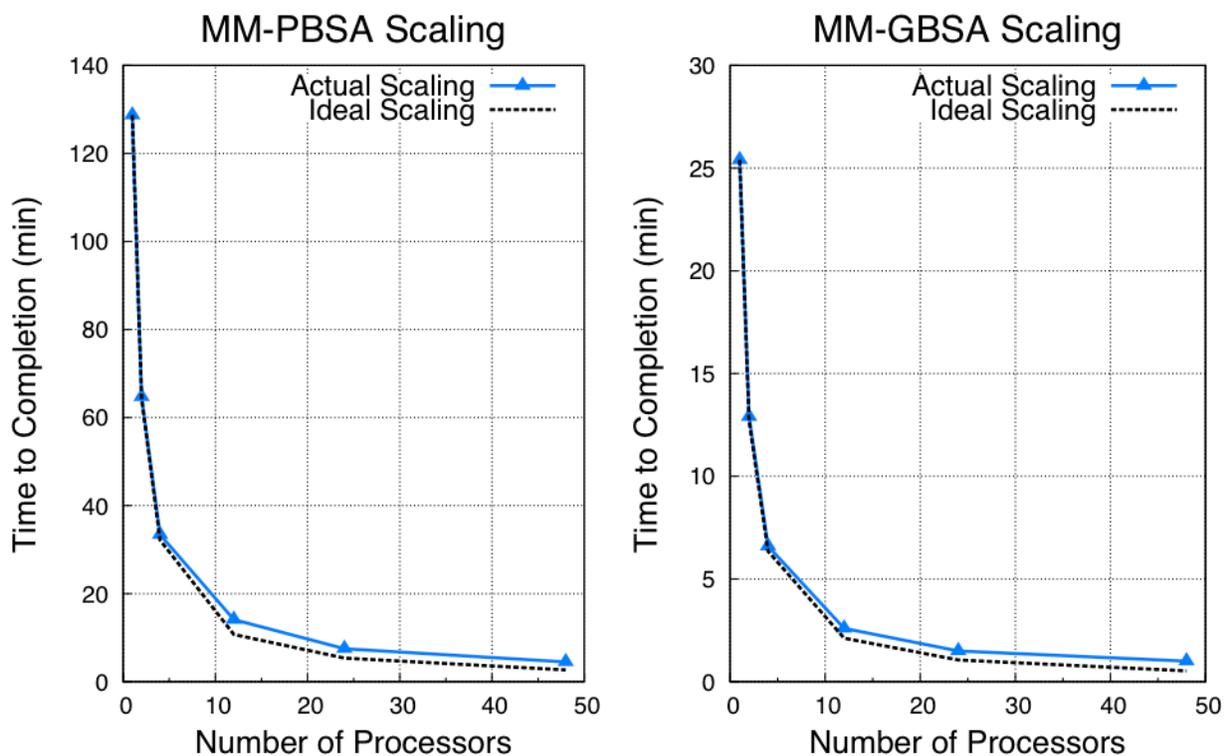


Figure 5-3. *MMPBSA.py* scaling comparison for MM-PBSA and MM-GBSA calculations on 200 frames of a 5910-atom complex. Times shown are the times required for the calculation to finish. Note that MM-GBSA calculations are approximately 5x faster than MM-PBSA calculations. All calculations were performed on NICS Keeneland (2 Intel Westmere 6-core CPUs per node, QDR infiniband interconnect).

MMPBSA.py is implemented in parallel so users with access to multiple processors can speed up their calculations. *MMPBSA.py.MPI* is the parallel implementation of *MMPBSA.py* that uses a Message Passing Interface (MPI) for Python (*mpi4py*). Since energy calculations for each frame are independent, the calculation can be trivially parallelized given enough available processors. *MMPBSA.py.MPI* divides frames evenly across all processors, which allows calculations using many frames to scale better than if *MMPBSA.py* invoked parallel executables to calculate free energies. However, perfect scaling is not attained because certain setups tasks and file input/output can only be done with a single processor. Figure 5-3 demonstrates scaling for a sample MM-PBSA and MM-GBSA calculation.

5.3.9 Differences to *mm_pbsa.pl*

Prior to the development of *MMPBSA.py*, end-state free energy calculations could be performed in Amber using a script written in Perl called *mm_pbsa.pl*. Both programs allow users to perform free energy calculations using the STP and MTP, although *MMPBSA.py* offers more flexibility when using the MTP. Both programs have the ability to use different PB and GB models contained within Amber and estimate entropic contributions. Finally, *MMPBSA.py* and *mm_pbsa.pl* can run free energy calculations in parallel, although only *MMPBSA.py* can run on distributed memory systems (*i.e.*, on multiple nodes connected over a network).

Despite their obvious similarities, there are many differences that exist in their accessibility, implementation, and capabilities. *MMPBSA.py* is available free-of-charge alongside AmberTools, while an Amber license is necessary to obtain *mm_pbsa.pl*. The usage of *MMPBSA.py* is intended to resemble Amber's MD engines for ease of the user, while *mm_pbsa.pl*'s input file and usage has its own syntax. Only *MMPBSA.py*

has an intuitive mechanism for guessing the ligand and receptor masks of a complex based on the topology files provided and analyzes topology files for parameter consistency. Furthermore, only *MMPBSA.py* can calculate entropic contributions to the free energy using the quasi-harmonic approximation. An interface to external PB solvers such as *Delphi*, *MEAD*, and *UHBD* is available with *mm_pbsa.pl* only, although both can use *apbs*. Surface area nonpolar solvation free energies calculated using the *molsurf* program is also only available with *mm_pbsa.pl*. *MMPBSA.py* allows users to provide their own input files for external programs, which gives users the ability to adjust all parameters, not just the variables described in the *MMPBSA.py* manual; in comparison, *mm_pbsa.pl* has no similar functionality without directly altering the source code. Finally, QM/MM-GBSA and MM/3D-RISM calculations are only available through the *MMPBSA.py* implementation.

5.4 Concluding Remarks

MMPBSA.py is a program written in Python and *nab* that allows users to easily and efficiently calculate free energies from MD and MC simulations. Using *MMPBSA.py* simplifies the end-state free energy calculation workflow by automating many tasks and checking topology files for consistency to minimize user error. The command line syntax and input file structure mimics those of other Amber programs, thereby reducing the learning curve. Accompanying test cases provide users with example inputs for performing each supported type of calculation. *MMPBSA.py.MPI* can be run in parallel to increase the speed of end-state free energy calculations.

MMPBSA.py is flexible enough to accommodate the needs of most users. Several methods may be used to calculate the free energy of solvation, such as PB, GB, and RISM. Furthermore, when using a GB model, part of the system can be treated

quantum mechanically. Specific interactions can be analyzed using alanine scanning or free energy decomposition. Solute entropy is approximated using a rigid rotor model in which the vibrational frequencies can be calculated using either normal mode or quasi-harmonic approaches. For increased flexibility, advanced users can supply custom input files to control all variables, including those not available in the *MMPBSA.py* input file.

MMPBSA.py is released with AmberTools under the GNU General Public License (version 2), available for download from <http://ambermd.org/>. Documentation and tutorials are also available on the Amber web site.

CHAPTER 6 CONCLUSIONS

Chagas' disease, often neglected because of its lack of presence in developed countries, was discovered over 100 years ago and currently affects millions of people worldwide, but there is still no single effective drug to treat the lethal disease.

Trypanosoma cruzi, the causative agent of Chagas', expresses a *trans*-sialidase (TcTS) enzyme that helps the parasite evade the host immune system and assists in host cell invasion. TcTS, which has no human analog, is an ideal drug target for the design and discovery of treatments against Chagas' disease. Despite its high sequence and structural similarity to neuraminidases, there are currently no known strong inhibitors of TcTS.

The goal of this study was to design and discover new potential inhibitors that could be used to develop treatments for Chagas' disease. Because of the lack of known actives, molecular docking results of compound fragments to three different TcTS conformations were used to develop energy-optimized pharmacophore (e-pharmacophore) models. The e-pharmacophore models detailed the important protein-ligand interactions that are hypothesized as vital to inhibiting the enzyme. Several motifs of TcTS—arginine triad, Arg53, and acceptor binding site—were already known to be important for catalytic activity from previous studies, but other enzyme residues—Arg93, Asp96, Glu230, and the hydrophobic pocket created by Val95 and Leu176—are suggested as new regions of the active site to target for novel inhibitor design. ZINC's Leads database was screened for hit lead compounds using the e-pharmacophore models and their corresponding features. The virtual screening discovered 82 new potential inhibitors of TcTS, including two compounds—ZINC13359679 and

ZINC02576132—that were calculated to have significantly stronger free energies of binding than the natural TcTS substrate, sialyl-lactose. Both compounds appear stable in the TcTS active site based on MD simulations during the time frame analyzed.

Independent molecular docking of over 35 million compounds was also performed using Schrödinger's Glide on a collection of molecular databases, as well as compounds designed using *de novo* software. The most promising compounds from docking results were further analyzed using MD simulations and MM-GBSA binding free energy calculations. Several of the top compounds from MM-GBSA calculations were designed using AutoGrow or LigBuilder, but these molecules are not necessarily ideal drugs because they are large, flexible, and potentially difficult to synthesize. Other compounds, such as PubChem 54353125 and ZINC04939766, are more easily obtainable for *in vitro* and *in vivo* assays, but their calculated binding free energies suggest they will be weaker inhibitors than those designed by *de novo* methods. Rational drug design was also performed on several potential inhibitors attempting to improve the binding of these lead compounds with mixed results.

The use of end-state free energy calculations has become an important tool for *in silico* drug discovery methods. To aid in the discovery of novel inhibitors of TcTS, a new program—*MMPBSA.py*—was developed that streamlines the calculation of end-state free energies, including binding free energy calculations of protein-ligand complexes. *MMPBSA.py* is user-friendly and provides many options for the calculation of solvation free energies and entropies. Residue interaction energies can also be calculated using energy decomposition and alanine scanning techniques. A parallel implementation of *MMPBSA.py* is available that can significantly reduce the wall time of free energy

calculations when multiple processors are available. This software is freely distributed alongside AmberTools and can be downloaded from the Amber molecular dynamics website (<http://ambermd.org/>).

Drug design is a difficult process that involves the use of many tools with the intention of discovering new inhibitors of enzymes. Unfortunately, computational methods are not rigorous and include many uncertainties that often give spurious results. In this study, the ZINC Leads molecular database was screened for potential TcTS inhibitors—once using Glide docking and once using Phase e-pharmacophore screening. Although both programs are distributed with Schrödinger, they gave very different results. For example, of the two best compounds from e-pharmacophore screening—ZINC02576132 and ZINC13359679—only ZINC02576132 made it through the initial Glide HTVS filtering, and docking of ZINC02576132 to TcTS with Glide XP gave the 7,056th best free energy of binding. These results suggest that given the uncertainty in results from drug design methods, multiple computational approaches should be utilized when attempting to discover novel inhibitors of enzymes. Furthermore, docking methods clearly have their own limitations that suggest the inclusion of more rigorous methods, such as MD and MM-GBSA. As Figure 4-1 shows, there is little correlation between docking results and MM-GBSA binding energies. MM-GBSA calculations have multiple advantages over docking results (e.g. averaging over many conformations, results that are more comparable to experiments, etc.) that suggest the MM-GBSA results are more dependable for drug discovery. The multiple computational strategies employed together in this study are unique in the pursuit of novel inhibitors of TcTS to treat Chagas' disease.

APPENDIX: SUPPLEMENTARY MATERIAL

A.1 Hit Compounds From e-pharmacophore Screening

Table A-1. All hit compounds from screening results of the 1S0I pharmacophore according to MM-GBSA binding free energies. The Phase Fitness scores are unitless, while the MM-GBSA energies have units of kcal/mol. The uncertainties reported represent the standard error of the mean.

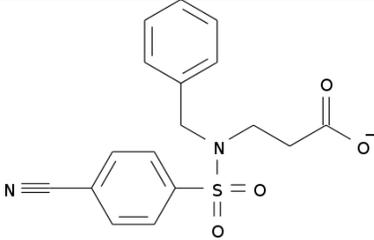
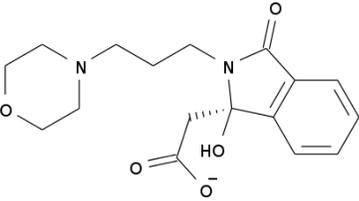
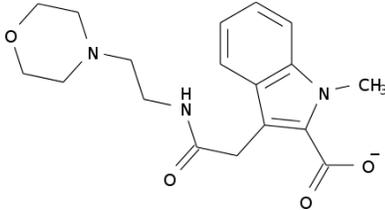
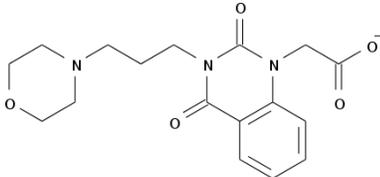
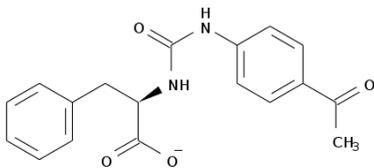
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
1	ZINC13359679		0.184	-76.3 ± 0.4
2	ZINC23187566		0.173	-58.0 ± 0.7
3	ZINC20107259		0.187	-56.4 ± 0.8
4	ZINC73658621		0.177	-53.5 ± 0.6
5	ZINC06614928		0.199	-49.2 ± 1.1

Table A-1. Continued

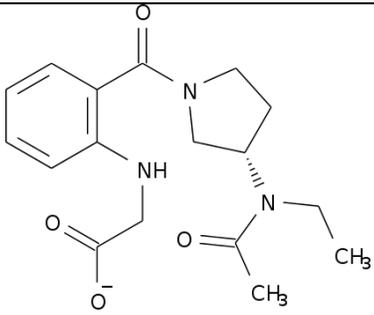
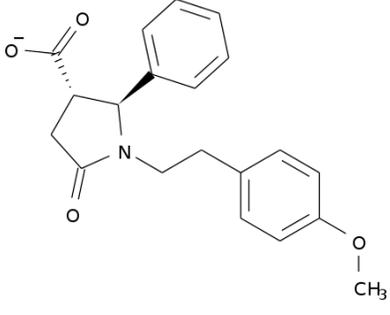
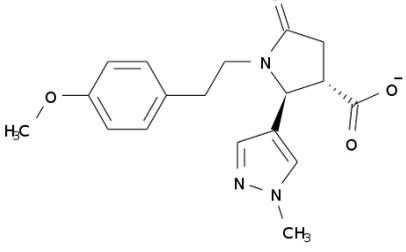
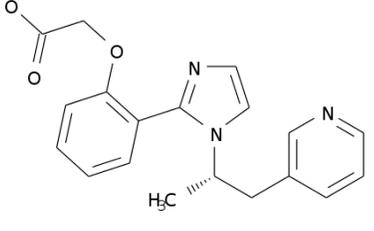
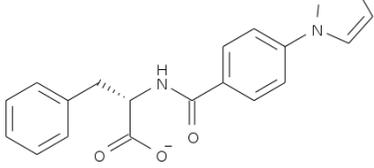
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
6	ZINC65394612		0.136	-47.2 ± 0.3
7	ZINC49466481		0.226	-45.2 ± 0.4
8	ZINC69491994		0.211	-44.8 ± 0.3
9	ZINC71416387		0.144	-42.1 ± 0.9
10	ZINC13259915		0.194	-40.3 ± 0.3

Table A-1. Continued

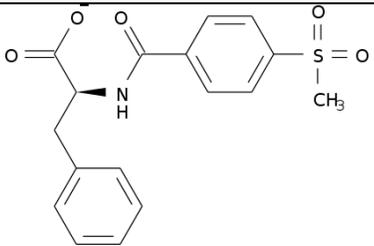
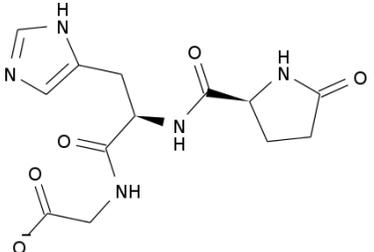
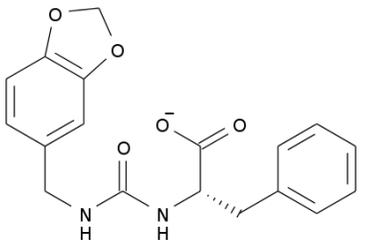
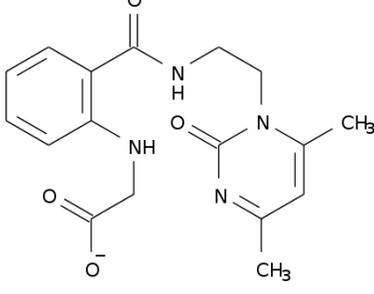
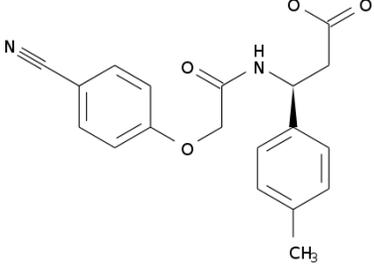
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
11	ZINC08132802		0.231	-38.9 ± 0.3
12	ZINC58184798		0.132	-37.5 ± 0.3
13	ZINC03863603		0.212	-36.7 ± 0.4
14	ZINC65373523		0.121	-36.5 ± 0.4
15	ZINC13179775		0.191	-34.9 ± 0.7

Table A-1. Continued

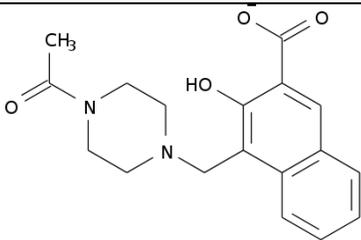
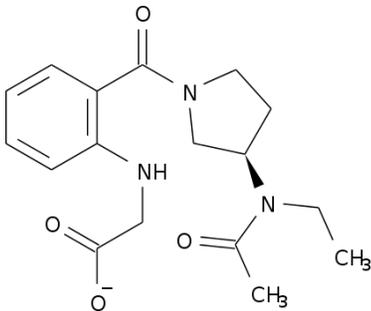
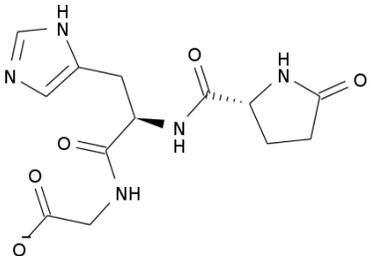
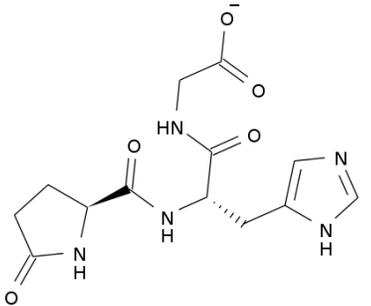
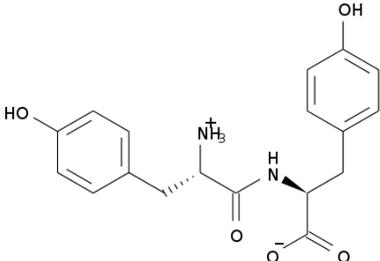
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
16	ZINC25703562		0.165	-34.3 ± 0.3
17	ZINC65394611		0.133	-34.2 ± 0.4
18	ZINC58184800		0.186	-34.2 ± 0.5
19	ZINC38192570		0.191	-33.3 ± 0.4
20	ZINC02004481		0.220	-27.5 ± 0.3

Table A-1. Continued

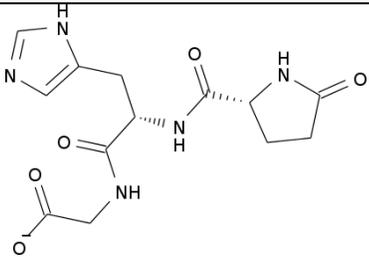
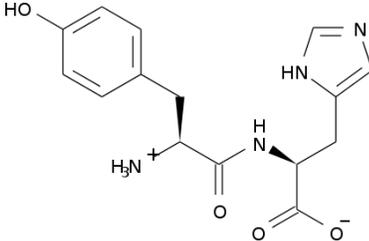
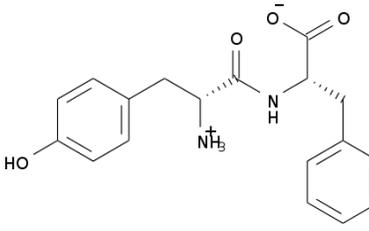
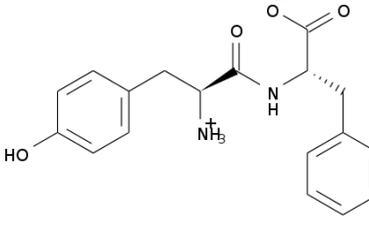
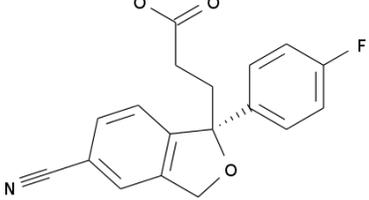
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
21	ZINC58184799		0.196	-26.9 ± 0.3
22	ZINC34349844		0.212	-26.2 ± 0.3
23	ZINC05273977		0.217	-21.0 ± 0.2
24	ZINC02384820		0.212	-16.6 ± 0.4
25	ZINC30731228		0.153	-4.6 ± 0.4

Table A-2. All hit compounds from screening results of the holo pharmacophore according to MM-GBSA binding free energies. The Phase Fitness scores are unitless, while the MM-GBSA energies have units of kcal/mol. The uncertainties reported represent the standard error of the mean.

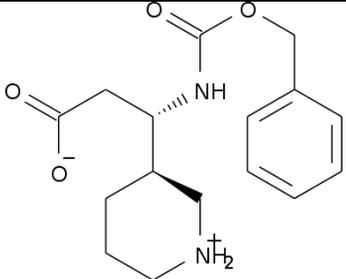
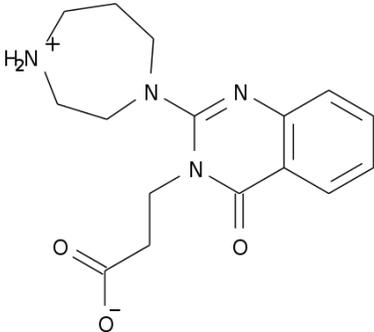
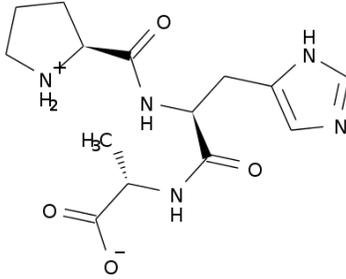
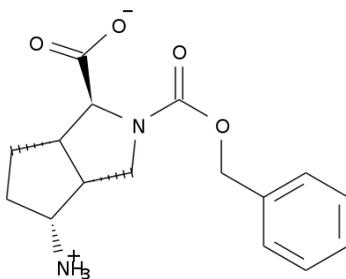
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
1	ZINC02576132		0.194	-73.2 ± 0.5
2	ZINC40351842		0.145	-55.1 ± 0.4
3	ZINC60181779		0.198	-52.2 ± 0.4
4	ZINC19795370		0.184	-48.5 ± 1.0

Table A-2. Continued

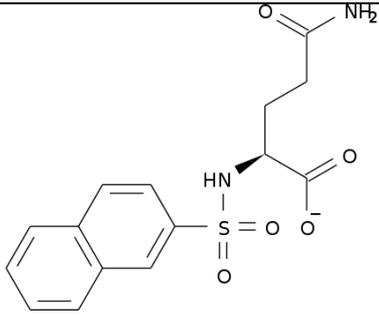
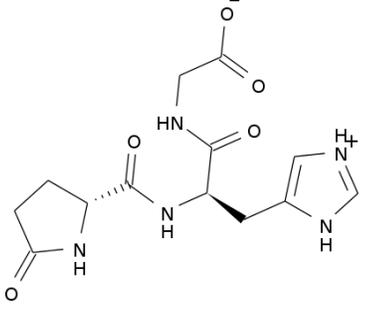
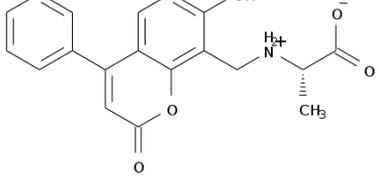
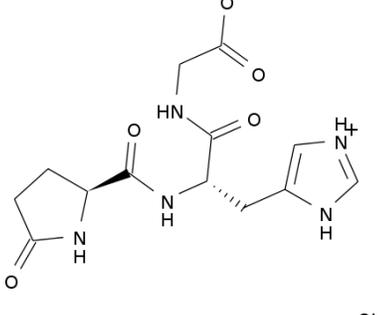
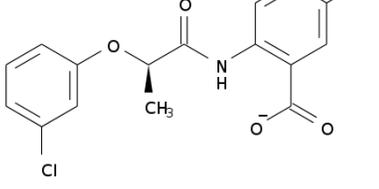
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
5	ZINC03406782		0.197	-45.0 ± 0.5
6	ZINC58184800		0.176	-40.5 ± 0.5
7	ZINC00405960		0.240	-39.4 ± 0.7
8	ZINC38192570		0.196	-38.1 ± 0.5
9	ZINC17903516		0.192	-36.7 ± 0.6

Table A-2. Continued

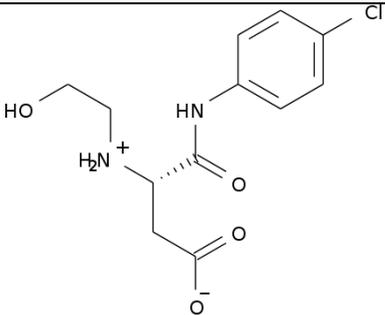
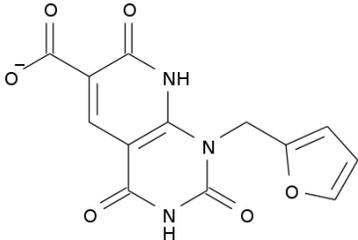
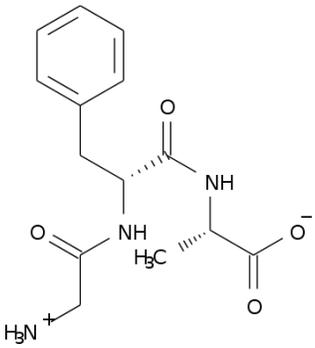
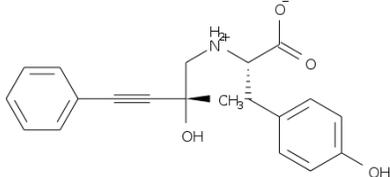
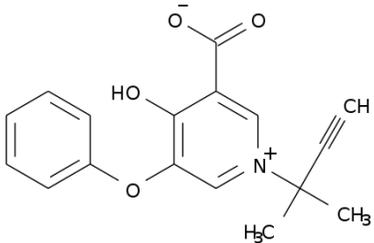
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
10	ZINC72334048		0.129	-36.4 ± 0.3
11	ZINC72417077		0.180	-35.9 ± 0.3
12	ZINC04545875		0.204	-33.4 ± 0.4
13	ZINC03168774		0.155	-33.1 ± 0.5
14	ZINC65395904		0.164	-32.8 ± 0.2

Table A-2. Continued

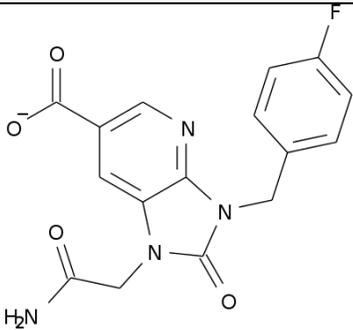
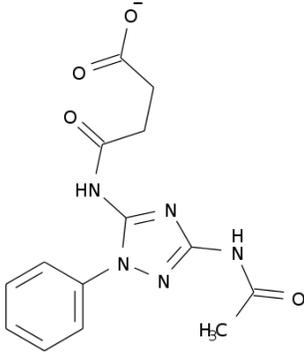
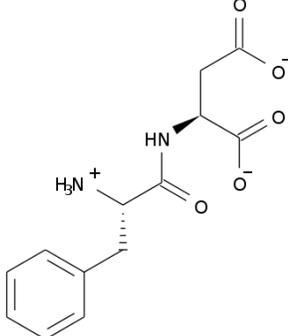
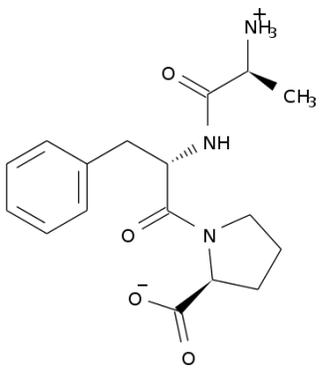
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
15	ZINC55211759		0.168	-32.6 ± 0.4
16	ZINC00550909		0.183	-32.2 ± 0.6
17	ZINC01579638		0.157	-31.8 ± 0.5
18	ZINC02560943		0.204	-29.8 ± 0.5

Table A-2. Continued

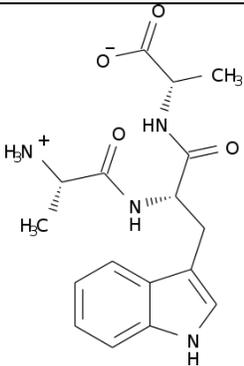
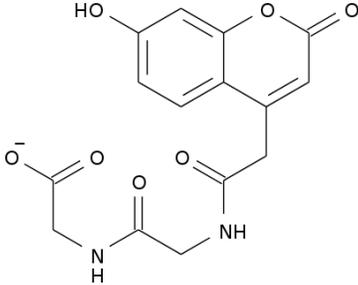
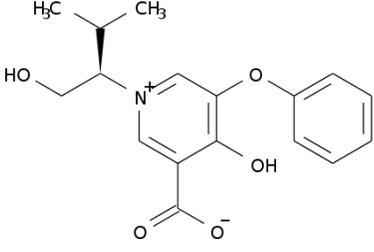
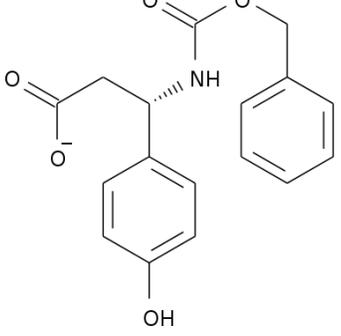
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
19	ZINC02560953		0.213	-29.8 ± 0.8
20	ZINC02107849		0.176	-29.4 ± 0.4
21	ZINC55405006		0.176	-29.4 ± 0.3
22	ZINC02512042		0.157	-28.9 ± 0.7

Table A-2. Continued

Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
23	ZINC38580567	<chem>O=C(NC1=CC=C(C=C1)COC(=O)C2=CC=C(O)C=C2)C3=CC=CC=C3</chem>	0.185	-28.2 ± 0.7
24	ZINC72320181	<chem>O=C(NC1=CC=C(C=C1)COC(=O)C2=CC=C(O)C=C2)C3=CC=CC=C3</chem>	0.133	-26.0 ± 0.3
25	ZINC39715047	<chem>O=C(NC1=CC=C(C=C1)COC(=O)C2=CC=C(O)C=C2)C3=CC=CC=C3</chem>	0.185	-23.4 ± 0.6
26	ZINC72154304	<chem>O=C(NC1=CC=C(C=C1)COC(=O)C2=CC=C(O)C=C2)C3=CC=CC=C3</chem>	0.197	-23.1 ± 0.5
27	ZINC71788361	<chem>O=C(NC1=CC=C(C=C1)COC(=O)C2=CC=C(O)C=C2)C3=CC=CC=C3</chem>	0.211	-22.5 ± 0.3

Table A-2. Continued

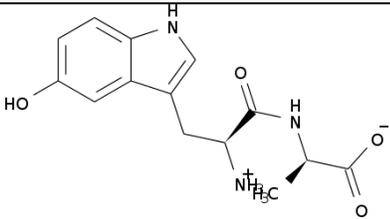
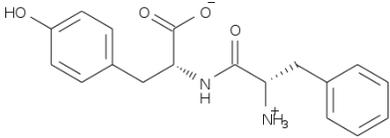
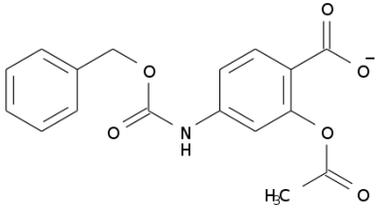
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
28	ZINC04016135		0.125	-22.4 ± 0.3
29	ZINC28631632		0.154	-19.4 ± 0.4
30	ZINC71769780		0.170	-18.8 ± 0.5

Table A-3. All hit compounds from screening results of the apo pharmacophore according to MM-GBSA binding free energies. The Phase Fitness scores are unitless, while the MM-GBSA energies have units of kcal/mol. The uncertainties reported represent the standard error of the mean.

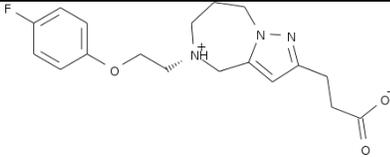
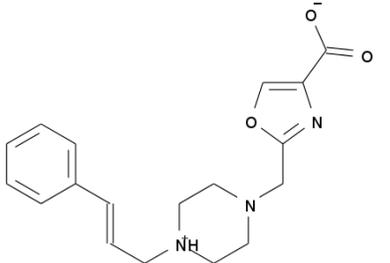
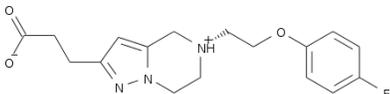
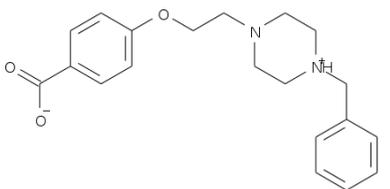
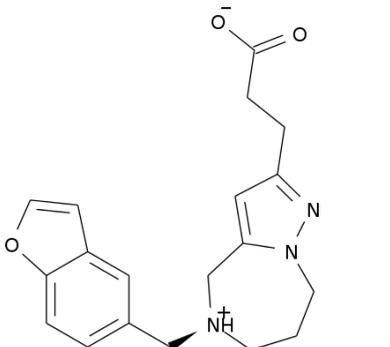
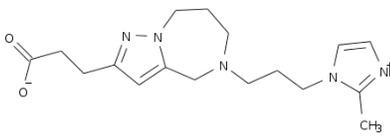
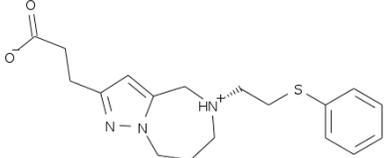
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
1	ZINC72447098		0.155	-54.8 ± 0.3
2	ZINC67946325		0.186	-50.1 ± 0.4
3	ZINC72145690		0.144	-44.0 ± 0.5
4	ZINC19850530		0.168	-42.0 ± 0.3
5	ZINC72432649		0.168	-41.8 ± 0.3
6	ZINC72477708		0.145	-38.6 ± 0.4
7	ZINC72407055		0.155	-37.0 ± 0.3

Table A-3. Continued

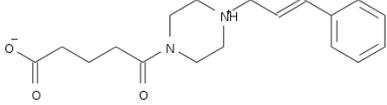
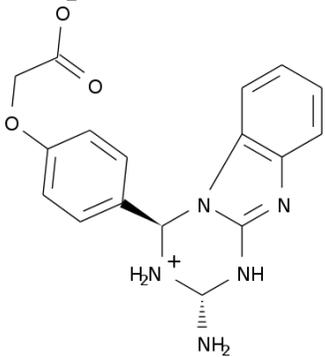
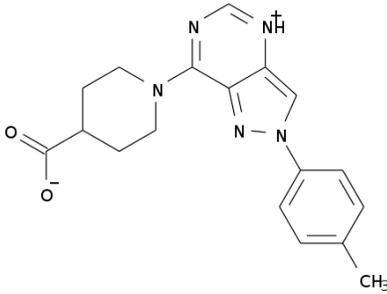
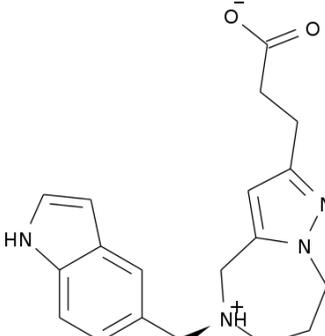
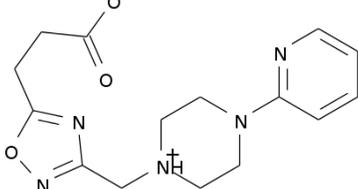
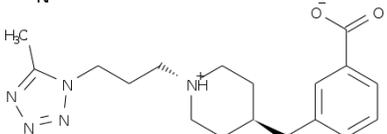
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
8	ZINC23379459		0.169	-36.8 ± 0.5
9	ZINC01432054		0.154	-34.3 ± 0.6
10	ZINC32919121		0.201	-33.8 ± 0.4
11	ZINC72429720		0.168	-33.6 ± 0.3
12	ZINC49453236		0.136	-32.6 ± 0.3
13	ZINC72173820		0.164	-32.5 ± 0.6

Table A-3. Continued

Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
14	ZINC72419466		0.155	-32.1 ± 0.3
15	ZINC19874293		0.140	-30.7 ± 0.5
16	ZINC72433326		0.170	-29.4 ± 0.4
17	ZINC13123936		0.174	-28.9 ± 0.3
18	ZINC44551566		0.143	-28.6 ± 0.3
19	ZINC72431004		0.163	-28.3 ± 0.3

Table A-3. Continued

Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
20	ZINC72369398		0.120	-26.8 ± 0.5
21	ZINC02384991		0.234	-24.3 ± 0.5
22	ZINC72480550		0.121	-22.5 ± 0.5
23	ZINC65384219		0.178	-22.2 ± 0.5
24	ZINC71614084		0.152	-19.6 ± 0.4
25	ZINC71751320		0.196	-15.5 ± 0.2
26	ZINC72160907		0.189	-15.0 ± 0.3
27	ZINC07518893		0.108	-11.1 ± 0.4

A.2 Potential Inhibitors of TcTS From Molecular Docking and *de novo* Design

Table A-4. Compounds from molecular docking and *de novo* design considered as potential inhibitors of TcTS. All compounds were subjected to at least 10 ns of MD and ranked according to the corresponding MM-GBSA free energies of binding. MM-GBSA energies have units of kcal/mol, and the uncertainties reported represent the standard error of the mean.

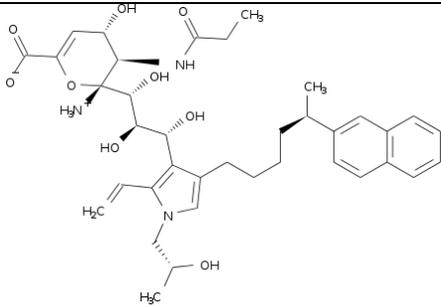
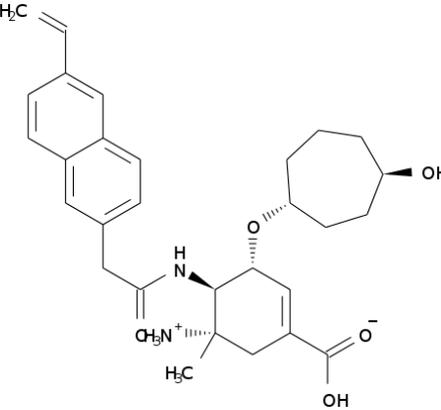
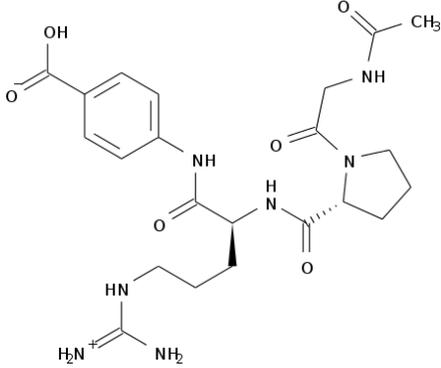
Rank	Ligand Name	Structure	Source	MM-GBSA
1	DANA Result 034	 The structure of DANA Result 034 is a complex molecule featuring a central indazole ring system. It is substituted with a 2-hydroxyethylamino group, a methyl group, a hydroxyl group, and a propyl chain that terminates in a naphthalene ring system. Other substituents include a methyl group and a hydroxyl group on the indazole ring.	LigBuilder	-91.2 ± 0.1
2	Oseltamivir Result 70	 The structure of Oseltamivir Result 70 is a complex molecule featuring a central piperidine ring system. It is substituted with a 4-vinylphenyl group, a methyl group, a hydroxyl group, and a carboxylate group. Other substituents include a methyl group and a hydroxyl group on the piperidine ring.	LigBuilder	-72.5 ± 0.2
3	PubChem 54353125	 The structure of PubChem 54353125 is a complex molecule featuring a central piperidine ring system. It is substituted with a 4-carboxyphenyl group, a methyl group, a hydroxyl group, and a carboxylate group. Other substituents include a methyl group and a hydroxyl group on the piperidine ring.	PubChem	-72.1 ± 0.2

Table A-4. Continued

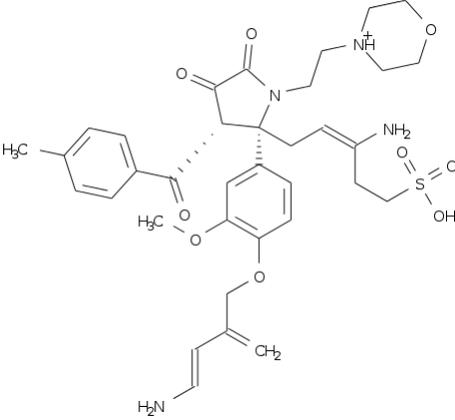
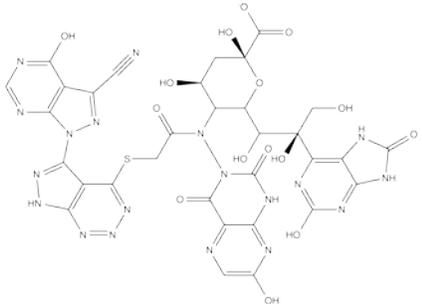
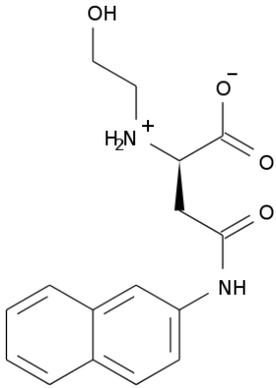
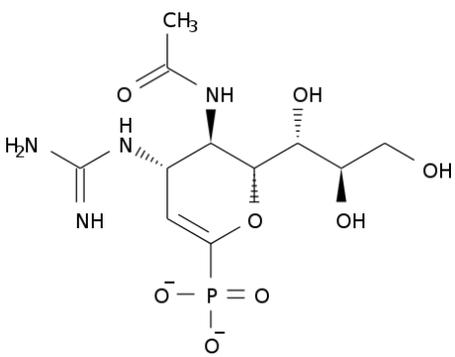
Rank	Ligand Name	Structure	Source	MM-GBSA
4	ZINC16247520 Rank 10		AutoGrow	-62.0 ± 0.2
5	Sialic Acid Rank 1		AutoGrow	-61.5 ± 0.3
6	ZINC04939766		ZINC Leads	-58.5 ± 0.2
7	Zanamivir phosphonate- Rank 62		Literature	-57.9 ± 0.3

Table A-4. Continued

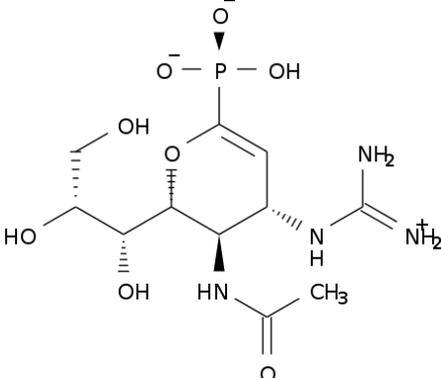
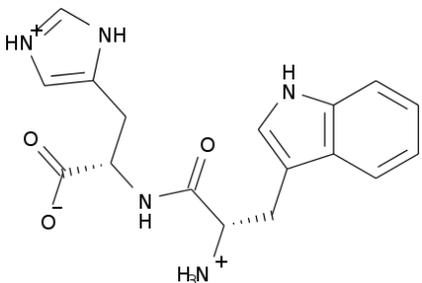
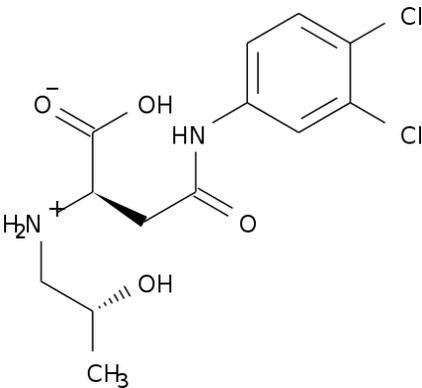
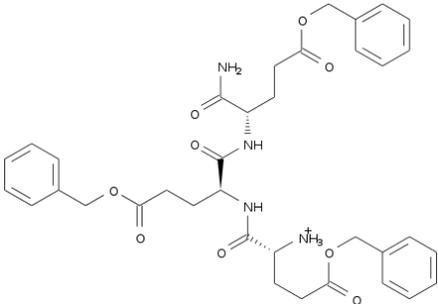
Rank	Ligand Name	Structure	Source	MM-GBSA
8	Zanamivir phosphonate+ Rank 1		Literature	-55.8 ± 0.3
9	ZINC34334192		ZINC Leads	-54.7 ± 0.2
10	ZINC04280204		ZINC Leads	-54.6 ± 0.2
11	Asinex 380341		Asinex	-54.3 ± 0.4

Table A-4. Continued

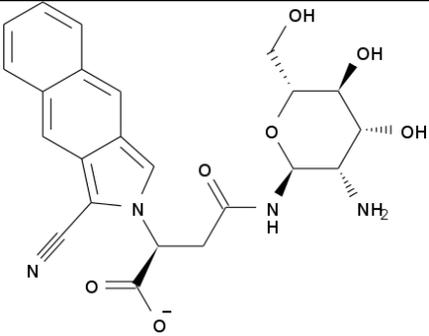
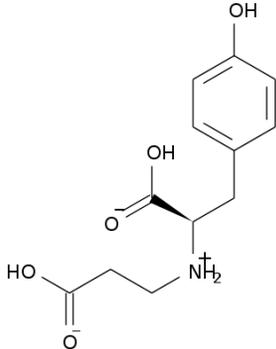
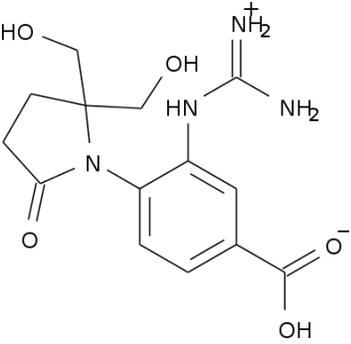
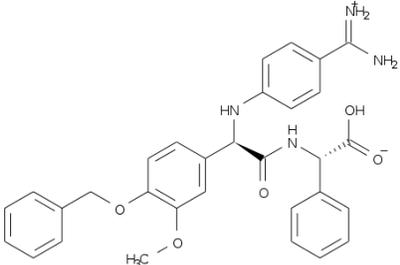
Rank	Ligand Name	Structure	Source	MM-GBSA
12	PubChem 12000964		PubChem	-54.1 ± 0.2
13	ZINC02044476		ZINC Leads	-53.0 ± 0.3
14	PubChem 444507		PubChem	-51.3 ± 0.3
15	PDB 1W7X ligand		1W7X PDB	-50.8 ± 0.2

Table A-4. Continued.

Rank	Ligand Name	Structure	Source	MM-GBSA
16	ZINC20472820		ZINC ChemBridge	-50.1 ± 0.1
17	ZINC31467426		ZINC DrugLike	-49.5 ± 0.5
18	ZINC59351056		ZINC Leads	-46.9 ± 0.3
19	ZINC16247520 Rank 5		AutoGrow	-46.8 ± 0.6

Table A-4. Continued.

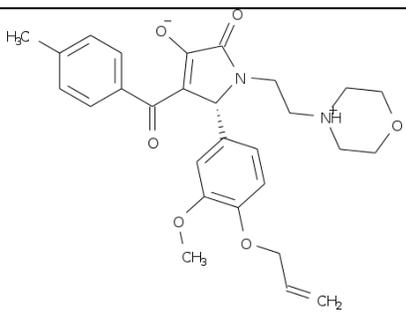
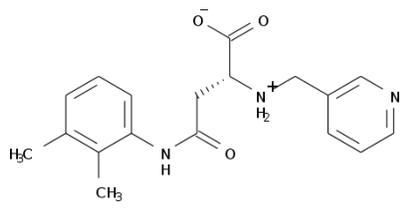
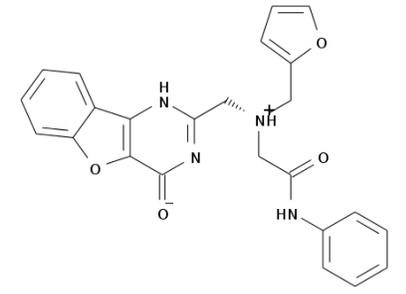
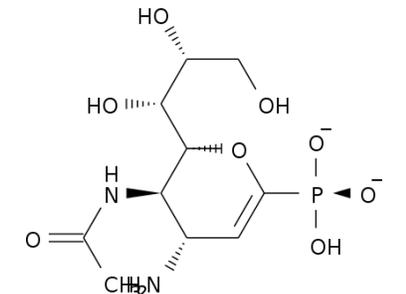
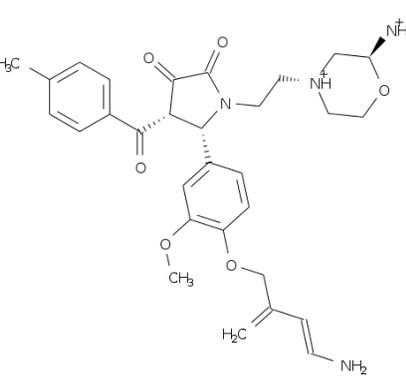
Rank	Ligand Name	Structure	Source	MM-GBSA
20	ZINC16247520		ZINC ChemBridge	-46.6 ± 0.3
21	ZINC00195758		ZINC Leads	-45.0 ± 0.2
22	ZINC24812755		ZINC DrugLike	-44.9 ± 0.3
23	4-amino DANA phosphonate		Literature	-44.2 ± 0.4
24	ZINC16247520 Rank 1		AutoGrow	-43.7 ± 0.2

Table A-4. Continued.

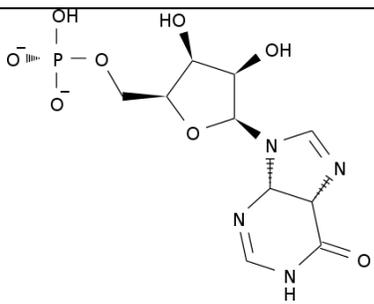
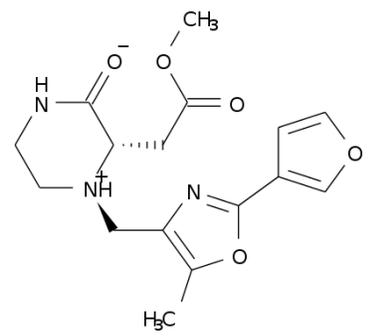
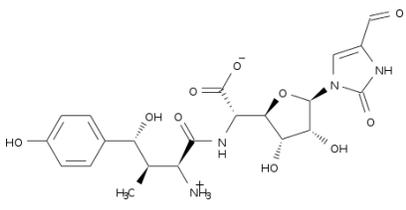
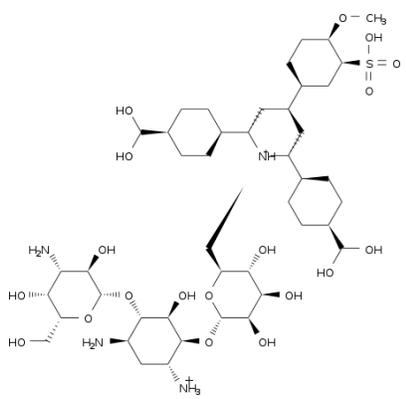
Rank	Ligand Name	Structure	Source	MM-GBSA
25	ZINC35112851		ZINC Leads	-43.6 ± 0.2
26	ZINC39394607		ZINC Leads	-43.4 ± 0.2
27	ZINC38577973		ZINC PubChem	-42.2 ± 0.4
28	Ligand 7442		Katritzky Set	-42.9 ± 0.3

Table A-4. Continued.

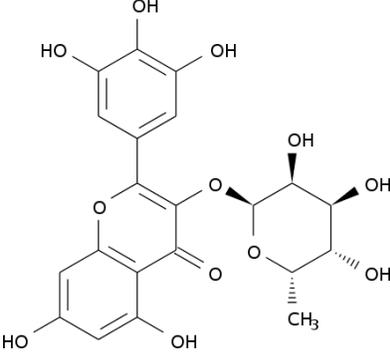
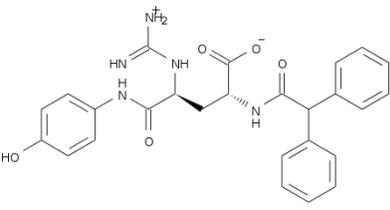
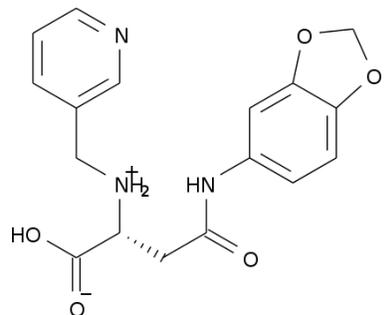
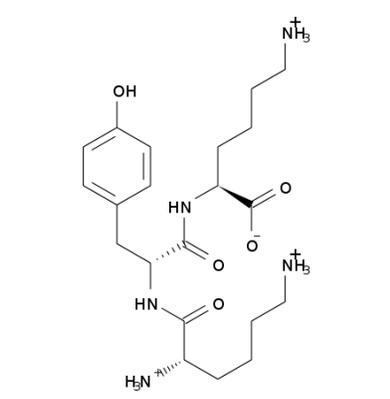
Rank	Ligand Name	Structure	Source	MM-GBSA
29	277		NCI Diversity Set II	-42.8 ± 0.1
30	ZINC38455214		ZINC PubChem	-42.2 ± 0.2
31	ZINC28232161		ZINC Leads	-42.1 ± 0.2
32	ZINC13298421		ZINC PubChem	-41.7 ± 0.8

Table A-4. Continued.

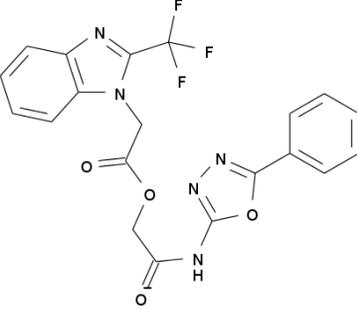
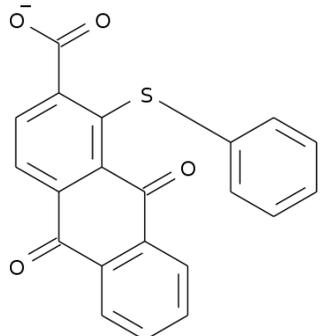
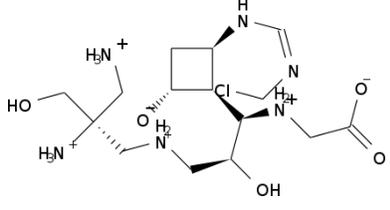
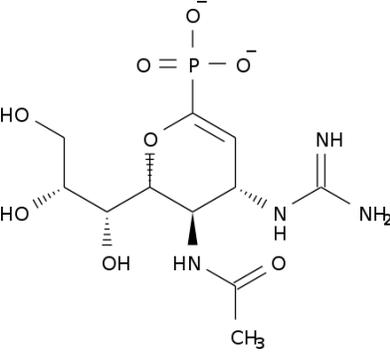
Rank	Ligand Name	Structure	Source	MM-GBSA
33	ZINC21151234		ZINC DrugLike	-41.2 ± 0.2
34	486		NCI Diversity Set II	-40.4 ± 0.1
35	Super Fragment		4 ZINC Fragments Combined	-39.4 ± 0.3
36	Zanamivir phosphonate- Rank 3		Literature	-39.3 ± 0.5

Table A-4. Continued.

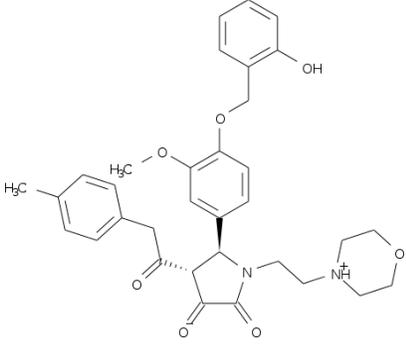
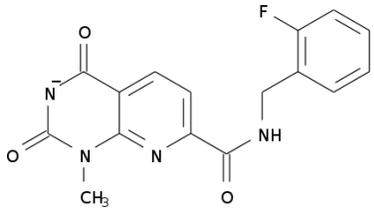
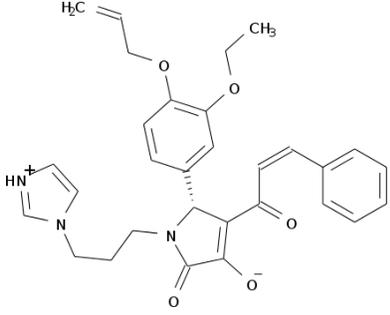
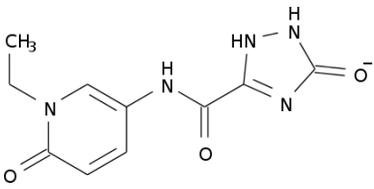
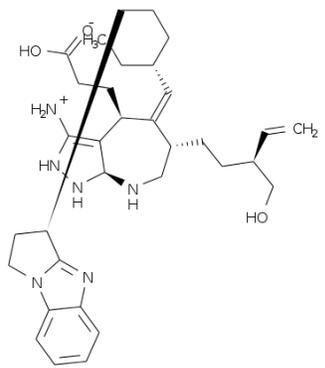
Rank	Ligand Name	Structure	Source	MM-GBSA
37	ZINC16247520 CH ₂ ArOH CH ₂		ZINC ChemBridge Modified	-38.8 ± 0.3
38	ZINC31471796		ZINC DrugLike	-38.4 ± 0.6
39	ZINC35444554		ZINC PubChem	-38.0 ± 0.4
40	ZINC48745899		ZINC Enamine	-37.8 ± 0.2
41	Fragments Combo		ZINC Fragments + LigBuilder	-37.7 ± 0.2

Table A-4. Continued.

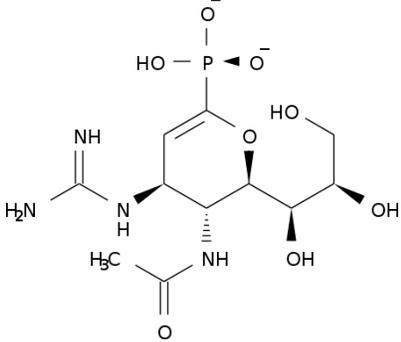
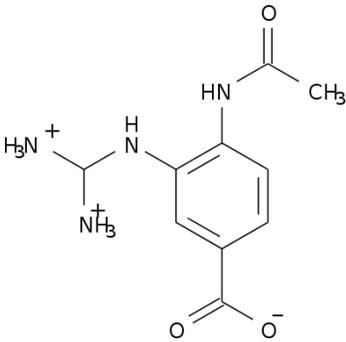
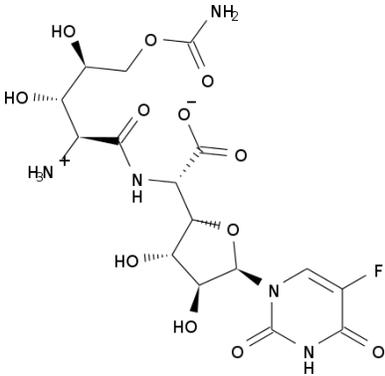
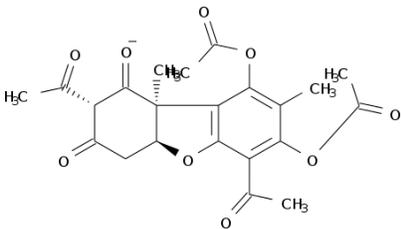
Rank	Ligand Name	Structure	Source	MM-GBSA
42	Zanamivir phosphonate- Rank 1		Literature	-37.6 ± 0.2
43	ZINC19328706		ZINC DrugBank Experimental	-37.5 ± 0.2
44	ZINC14676974		ZINC PubChem	-37.4 ± 0.2
45	571		NCI Diversity Set II	-36.7 ± 0.2

Table A-4. Continued.

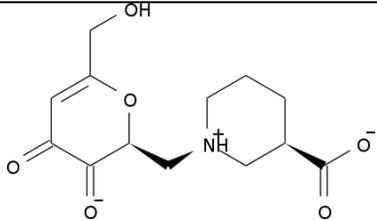
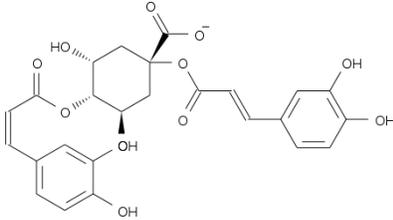
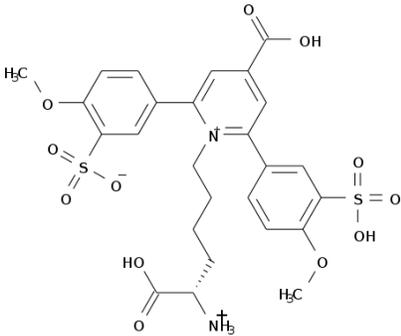
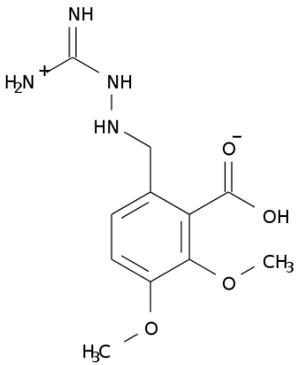
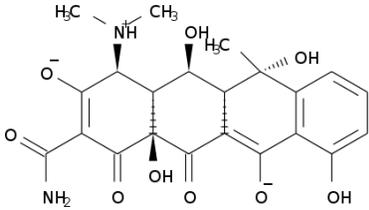
Rank	Ligand Name	Structure	Source	MM-GBSA
46	ZINC62001775		ZINC Leads	-36.1 ± 0.2
47	451		NCI Diversity Set II	-35.5 ± 0.1
48	Ligand 6400		Katritzky Set	-35.3 ± 0.4
49	ZINC17192809		ZINC Leads	-35.2 ± 0.2
50	ZINC18202350		ZINC Pharmeks	-34.6 ± 0.3

Table A-4. Continued.

Rank	Ligand Name	Structure	Source	MM-GBSA
51	ZINC35183118		ZINC Leads	-34.2 ± 0.1
52	ZINC31572053		ZINC DrugLike	-33.9 ± 0.2
53	ZINC41540804		ZINC PBMR Labs	-33.9 ± 0.2
54	ZINC00119770		ZINC Leads	-32.9 ± 0.2
55	PubChem 44428315		PubChem	-32.9 ± 0.2

Table A-4. Continued.

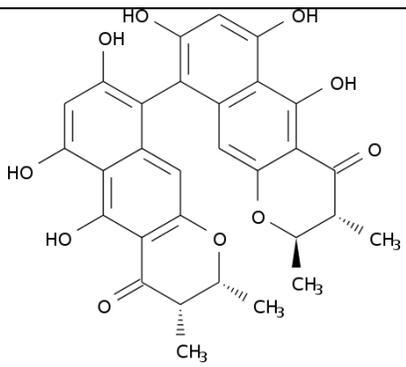
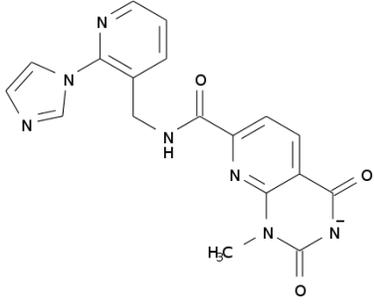
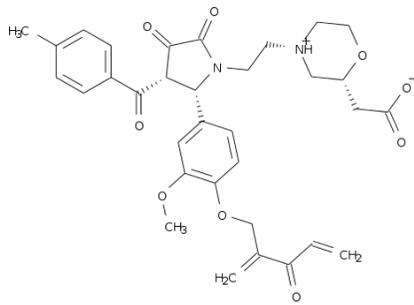
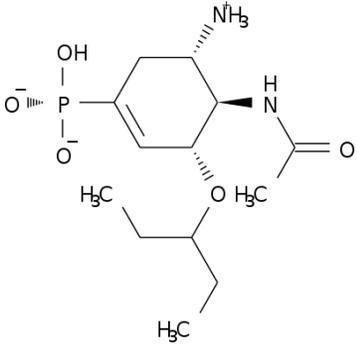
Rank	Ligand Name	Structure	Source	MM-GBSA
56	1385		NCI Diversity Set II	-32.2 ± 0.2
57	ZINC29022478		ZINC DrugLike	-31.8 ± 0.4
58	ZINC16247520 Rank 2		AutoGrow	-31.3 ± 0.2
59	Oseltamivir phosphonate		Literature	-31.0 ± 0.2

Table A-4. Continued.

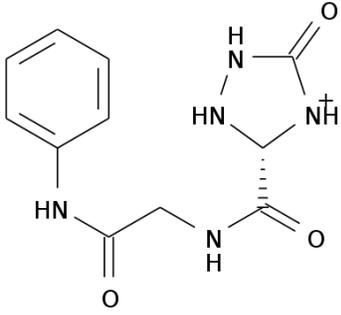
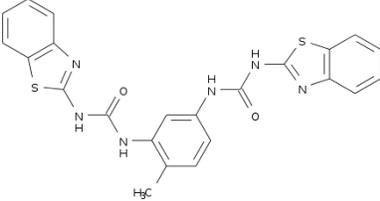
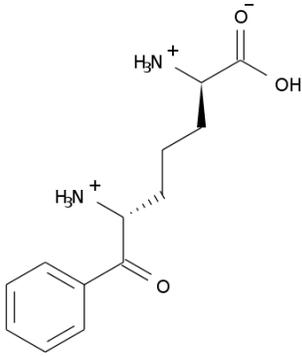
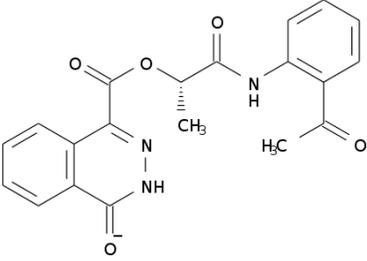
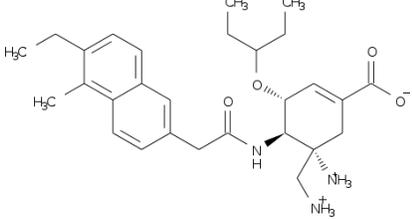
Rank	Ligand Name	Structure	Source	MM-GBSA
60	ZINC48736239		ZINC Enamine	-30.4 ± 0.3
61	639		NCI Diversity Set II	-30.0 ± 0.1
62	ZINC01675031		ZINC Leads	-29.2 ± 0.2
63	ZINC07057776		ZINC DrugLike	-28.5 ± 0.3
64	Oseltamivir phosphonate Rank 6		LigBuilder	-27.8 ± 0.2

Table A-4. Continued.

Rank	Ligand Name	Structure	Source	MM-GBSA
65	ZINC03945038		ZINC Pharmeks	-27.8 ± 0.3
66	ZINC29020993		ZINC DrugLike	-27.5 ± 0.5
67	ZINC08995828		ZINC Pharmeks	-26.9 ± 0.2
68	ZINC05819288		ZINC Leads	-26.2 ± 0.1
69	ZINC31000655		ZINC DrugLike	-23.6 ± 0.1
70	ZINC50027772		ZINC Leads	-20.7 ± 0.2

Table A-4. Continued.

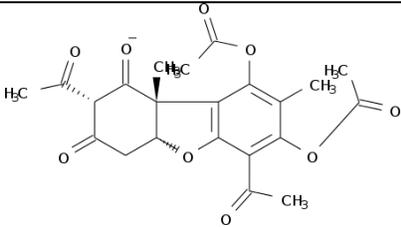
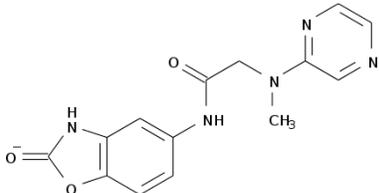
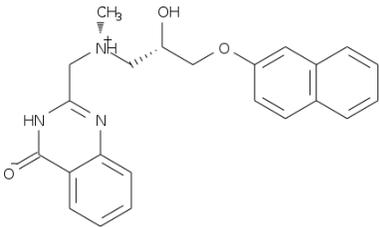
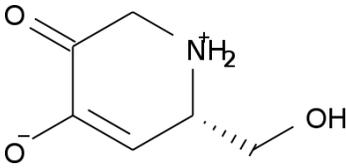
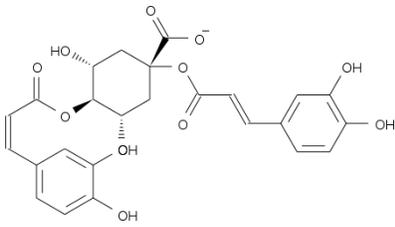
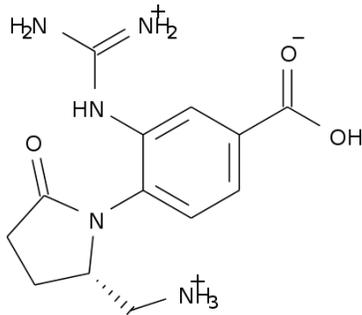
Rank	Ligand Name	Structure	Source	MM-GBSA
71	572		NCI Diversity Set II	-19.6 ± 0.1
72	ZINC30600361		ZINC DrugLike	-19.5 ± 0.2
73	ZINC33015084		ZINC DrugLike	-19.4 ± 0.4
74	ZINC09915898		ZINC Fragments	-18.9 ± 0.1
75	449		NCI Diversity Set II	-18.8 ± 0.3
76	chEMBL 198898		chEMBL	-18.1 ± 0.2

Table A-4. Continued.

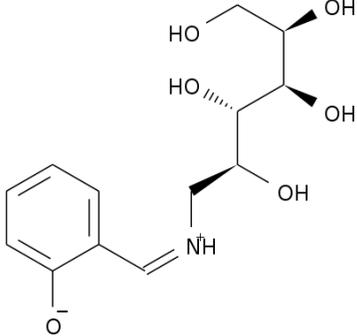
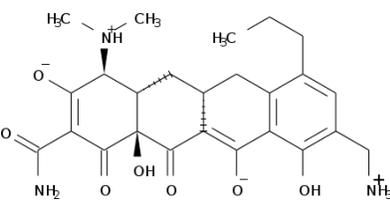
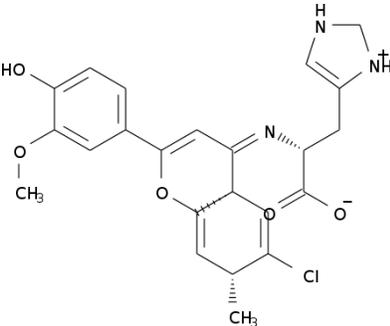
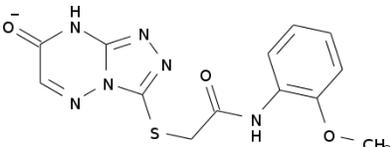
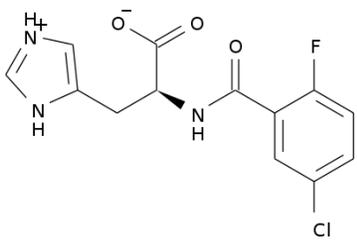
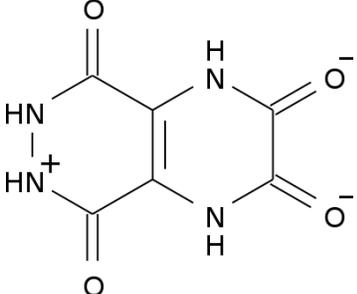
Rank	Ligand Name	Structure	Source	MM-GBSA
77	ZINC17001340		ZINC Leads	-17.9 ± 0.2
78	ZINC34312910		ZINC PubChem	-16.4 ± 0.4
79	ZINC17043337		ZINC DrugLike	-14.8 ± 0.1
80	ZINC05144331		ZINC DrugLike	-12.7 ± 0.3
81	ZINC20555579		ZINC DrugLike	-10.2 ± 0.1

Table A-4. Continued.

Rank	Ligand Name	Structure	Source	MM-GBSA
82	ZINC08655738		ZINC DrugLike	-8.8 ± 0.6

LIST OF REFERENCES

- (1) Chagas, C. *Memórias Inst. Oswaldo Cruz* **1909**, 1, 159–218.
- (2) Moncayo, Á.; Silveira, A. C. *Memórias Inst. Oswaldo Cruz* **2009**, 104, 17–30.
- (3) Organization, W. H. *Control of Chagas Disease: Second Report of the WHO Expert Committee [on the Control of Chagas Disease, Brasilia, 20-28 November 2000]*; World Health Organization, 2002.
- (4) Dodd, R. Y.; Leiby, D. A. *Annu. Rev. Med.* **2004**, 55, 191–207.
- (5) Gascon, J.; Bern, C.; Pinazo, M.-J. *Acta Trop.* **2010**, 115, 22–27.
- (6) Crompton, D. W. D. W. T.; Peters, P. *Working to Overcome the Global Impact of Neglected Tropical Diseases: First WHO Report on Neglected Tropical Diseases 2010*; World Health Organization, 2010.
- (7) Hotez, P. J.; Dumonteil, E.; Woc-Colburn, L.; Serpa, J. A.; Bezek, S.; Edwards, M. S.; Hallmark, C. J.; Musselwhite, L. W.; Flink, B. J.; Bottazzi, M. E. *PLoS Negl Trop Dis* **2012**, 6, e1498.
- (8) Leiby, D. A.; Herron, R. M.; Read, E. J.; Lenes, B. A.; Stumpf, R. J. *Transfusion (Paris)* **2002**, 42, 549–555.
- (9) Nowicki, M. J.; Chinchilla, C.; Corado, L.; Matsuoka, L.; Selby, R.; Steurer, F.; Mone, T.; Mendez, R.; Aswad, S. *Transplantation* **2006**, 81, 477–479.
- (10) Kun, H.; Moore, A.; Mascola, L.; Frank, S.; Gena, L.; Kubak, B.; Radhakrishna, S.; Leiby, D.; Herron, R.; Mone, T.; Hunter, R.; Kuehnert, M. *Clin. Infect. Dis.* **2009**, 48, 1534–1540.
- (11) Riera, C.; Guarro, A.; Kassab, H. E.; Jorba, J. M.; Castro, M.; Angrill, R.; Gállego, M.; Fisa, R.; Martin, C.; Lobato, A.; Portús, M. *Am. J. Trop. Med. Hyg.* **2006**, 75, 1078–1081.
- (12) Pereira, K. S.; Schmidt, F. L.; Guaraldo, A. M. A.; Franco, R. M. B.; Dias, V. L.; Passos, L. A. C. *J. Food Prot.* **2009**, 72, 441–446.
- (13) Pereira, K. S.; Schmidt, F. L.; Barbosa, R. L.; Guaraldo, A. M. A.; Franco, R. M. B.; Dias, V. L.; Passos, L. A. C. In *Advances in Food and Nutrition Research*; Steve L. Taylor, Ed.; Academic Press, 2010; Vol. Volume 59, pp. 63–85.
- (14) Chagas, C.; Chagas, C. *Memórias Inst. Oswaldo Cruz* **1916**, 8, 37–60.
- (15) Chagas, C.; Villela, E. *Inst Osw Cruz—14* **1922**, 5–61.
- (16) Dawes, B. *Advances in Parasitology Volume 6 APL*; Academic Press, 1968.

- (17) Barrett, M. P.; Burchmore, R. J.; Stich, A.; Lazzari, J. O.; Frasc, A. C.; Cazzulo, J. J.; Krishna, S. *The Lancet* **2003**, *362*, 1469–1480.
- (18) Zhang, L.; Tarleton, R. L. *J. Infect. Dis.* **1999**, *180*, 480–486.
- (19) A ez, N.; Carrasco, H.; Parada, H.; Crisante, G.; Rojas, A.; Fuenmayor, C.; Gonzalez, N.; Percoco, G.; Borges, R.; Guevara, P. *Am. J. Trop. Med. Hyg.* **1999**, *60*, 726–732.
- (20) Garcia, S.; Ramos, C. O.; Senra, J. F. V.; Vilas-Boas, F.; Rodrigues, M. M.; Campos-de-Carvalho, A. C.; Ribeiro-dos-Santos, R.; Soares, M. B. P. *Antimicrob. Agents Chemother.* **2005**, *49*, 1521–1528.
- (21) Jackson, Y.; Chatelain, E.; Mauris, A.; Holst, M.; Miao, Q.; Chappuis, F.; Ndao, M. *BMC Infect. Dis.* **2013**, *13*, 85.
- (22) Castro, J. A.; deMecca, M. M.; Bartel, L. C. *Hum Exp Toxicol* **2006**, *25*, 471–479.
- (23) Filardi, L. S.; Brener, Z. *Trans. R. Soc. Trop. Med. Hyg.* **1987**, *81*, 755–759.
- (24) Docampo, R.; Schmuñis, G. A. *Parasitol. Today* **1997**, *13*, 129–130.
- (25) Engel, J. C.; Doyle, P. S.; Hsieh, I.; McKerrow, J. H. *J. Exp. Med.* **1998**, *188*, 725–734.
- (26) McKerrow, J. H. *Int. J. Parasitol.* **1999**, *29*, 833–837.
- (27) Bond, C. S.; Zhang, Y.; Berriman, M.; Cunningham, M. L.; Fairlamb, A. H.; Hunter, W. N. *Structure* **1999**, *7*, 81–89.
- (28) Freymann, D. M.; Wenck, M. A.; Engel, J. C.; Feng, J.; Focia, P. J.; Eakin, A. E.; Craig III, S. P. *Chem. Biol.* **2000**, *7*, 957–968.
- (29) Lira, R.; Contreras, L. M.; Rita, R. M. S.; Urbina, J. A. *J. Antimicrob. Chemother.* **2001**, *47*, 537–546.
- (30) Martin, M. B.; Grimley, J. S.; Lewis, J. C.; Heath, H. T.; Bailey, B. N.; Kendrick, H.; Yardley, V.; Caldera, A.; Lira, R.; Urbina, J. A.; Moreno, S. N. J.; Docampo, R.; Croft, S. L.; Oldfield, E. *J. Med. Chem.* **2001**, *44*, 909–916.
- (31) Oli, M. M.; Einicker-Lamas, M. *An. Acad. Bras. Ciências* **2000**, *72*, 413–419.
- (32) Marini, V.; Moretti, E.; Bermejo, D.; Basso, B. *Memórias Inst. Oswaldo Cruz* **2011**, *106*, 32–37.
- (33) Ramsey, J. M.; Cruz-Celis, A.; Salgado, L.; Espinosa, L.; Ordoñez, R.; Lopez, R.; Schofield, C. J. *J. Med. Entomol.* **2003**, *40*, 912–920.
- (34) Friedman-Rudovsky, J. *Science* **2012**, *336*, 666–667.

- (35) Bern, C.; Montgomery, S. P.; Katz, L.; Caglioti, S.; Stramer, S. L. *Curr. Opin. Infect. Dis.* **2008**, *21*, 476–482.
- (36) Andrews, N. W.; Hong, K.; Robbins, E. S.; Nussenzweig, V. *Exp. Parasitol.* **1987**, *64*, 474–484.
- (37) Ley, V.; Andrews, N. W.; Robbins, E. S.; Nussenzweig, V. *J. Exp. Med.* **1988**, *168*, 649–659.
- (38) Bonaldo, M. C.; Souto-Padron, T.; Souza, W. de; Goldenberg, S. *J. Cell Biol.* **1988**, *106*, 1349–1358.
- (39) Kleffmann, T.; Schmidt, J.; Schaub, G. A. *J. Eukaryot. Microbiol.* **1998**, *45*, 548–555.
- (40) Andrews, N. W. *Biol. Res.* **1993**, *26*, 65–67.
- (41) Schauer, R. *Zoology* **2004**, *107*, 49–64.
- (42) Pilatte, Y.; Bignon, J.; Lambré, C. R. *Glycobiology* **1993**, *3*, 201–218.
- (43) Previato, J.; Andrade, A. F. B.; Pessolani, M. C. V.; Mendonça-Previato, L. *Mol. Biochem. Parasitol.* **1985**, *16*, 85–96.
- (44) Schenkman, S.; Eichinger, D.; Pereira, M. E. A.; Nussenzweig, V. *Annu. Rev. Microbiol.* **1994**, *48*, 499–523.
- (45) Todeschini, A. R.; Mendonca-Previato, L.; Previato, J. O.; Varki, A.; Halbeek, H. *van Glycobiology* **2000**, *10*, 213–221.
- (46) Buschiazzo, A.; Amaya, M. F.; Amaya, M. L.; Frasch, A. C.; Alzari, P. M. *Mol. Cell* **2002**, *10*, 757–768.
- (47) Buscaglia, C. A.; Alfonso, J.; Campetella, O.; Frasch, A. C. C. *Blood* **1999**, *93*, 2025–2032.
- (48) Crennell, S. J.; Garman, E. F.; Laver, W. G.; Vimr, E. R.; Taylor, G. L. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 9852–9856.
- (49) Luo, Y.; Li, S.-C.; Li, Y.-T.; Luo, M. *J. Mol. Biol.* **1999**, *285*, 323–332.
- (50) Taylor, G. *Curr. Opin. Struct. Biol.* **1996**, *6*, 830–837.
- (51) Horenstein, B. A.; Yang, J.; Bruner, M. *NUKLEONIKA* **2002**, *47*, S25–S28.
- (52) Watts, A. G.; Damager, I.; Amaya, M. L.; Buschiazzo, A.; Alzari, P.; Frasch, A. C.; Withers, S. G. *J Am Chem Soc* **2003**, *125*, 7532–7533.

- (53) Amaya, M. F.; Watts, A. G.; Damager, I.; Wehenkel, A.; Nguyen, T.; Buschiazzo, A.; Paris, G.; Frasch, A. C.; Withers, S. G.; Alzari, P. M. *Structure* **2004**, *12*, 775–784.
- (54) Pierdominici-Sottile, G.; Roitberg, A. E. *Biochem.* **2011**, *50*, 836–842.
- (55) Mitchell, F. L.; Miles, S. M.; Neres, J.; Bichenkova, E. V.; Bryce, R. A. *Biophys. J.* **2010**, *98*, L38–L40.
- (56) Pierdominici-Sottile, G.; Horenstein, N. A.; Roitberg, A. E. *Biochem.* **2011**, *50*, 10150–10158.
- (57) Pauling, L. *Chem. Eng. News Arch.* **1946**, *24*, 1375–1377.
- (58) Pauling, L. *Nature* **1948**, *161*, 707–709.
- (59) Demir, O.; Roitberg, A. E. *Biochem.* **2009**, *48*, 3398–3406.
- (60) Amaro, R. E.; Schnaufer, A.; Interthal, H.; Hol, W.; Stuart, K. D.; McCammon, J. A. *Proc. Natl. Acad. Sci.* **2008**, *105*, 17278–17283.
- (61) Damm, K. L.; Ung, P. M. U.; Quintero, J. J.; Gestwicki, J. E.; Carlson, H. A. *Biopolymers* **2008**, *89*, 643–652.
- (62) Meindl, P.; Bodo, G.; Palese, P.; Schulman, J.; Tuppy, H. *Virology* **1974**, *58*, 457–463.
- (63) Vandekerckhove, F.; Schenkman, S.; Carvalho, L. P. de; Tomlinson, S.; Kiso, M.; Yoshida, M.; Hasegawa, A.; Nussenzweig, V. *Glycobiology* **1992**, *2*, 541–548.
- (64) Paris, G.; Ratier, L.; Amaya, M. F.; Nguyen, T.; Alzari, P. M.; Frasch, A. C. C. *J. Mol. Biol.* **2005**, *345*, 923–934.
- (65) Streicher H. *Curr. Med. Chem. - Anti-Infect. Agents* **2004**, *3*, 149–161.
- (66) Harrison, J. A.; Kartha, K. P. R.; Turnbull, W. B.; Scheuerl, S. L.; Naismith, J. H.; Schenkman, S.; Field, R. A. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 141–144.
- (67) Buchini, S.; Buschiazzo, A.; Withers, S. G. *Angew. Chem.* **2008**, *120*, 2740–2743.
- (68) Carvalho, S. T.; Sola-Penna, M.; Oliveira, I. A.; Pita, S.; Goncalves, A. S.; Neves, B. C.; Sousa, F. R.; de-Lima, L. F.; Kurogochi, M.; Hinou, H.; Nishimura, S.-I.; Mendonca-Previato, L.; Previato, J. O.; Todeschini, A. R. *Glycobiology* **2010**, *20*, 1034–1045.
- (69) Agusti, R.; Paris, G.; Ratier, L.; Frasch, A. C. C.; de Lederkremer, R. M. *Glycobiology* **2004**, *14*, 659–670.

- (70) Giorgi, M. E.; Ratier, L.; Agusti, R.; Frasc, A. C. C.; Lederkremer, R. M. de *Glycoconj. J.* **2010**, *27*, 549–559.
- (71) Giorgi, M. E.; Piuselli, D.; Agusti, R.; de Lederkremer, R. M.; Giordano, S. *ARKIVOC* **2011**, *7*, 260–271.
- (72) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021.
- (73) Carvalho, I.; Andrade, P.; Campo, V. L.; Guedes, P. M. M.; Sesti-Costa, R.; Silva, J. S.; Schenkman, S.; Dedola, S.; Hill, L.; Rejzek, M.; Nepogodiev, S. A.; Field, R. A. *Bioorg. Med. Chem.* **2010**, *18*, 2412–2427.
- (74) Harrison, J. A.; Kartha, K. P. R.; Fournier, E. J. L.; Lowary, T. L.; Malet, C.; Nilsson, U. J.; Hindsgaul, O.; Schenkman, S.; Naismith, J. H.; Field, R. A. *Org. Biomol. Chem.* **2011**, *9*, 1653.
- (75) Meinke, S.; Schroven, A.; Thiem, J. *Org. Biomol. Chem.* **2011**, *9*, 4487–4497.
- (76) Neres, J.; Bonnet, P.; Edwards, P. N.; Kotian, P. L.; Buschiazzo, A.; Alzari, P. M.; Bryce, R. A.; Douglas, K. T. *Bioorg. Med. Chem.* **2007**, *15*, 2106–2119.
- (77) Neres, J.; Brewer, M. L.; Ratier, L.; Botti, H.; Buschiazzo, A.; Edwards, P. N.; Mortenson, P. N.; Charlton, M. H.; Alzari, P. M.; Frasc, A. C.; Bryce, R. A.; Douglas, K. T. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 589–596.
- (78) Kim, J. H.; Ryu, H. W.; Shim, J. H.; Park, K. H.; Withers, S. G. *ChemBioChem* **2009**, *10*, 2475–2479.
- (79) Lima, A. H.; Lameira, J.; Alves, C. N. *Struct. Chem.* **2012**, *23*, 147–152.
- (80) Arioka, S.; Sakagami, M.; Uematsu, R.; Yamaguchi, H.; Togame, H.; Takemoto, H.; Hinou, H.; Nishimura, S.-I. *Bioorg. Med. Chem.* **2010**, *18*, 1633–1640.
- (81) Fischer, E. *Berichte Dtsch. Chem. Ges.* **1894**, *27*, 2985–2993.
- (82) Levinthal, C.; Wodak, S. J.; Kahn, P.; Dadvanian, A. K. *Proc. Natl. Acad. Sci.* **1975**, *72*, 1330–1334.
- (83) Salemme, F. R. *J. Mol. Biol.* **1976**, *102*, 563–568.
- (84) Wodak, S. J.; Janin, J. *J. Mol. Biol.* **1978**, *124*, 323–342.
- (85) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.
- (86) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. A. *Bioorg. Med. Chem.* **1995**, *3*, 411–428.

- (87) Koshland, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98–104.
- (88) DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. *J. Med. Chem.* **1986**, *29*, 2149–2153.
- (89) Busetta, B.; Tickle, I. J.; Blundell, T. L. *J. Appl. Crystallogr.* **1983**, *16*, 432–437.
- (90) Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849–857.
- (91) Pattabiraman, N.; Levitt, M.; Ferrin, T. E.; Langridge, R. *J. Comput. Chem.* **1985**, *6*, 432–436.
- (92) Tomioka, N.; Itai, A.; Iitaka, Y. *J. Comput. Aided Mol. Des.* **1987**, *1*, 197–210.
- (93) Laurie, A. T. R.; Jackson, R. M. *Bioinformatics* **2005**, *21*, 1908–1916.
- (94) Neuvirth, H.; Raz, R.; Schreiber, G. *J. Mol. Biol.* **2004**, *338*, 181–199.
- (95) Liu, X. S.; Brutlag, D. L.; Liu, J. S. *Nat. Biotechnol.* **2002**, *20*, 835–839.
- (96) Singh, T.; Biswas, D.; Jayaram, B. *J. Chem. Inf. Model.* **2011**, *51*, 2515–2527.
- (97) Jiang, F.; Kim, S.-H. *J. Mol. Biol.* **1991**, *219*, 79–102.
- (98) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. *J. Med. Chem.* **2006**, *49*, 534–553.
- (99) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. *J. Med. Chem.* **2004**, *47*, 5076–5084.
- (100) Leach, A. *J. Mol. Biol.* **1994**, *235*, 345–356.
- (101) Anderson, A. C.; O'Neil, R. H.; Surti, T. S.; Stroud, R. M. *Chem. Biol.* **2001**, *8*, 445–457.
- (102) Yang, A. Y.-C.; Källblad, P.; Mancera, R. L. *J. Comput. Aided Mol. Des.* **2004**, *18*, 235–250.
- (103) Cavasotto, C. N.; Abagyan, R. A. *J. Mol. Biol.* **2004**, *337*, 209–225.
- (104) Totrov, M.; Abagyan, R. *Proteins* **1997**, *Suppl 1*, 215–220.
- (105) Meiler, J.; Baker, D. *Proteins Struct. Funct. Bioinforma.* **2006**, *65*, 538–548.
- (106) Schrödinger <http://www.schrodinger.com/> (accessed Jul 5, 2013).
- (107) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. *Med. Res. Rev.* **2006**, *26*, 531–568.

- (108) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. *J. Am. Chem. Soc.* **2002**, *124*, 5632–5633.
- (109) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. *Biopolymers* **2003**, *68*, 47–62.
- (110) Amaro, R. E.; Baron, R.; McCammon, J. A. *J. Comput. Aided Mol. Des.* **2008**, *22*, 693–705.
- (111) Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit. JMR* **1996**, *9*, 1–5.
- (112) Fernández-Recio, J.; Totrov, M.; Abagyan, R. *Proteins* **2003**, *52*, 113–117.
- (113) Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M. *J. Mol. Struct. THEOCHEM* **1994**, *308*, 191–206.
- (114) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J Comp Chem* **1998**, *19*, 1639–1662.
- (115) Taylor, J. S.; Burnett, R. M. *Proteins* **2000**, *41*, 173–191.
- (116) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727–748.
- (117) Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. *J. Comput. Aided Mol. Des.* **1995**, *9*, 113–130.
- (118) Hou, T.; Wang, J.; Chen, L.; Xu, X. *Protein Eng.* **1999**, *12*, 639–648.
- (119) Yang, J. M.; Kao, C. Y. *IEEE Trans. Inf. Technol. Biomed. Publ. IEEE Eng. Med. Biol. Soc.* **2000**, *4*, 225–237.
- (120) Feher, M.; Williams, C. I. *J. Chem. Inf. Model.* **2009**, *49*, 1704–1714.
- (121) Williams, C. I.; Feher, M. *J. Comput. Aided Mol. Des.* **2008**, *22*, 39–51.
- (122) Goodsell, D. S.; Olson, A. J. *Proteins Struct. Funct. Bioinforma.* **1990**, *8*, 195–202.
- (123) Metropolis, N.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341.
- (124) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (125) Trosset, J.-Y.; Scheraga, H. A. *Proc. Natl. Acad. Sci.* **1998**, *95*, 8011–8015.
- (126) Pak, Y.; Wang, S. *J. Phys. Chem. B* **2000**, *104*, 354–359.

- (127) Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. *Chem. Phys. Lett.* **1997**, *278*, 297–301.
- (128) Nakajima, N.; Nakamura, H.; Kidera, A. *J. Phys. Chem. B* **1997**, *101*, 817–824.
- (129) Di Nola, A.; Roccatano, D.; Berendsen, H. J. C. *Proteins Struct. Funct. Bioinforma.* **1994**, *19*, 174–182.
- (130) Mangoni, M.; Roccatano, D.; Di Nola, A. *Proteins Struct. Funct. Bioinforma.* **1999**, *35*, 153–162.
- (131) 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. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (132) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (133) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470–489.
- (134) Leach, A. R.; Kuntz, I. D. *J. Comput. Chem.* **1992**, *13*, 730–748.
- (135) Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput. Aided Mol. Des.* **2001**, *15*, 411–428.
- (136) Miranker, A.; Karplus, M. *Proteins Struct. Funct. Genet.* **1991**, *11*, 29–34.
- (137) Bohm, H.-J. *J. Comput. Aided Mol. Des.* **1992**, *6*, 593–606.
- (138) Durrant, J. D.; Amaro, R. E.; McCammon, J. A. *Chem. Biol. Drug Des.* **2009**, *73*, 168–178.
- (139) Wang, R.; Gao, Y.; Lai, L. *Mol. Model. Annu.* **2000**, *6*, 498–516.
- (140) Yuan, Y.; Pei, J.; Lai, L. *J. Chem. Inf. Model.* **2011**, *51*, 1083–1091.
- (141) Judson, R. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; John Wiley & Sons, Inc., 2007; pp. 1–73.
- (142) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (143) Bash, P. A.; Field, M. J.; Karplus, M. *J. Am. Chem. Soc.* **1987**, *109*, 8092–8094.
- (144) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat Rev Drug Discov* **2004**, *3*, 935–949.
- (145) Jones, J. E. *Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character* **1924**, *106*, 441–462.

- (146) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins Struct. Funct. Bioinforma.* **2003**, *52*, 609–623.
- (147) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425–445.
- (148) Muegge, I.; Martin, Y. C. *J. Med. Chem.* **1999**, *42*, 791–804.
- (149) Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000**, *295*, 337–356.
- (150) DeWitte, R. S.; Shakhnovich, E. I. *J. Am. Chem. Soc.* **1996**, *118*, 11733–11744.
- (151) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (152) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750–1759.
- (153) Plewczynski, D.; Łażniewski, M.; Augustyniak, R.; Ginalski, K. *J. Comput. Chem.* **2011**, *32*, 742–755.
- (154) Kier, L. B. *Mol. Pharmacol.* **1967**, *3*, 487–494.
- (155) Kier, L. B. *Molecular orbital theory in drug research*; Academic Press, 1971.
- (156) Güner, O. F. *Pharmacophore: Perception, Development, and Use in Drug Design*; Internat'l University Line, 2000.
- (157) Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563–571.
- (158) Chen, J.; Lai, L. *J. Chem. Inf. Model.* **2006**, *46*, 2684–2691.
- (159) Dixon, S. L.; Smondirev, A. M.; Rao, S. N. *Chem. Biol. Drug Des.* **2006**, *67*, 370–372.
- (160) Dixon, S. L.; Smondirev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. *J. Comput. Aided Mol. Des.* **2006**, *20*, 647–671.
- (161) Yang, S.-Y. *Drug Discov. Today* **2010**, *15*, 444–450.
- (162) Loving, K.; Salam, N. K.; Sherman, W. *J. Comput. Aided Mol. Des.* **2009**, *23*, 541–554.
- (163) Salam, N. K.; Nuti, R.; Sherman, W. *J. Chem. Inf. Model.* **2009**, *49*, 2356–2368.
- (164) Wang, H.; Duffy, R. A.; Boykow, G. C.; Chackalamannil, S.; Madison, V. S. *J. Med. Chem.* **2008**, *51*, 2439–2446.

- (165) Schuster, D.; Nashev, L. G.; Kirchmair, J.; Laggner, C.; Wolber, G.; Langer, T.; Odermatt, A. *J. Med. Chem.* **2008**, *51*, 4188–4199.
- (166) Neves, M. A. C.; Dinis, T. C. P.; Colombo, G.; Sá e Melo, M. L. *J. Med. Chem.* **2009**, *52*, 143–150.
- (167) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957**, *27*, 1208–1209.
- (168) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459–466.
- (169) Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.
- (170) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; CRC Press, 2010.
- (171) Leach, A. R. *Molecular Modelling: Principles and Applications*; Pearson Prentice Hall, 2001.
- (172) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. . *J. Comput. Phys.* **1977**, *23*, 327–341.
- (173) Andrea, T. A.; Swope, W. C.; Andersen, H. C. *J. Chem. Phys.* **1983**, *79*, 4576–4584.
- (174) Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. *J. Chem. Theory Comput.* **2009**, *5*, 1624–1631.
- (175) Homeyer, N.; Gohlke, H. *Mol. Informatics* **2012**, *31*, 114–122.
- (176) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. *J. Phys. Chem. B* **2010**, *114*, 8505–8516.
- (177) Price, D. J.; Brooks, C. L. *J. Chem. Phys.* **2004**, *121*, 10096–10103.
- (178) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (179) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (180) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–725.
- (181) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817–3829.
- (182) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (183) Foloppe, N.; Mackerell, Jr., A. D. *J. Comput. Chem.* **2000**, *21*, 86–104.

- (184) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- (185) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (186) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. *J. Med. Chem.* **2006**, *49*, 4805–4808.
- (187) Rastelli, G.; Rio, A. D.; Degliesposti, G.; Sgobba, M. *J. Comput. Chem.* **2010**, *31*, 797–810.
- (188) Hou, T.; Wang, J.; Li, Y.; Wang, W. *J. Chem. Inf. Model.* **2011**, *51*, 69–82.
- (189) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.
- (190) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. *Curr. Comput. - Aided Drug Des.* **2011**, *7*, 146–157.
- (191) Chen, Y.; Shoichet, B. K. *Nat. Chem. Biol.* **2009**, *5*, 358–364.
- (192) Babaoglu, K.; Shoichet, B. K. *Nat. Chem. Biol.* **2006**, *2*, 720–723.
- (193) Marcou, G.; Rognan, D. *J. Chem. Inf. Model.* **2007**, *47*, 195–207.
- (194) Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. *Drug Discov. Today* **2008**, *13*, 23–29.
- (195) Guner, O. F. *Curr. Top. Med. Chem.* **2002**, *2*, 1321–1332.
- (196) Massova, I.; Kollman, P. *Perspect. Drug Discov. Des.* **2000**, *18*, 113–135.
- (197) Okimoto, N.; Futatsugi, N.; Fujii, H.; Suenaga, A.; Morimoto, G.; Yanai, R.; Ohno, Y.; Narumi, T.; Taiji, M. *Biophys. J.* **2010**, *98*, 460a.
- (198) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768.
- (199) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (200) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. *J. Comput. Aided Mol. Des.* 1–14.
- (201) *LigPrep v2.5*; Schrodinger, Inc.: Portland, OR.
- (202) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory Comput.* **2012**, *8*, 1542–1555.

- (203) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A. E.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; AMBER; University of California, San Francisco, 2012.
- (204) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.
- (205) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.
- (206) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591–607.
- (207) Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y.-P. *J. Med. Chem.* **2000**, *43*, 401–408.
- (208) Grüneberg, S.; Stubbs, M. T.; Klebe, G. *J. Med. Chem.* **2002**, *45*, 3588–3602.
- (209) Desai, P. V.; Patny, A.; Sabnis, Y.; Tekwani, B.; Gut, J.; Rosenthal, P.; Srivastava, A.; Avery, M. *J. Med. Chem.* **2004**, *47*, 6609–6615.
- (210) Brunskole Švegelj, M.; Turk, S.; Brus, B.; Lanišnik Rižner, T.; Stojan, J.; Gobec, S. *J. Chem. Inf. Model.* **2011**, *51*, 1716–1724.
- (211) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *J. Cheminformatics* **2012**, *4*, 17.
- (212) Trott, O.; Olson, A. J. *J Comp Chem* **2010**, *31*, 455–461.
- (213) Shie, J.-J.; Fang, J.-M.; Lai, P.-T.; Wen, W.-H.; Wang, S.-Y.; Cheng, Y.-S. E.; Tsai, K.-C.; Yang, A.-S.; Wong, C.-H. *J. Am. Chem. Soc.* **2011**, *133*, 17959–17965.
- (214) Neres, J.; Buschiazzo, A.; Alzari, P. M.; Walsh, L.; Douglas, K. T. *Anal. Biochem.* **2006**, *357*, 302–304.
- (215) De Ruiter, A.; Oostenbrink, C. *Curr. Opin. Chem. Biol.* **2011**, *15*, 547–552.
- (216) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (217) Meng, Y.; Sabri Dashti, D.; Roitberg, A. E. *J. Chem. Theory Comput.* **2011**, *7*, 2721–2727.
- (218) McQuarrie, D. A.; McQuarrie, D. A.; McQuarrie, D. A.; McQuarrie, D. A. *Statistical thermodynamics*; Harper and Row New York, 1973.
- (219) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891–914.

- (220) Genheden, S.; Ryde, U. *J. Comput. Chem.* **2010**, *31*, 837–846.
- (221) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.
- (222) Steinbrecher, T.; Labahn, A. *Curr. Med. Chem.* **2010**, *17*, 767–785.
- (223) Vargiu, A. V.; Magistrato, A. *Inorg Chem* **2012**, *51*, 2046–2057.
- (224) Kar, P.; Knecht, V. *J Phys Chem B* **2012**, *116*, 2605–2614.
- (225) Wang, J.; Hou, T.; Xu, X. *Curr. Comput. Aided Drug Des.* **2006**, *2*, 287–306.
- (226) Metz, A.; Pflieger, C.; Kopitz, H.; Pfeiffer-Marek, S.; Baringhaus, K.-H.; Gohlke, H. *J Chem Inf Model* **2011**, *52*, 120–133.
- (227) Massova, I.; Kollman, P. A. *J Am Chem Soc* **1999**, *121*, 8133–8143.
- (228) McQuarrie, D. A. *Statistical Mechanics*; University Science Books, 2000.
- (229) Brooks, B. R.; Janežič, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522–1542.
- (230) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.
- (231) Macke Thomas J.; Case David A. In *Molecular Modeling of Nucleic Acids*; ACS Symposium Series; American Chemical Society, 1997; Vol. 682, pp. 379–393.
- (232) Tan, C.; Tan, Y.-H.; Luo, R. *J. Phys. Chem. B* **2007**, *111*, 12263–12274.
- (233) Tan, C.; Yang, L.; Luo, R. *J. Phys. Chem. B* **2006**, *110*, 18680–18687.
- (234) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244–1253.
- (235) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci.* **2001**, *98*, 10037–10041.
- (236) Konecny, R.; Baker, N. A.; McCammon, J. A. *Comput. Sci. Discov.* **2012**, *5*, 015005.
- (237) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.
- (238) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
- (239) Nguyen, H.; Roe, D. R.; Simmerling, C. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034.

- (240) Gleeson, M. P.; Gleeson, D. *J. Chem. Inf. Model.* **2009**, *49*, 1437–1448.
- (241) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019–1031.
- (242) Dewar, M. J. S.; Zebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (243) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.
- (244) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220.
- (245) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*; Academic Press, 2006.

BIOGRAPHICAL SKETCH

Bill Miller III was born in Fort Scott, Kansas in 1985 to Bill and Robin Miller, Jr. He has one sibling, his sister Crystal. He grew up in Nevada, a small town in southwest Missouri. He attended Nevada High School where Mr. Brian Norton first introduced him to chemistry during his junior year. He graduated valedictorian from Nevada High in 2004, and proceeded to Truman State University, a small liberal arts university in Kirksville, Missouri. At Truman State, he became active in computational chemistry research working under the advisement of Dr. Maria Nagan. He graduated summa cum laude from Truman State in 2008, and then moved to Gainesville, Florida to work with Dr. Adrian Roitberg at the University of Florida. He received his Ph.D. from the University of Florida in the summer of 2013.

In June 2007, he married his loving wife, Ryan. They were blessed with the arrival of their daughter, Rylee Ann Miller, in June 2010. And they recently had their second child, Billy Ray Miller IV, in December 2012.