Title
Computational Methods for Studying Molecular Recognition, and Applications for Drug Discovery and Improved Treatment Options

Permalink
https://escholarship.org/uc/item/1t3494h5

Author
Roger, Kathleen Elizabeth

Publication Date
2016-03-23

Peer reviewed|Thesis/dissertation
Computational Methods for Studying Molecular Recognition, and Applications for Drug Discovery and Improved Treatment Options

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy

in

Biomedical Sciences

by

Kathleen Elizabeth Rogers

Committee In Charge:

Professor J. Andrew McCammon, Chair
Professor Philip E. Bourne
Professor Michael K. Gilson
Professor Alexandra C. Newton
Professor Renate Pilz

2013
Copyright
Kathleen Elizabeth Rogers, 2013
All Rights Reserved
The dissertation of Kathleen Elizabeth Rogers is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

[Signatures]

________________________________________
________________________________________
________________________________________
________________________________________

Chair

University of California, San Diego

2013
Dedication

I wish to dedicate this dissertation to my family: past, present and future. In loving memory of my amazing grandparents, I acknowledge their support and strong belief in education. As the youngest of four, I also must thank my brothers and sister for their encouragement and inspiration over the years. To my parents, without whom I would not be who I am or where I am today, I gratefully acknowledge all of their hard work, strength, sacrifice, wisdom, selflessness, and praise. Lastly, it gives me great pleasure to acknowledge my thoughtful and brilliant fiancé for all the support, encouragement, and helpful discussions, without which completing this work would have been much more difficult and unpleasant.

I owe my deepest gratitude to Professor J. Andrew McCammon for providing such a great research environment and allowing me to complete my graduate work under his wing. It was an honor to work with him and the many incredibly intelligent people in his laboratory.

I would also like to dedicate this in memory of my dear friend, Anne Smedinghoff. Her strength, courage, compassion, and kind heart continue to inspire me to be as great as the legacy she leaves behind.
I believe in intuition and inspiration. Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution. It is, strictly speaking, a real factor in scientific research.

*Albert Einstein*
Table of Contents

Signature Page ........................................................................................................... iii
Dedication ................................................................................................................ iv
Epigraph ................................................................................................................... v
Table of Contents .................................................................................................... vi
List of Figures .......................................................................................................... ix
List of Tables ........................................................................................................... xii
Acknowledgements ................................................................................................. xiii
Vita ............................................................................................................................. xvi
Abstract of the Dissertation ..................................................................................... xvii
Chapter 1: Current computational approaches towards discovery and development of improved, targeted chemotherapy ......................................................... 1
   Abstract ................................................................................................................. 1
   Abbreviations ....................................................................................................... 2
   Problems associated with cytotoxic chemotherapeutics .................................... 2
   The promise of selective treatment against disease-driving targets .................. 4
   Specific drug delivery systems for localized release ........................................... 5
   Structure-based, computer-aided modeling and discovery of cancer or viral and parasitic-infection drug leads ................................................................. 6
   Accurate computational free energy calculations .................................................. 7
   Concluding remarks ............................................................................................... 8
Chapter 2: On the Role of Dewetting Transitions in Host-Guest Binding Free Energy Calculations ........................................................................................................ 10
   Abstract ................................................................................................................. 10
   Abbreviations ....................................................................................................... 11
   Introduction .......................................................................................................... 11
   Computational Details ........................................................................................... 14
      Molecular Dynamics Simulations ..................................................................... 15
   Thermodynamic Integration ............................................................................... 15
      Statistical Analysis of Uncertainties ................................................................. 16
   Results and Discussion ......................................................................................... 19
   Conclusions .......................................................................................................... 32
   Acknowledgments ................................................................................................. 33
TEAD Computational Model System Preparation .......................................................... 72
Molecular Dynamics (MD) Simulations of the TEAD YBD ........................................ 72
TEAD Conformational Clustering and Docking ............................................................. 73
Cell Culture ................................................................................................................. 75
Luciferase Reporter Assay ........................................................................................... 76
YAP/TEAD co-transfection, immunoprecipitation and immunoblotting ................. 76
Acknowledgments ....................................................................................................... 76
Supporting Information ............................................................................................... 77
Chapter 5 : Conclusions ............................................................................................. 80
References ................................................................................................................... 87
List of Figures

Figure 1.1. A long but incomplete list of chemotherapy side effects and sequelae of varying severity and duration. ................................................................. 3

Figure 1.2. Thermodynamic Integration (TI) cycle using “computational alchemy” to estimate absolute binding free energy. ......................................................... 8

Figure 2.1. Host and guest molecules studied. (a) Chemical representation of glycoluril repeat unit constituting the host. (b) Side view of the [n=7] non-polar host cucurbit[7]uril (CB[7]) structure. (c) B2 non-polar guest molecule. (d) Top view of CB[7]-B2 host-guest ........................................ 13

Figure 2.2. Thermodynamic Integration along the λ coupling parameter. vDW-decoupling TI curves of (a) guest and (b) host-guest system simulations in TIP3P (red) and TIP4P (blue). <∂V(λ)/∂λ>2 values and corresponding uncertainties, σ_sim(t), are shown at each λ sampled. ............................................. 18

Figure 2.3. Convergence monitoring of <∂V(λ)/∂λ> and the potential energy derivative backward average, across time with different criteria, τ_sim, during B2-CB[7] vDW decoupling simulations where the highest ∂V/∂λ fluctuations occur using either TIP3P (left) or TIP4P (right) water models................................... 18

Figure 2.4. Average hydration surfaces of 50 ns vDW host-guest simulations for unbound (λ=1.0; top) and bound (λ=0.0; bottom) thermodynamic states. Corresponding maps are also shown for the λ=0.85 intermediate unphysical state (middle). ......................................................... 25

Figure 2.5. Average hydration surfaces (as in Figure 2.4) for host-guest vDW simulation bound (λ=0.0) thermodynamic state with TIP3P (red wireframe, left side) and TIP4P (blue wireframe, right side) water oxygen atom densities shown in and around the host structure (licorice representation). .................. 28

Figure 2.6. Fluctuations of (a,c) <∂V/∂λ> values and (b,d) number of water molecules in the CB[7] host system over time for the TIP3P (red, a-b) and TIP4P (blue, c-d) water models during 50 ns host-guest vDW simulations. ...... 29

Figure 2.7. Correlation between <∂V/∂λ> values and number of waters within host for λ=0.85 vDW simulation with TIP3P and TIP4P water models. Vertical error bars are the standard deviation values. ................................................................. 31

Figure SI-2.1. vs. λ TI curves of partial charge (ele) decoupling steps for both systems and water models. Uncertainties are reported as vertical error bars. ........................................................................................................ 35

Figure SI-2.2. Correlation between the and the number of waters within the host for vDW TI simulations at endpoint λ values and those with slow convergence. Vertical error bars are the standard deviation values. .................................................. 36

Figure 3.1. The structures of all experimentally validated compounds. Associated IC50 values are given for the top three non-aggregating inhibitors. Means and standard deviations were calculated by considering all IC50 values associated with each compound, measured under varying experimental conditions........ 49
Figure 3.2. The Predicted Binding Poses and Receptor-Ligand Interactions of the Experimentally Validated Inhibitors. a, c, and e: Compounds and cruzain active sites are shown in stick and surface representation, respectively. Active site domains are also labeled (S1-S3, S1', S3').

Figure SI-3.1. IC50 curves for selected NSC compounds under various experimental conditions, used to investigate the mechanisms of inhibition and determine aggregation-based activity.

Figure 4.1. New pocket formed in TEAD model near edge of YAP segment. (A) TEAD solved structure is missing several residues, including 230-235, which are near edge of important interface where a YAP segment binds. (B) Filling in loop here for the TEAD model helps enclose a cavity, left of Y406.

Figure 4.2 Changes in TEAD YBD surface during MD suggest flexibility that may allow for and adapt to small molecule docking. Conformations adopted by the TEAD YBD surface near Y406 during the simulation show a variety of more promising forms.

Figure 4.3. Reproduced pose of pseudo-ligand YAP segment bound to TEAD. A small segment of YAP (pink sticks), residues L91-F95, docked similarly compared to the crystal structure of the YAP TEAD binding domain (yellow sticks). TEAD model shown as green surface except for yellow Y406.

Figure 4.4. Inhibition of YAP-TEAD4 luciferase activity by three NCI Diversity Set II compounds indicate effect on YAP activity. A dose-dependent decrease in YAP reporter activity can be seen by three of the first ten compounds tested. Luciferase activity normalized to Renilla.

Figure 4.5. NSC 105827 decreases YAP-TEAD4 interaction both by treatment of cells directly or in the post-lysate lysate. NSC 81750 and NSC 134159 do not appear to inhibit YAP binding in Co-IP experiments but NSC 105827 looks to decrease YAP in a dose-dependent manner.

Figure 4.6. Effect of NCI compounds on protein level of YAP and phosphorylated YAP in various cells lines. NSC 81750 and NSC 134159 both appear to slightly inhibit pYAP in at least one cell type but NSC 105827 appears to be the most effective overall. α-Tubulin shown as loading control.

Figure 4.7. Induced Fit Docking Pose of NSC105827 with an MD Ensemble Member. NSC 105827 screened with an ensemble member (that performed well overall in the Vina screen) using IFD shown here lays right on top of Y406.

Figure SI-4.1. Root-mean-square fluctuations (RMSF) of α-C atoms during TEAD Trajectory. Some of the highest fluctuations of α-C atoms are seen near residues 230-235. This region became an interesting focus in the dynamics and virtual screening studies due to pocket-forming surface flexibility there.

Figure SI-4.2. Dynamic Residues Near Modeled Pocket. (A) Residues 230-235 (purple), many of which also have high RMSF, are near the end of the YAP (yellow) PDB segment at interface 3 and form part of the pocket filled in for the TEAD model (green). (B) Rotated view of (A).

Figure SI-4.3. NCI Compounds Appear to Dock in Various Ways to TEAD. Many of the top hits from the Autodock Vina (A-B) and Glide (C-D) RCS screens of the NCI Diversity Set II library to the TEAD model (C) and ensemble.
members (A, B, D) are noticeably unique. TEAD Y406 shown as yellow surface.  

Figure SI-4.4. Additional NCI Compounds Show Preliminary Inhibitory Effects on TEAD-YAP Association. Compounds labeled here as A and H appear to inhibit TEAD-YAP binding in both TEAD and YAP pull-down experiments. Guan lab continues to further validate these and other potential hits.  

Figure 5.1. User-generated report and map on recent U.S. HIV statistics, available from the CDC’s National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP) Atlas.  

Figure 5.2. Absolute Free Energy of Host-Guest Binding. Pictured are the host and host-guest systems with the average water density surfaces (TIP3P and TIP4P water models) shown for both states.  

Figure 5.3. Induced Fit Docking pose of Cruzain with top lead, NSC 227186, which had an IC$_{50}$ value of 16.6 µM in enzymatic inhibition assays described in Chapter 3.  

Figure 5.4. A model of the Hippo pathway and spatial/temporal regulation of YAP activity via phosphorylation by Lats.
List of Tables

Table 1.1. Targeted drugs designed to treat specific types of cancer ........................................4
Table 2.1. Absolute Host-Guest Binding Free Energy Estimates ..............................................21
Table R2.2. Hydration During the Endpoint and Unphysical States Employed for Thermodynamic Integration .................................................................................................31
Table SI-2.1. Statistical analysis of free energy uncertainties, σ, for the host-guest van der Waals decoupling TI component (ΔG_{vdW}) using different convergence monitoring criterions ........................................................................................................34
Table SI-3.1. Preliminary experimental validation results for compounds with IC50 values of 100 µM or lower ...................................................................................................................55
Acknowledgements

I must first thank Professor J. Andrew McCammon for the wonderful opportunity to complete my graduate research in such a supportive and collaborative laboratory. The encouragement, guidance, and mentoring I received from Andy were essential to my success. I would also like to thank Professor Riccardo Baron, Dr. César de Oliveira, Dr. Jacob Durrant, Dr. Mikolai Fajer, Henrik Keränen, and all the McCammon lab members whom have ever helped with any of my projects. I must thank my collaborators Professor Michelle Arkin, Allison Doak, Professor Kun-Liang Guan, Dr. Ian Lian, and Joseline Ratnam, as well as Palmer Taylor and Ákos Nimenecz for the use of and guidance with their FlexStation. Last but not least, I would like to thank my thesis committee for their continued guidance and support.

This work was carried out with funding from the National Institutes of Health grants GM31749 (to J.A.M.) and NIH 5T32GM007752-32 (to K.E.R) and the National Science Foundation grants MCB-1020765 and MCA93S013 (to J.A.M.). Additional support from the ARCS foundation, the Howard Hughes Medical Institute, the NSF Supercomputer Centers, the San Diego Supercomputer Center, the W. M. Keck Foundation, the National Biomedical Computational Resource, and the Center for Theoretical Biological Physics is gratefully acknowledged. Compounds tested in this work were obtained free of charge from the National Cancer Institute’s Development Therapeutics Program.
Chapter 2 is a minimally modified reprint of the material as it appears in Rogers, Kathleen E.; Ortiz-Sanchez, Juan M.; Baron, Riccardo; Fajer, Mikolai; de Oliveira, César A.F., McCammon, J. Andrew. “On the Role of Dewetting Transitions in Host–Guest Binding Free Energy Calculations.” *Journal of Chemical Theory and Computation* 2013, 9 (1), 46-53. I was the first author and principal investigator on this paper. Dr. Ortiz-Sanchez provided useful discussions, Professor Baron was instrumental in much of the design of this work, Dr. Fajer made essential modifications to the AMBER code used, Dr. Oliveira completed all system parameterizations, and Professor McCammon suggested and further guided the project.

Chapter 3 is a minimally modified reprint of the material as it appears in Rogers, Kathleen E.; Keränen, Henrik; Durrant, Jacob D.; Ratnam, Joseline; Doak, Allison; Arkin, Michelle R.; McCammon, J. Andrew. “Novel cruzain inhibitors for the treatment of Chagas' disease.” *Chemical Biology & Drug Design*. 2012, Sep;80(3):398-405. I was the first author and principal investigator on this paper. Mr. Keränen and Dr. Durrant began the Relaxed Complex Scheme work with molecular dynamics and initial virtual screens. Dr. Ratnam, Ms. Doak, and Professor Arkin provided essential guidance for the enzymatic inhibition assays. Professor McCammon supported this work and advised during the project.
Chapter 4 contains text from the article in preparation Rogers, Kathleen E.; Lian, Ian; Guan, Kun-Liang; McCammon, J. Andrew. “Computational Drug Discovery of TEAD-YAP inhibitors” (In preparation for Chemical Biology & Drug Design). I am the first author and principal investigator on this paper. Dr. Lian provided the experimental results while Professor Guan and Professor McCammon supported the collaborative work.
Vita

2009  B. S. Chemical & Biomolecular Engineering, Johns Hopkins University

2013  Ph.D. Biomedical Sciences University of California, San Diego

Publications


ABSTRACT OF THE DISSERTATION

Computational Methods for Studying Molecular Recognition, and Applications for Drug Discovery and Improved Treatment Options

by

Kathleen Elizabeth Rogers

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2013

Professor J. Andrew McCammon, Chair

Molecular dynamics and computational drug discovery have both provided invaluable insight into numerous important biomolecular systems and even launched the design of a number of new therapeutics; however, significant gaps in our knowledge and insufficiencies in the treatment options available to patients still
remain. The latter is especially challenging for individuals suffering from diseases, such as cancer or Chagas disease, where there often is no alternative to highly toxic and unacceptable chemotherapeutics. The protozoan parasite *Trypanosoma cruzi*, the etiological agent of Chagas disease, affects millions of individuals and continues to be an important global health concern, yet available drugs are so poorly efficacious and produce such severe side effects that the FDA does not approve them. Chemotherapy has broad indications, including various types of cancers and infections, but the side effects are generally only acceptable in dire situations when the benefits outweigh the risks to the patient. Inherent to their mechanism, many of these adverse consequences are a result of broad, off target effects. An exciting approach to overcome this is through the use of drug delivery systems to improve selectivity of drugs such as chemotherapy. One such system is a very promising synthetic host, Cucurbit[7]uril (CB[7]), which has already shown great potential as a drug carrier for chemotherapeutics by improving stability and selectivity, thus decreasing negative side effects of these drugs.

Thermodynamic Integration (TI) with explicit solvent molecular dynamics (MD) simulations of a CB[7] host-guest system to estimate absolute binding free energy, presented in this work, has enhanced our understanding of protein-ligand binding and the effects of solvation. Another approach toward development of improved treatment options is to identify more specific targets, such as *T. cruzi* cruzain for Chagas disease and the oncogenic YAP-TEAD complex. The former is the major cysteine protease essential to parasite survival and the latter is an intriguing new focus with potential to improve cancer treatment because it is implicated in several forms of
cancer and metastasis. This work also describes the use of Computational Drug Discovery (CADD) methods to uncover novel chemical scaffolds against these two encouraging disease targets, *T. cruzi* cruzain and the human TEAD-YAP complex.
Chapter 1: Current computational approaches towards discovery and development of improved, targeted chemotherapy.

Abstract

There is a clear need for improved chemotherapy and treatment options but Next Generation Sequencing (NGS) is poised to help revolutionize how we treat cancer and other diseases with a genetic cause. Genomic-medicine has the potential not only to improve treatment efficacy but also limit unnecessary side effects, especially when a patient might be genetically predisposed to more severe toxicities. As more and more data is available on human oncogenes and other disease-causing mutations, their mechanism, and key regulators up- and down-stream of their products, we will become better informed to develop novel therapeutics against these new targets. In this way, we can approach more selective treatment to inhibit these disease-driving factors with higher efficiency and fewer adverse effects. Computational approaches for discovery of lead compounds and modeling of promising drug delivery systems help us explore our therapeutic options and better understand molecular recognition in protein-ligand binding.
Abbreviations

Next Generation Sequencing (NGS); multidrug resistance (MDR); high throughput screening (HTS); protein-protein inhibitors (PPI); absorption, distribution, metabolism, and excretion (ADME); free energy perturbation (FEP); thermodynamic integration (TI); molecular dynamics (MD)

Problems associated with cytotoxic chemotherapeutics

Although there are some potent and effective drugs available for many common cancers and infectious diseases, under certain conditions (e.g. pregnancy (1), impaired liver (2) or kidney function (3, 4), other infections such as hepatitis B and C viruses (5, 6)) they are unacceptable or may increase risks due to severe side effects and possible complications. Traditional cytotoxics exert their effects via several mechanisms including modifying DNA or interfering with microtubule polymerization, metabolite synthesis, or chromosome topology. In addition to targeting cancer cells, these cytotoxic agents affect normal cells including mucosal cells in the mouth and the lining of the intestines. This may lead to temporary yet painful sores, changes in taste, inability to swallow, opportunistic infections, inflammation and ulcers in the mouth as well as nausea, loss of appetite, vomiting, and diarrhea (7, 8). Additionally, chemotherapy targets bone marrow cells, causing myelosuppression, which lowers patient blood cell counts and leads to another collection of adverse events including increased risk of infection (9-11). Some cytotoxic treatments may also affect organs including the heart, lungs, bladder, kidneys, and the nervous system (12), on top of a variety of other damaging consequences on hair, skin, and nails (13, 14).
Several “neocytotoxics” (15) have been developed that target protein processing, chromatin or histone modification, and cell division in new ways. These drugs, although more selective towards a particular protein, are still directed toward essential targets that may be critical in many if not all cells. Thus, we may run into similar problems to those associated with traditional cytotoxics. Unfortunately, chemotherapy is not always highly effective for every patient and, even if it is, patients often relapse, where new tumor (16), parasite (17, 18), or viral (19) growth contains a heterogeneous mix of resistant cells or viruses generating further genetic modifications (20, 21). The risks vs. benefits of chemotherapy for each patient must be considered before proceeding with any treatment due to the inevitable side effects (Figure 1.1) and sometimes limited efficacy.

Figure 1.1. A long but incomplete list of chemotherapy side effects and sequelae of varying severity and duration.
The promise of selective treatment against disease-driving targets

In many cases now, a patient’s treatment may be best prescribed after first determining the genetic profile of their disease or infection. Several current or developing drugs exist that target specific cell signaling, disease-related components (See Table 1.1 for examples in oncology) such as kinases that help drive tumorigenesis (22-26) or parasitic infection (27) and viral enzymes essential for replication (28). Some of these targets, especially those of an invading parasite, are nonessential in adult cells but ideally small molecules should be designed to selectively inhibit only the one mutated or parasitic target and not any others that may be more crucial to adult cells. This can be difficult due to sequence and structural similarities, especially at highly conserved catalytic domains.

Table 1.1. Targeted drugs designed to treat specific types of cancer

<table>
<thead>
<tr>
<th>Drug Target</th>
<th>Drug(s)</th>
<th>Cancer Type(s) Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aromatase</td>
<td>Exemestane, Letrozole (29)</td>
<td>Breast</td>
</tr>
<tr>
<td>EGFR</td>
<td>Cetuximab(30)</td>
<td>Colorectal, head &amp; neck</td>
</tr>
<tr>
<td></td>
<td>Erlotinib(31)</td>
<td>Lung, Pancreatic</td>
</tr>
<tr>
<td></td>
<td>Gefitinib(32)</td>
<td>Lung</td>
</tr>
<tr>
<td></td>
<td>Panitumumab(33)</td>
<td>Colorectal</td>
</tr>
<tr>
<td>BCR-ABL, KIT, &amp; PDGFRα tyrosine kinases</td>
<td>Imatinib(34)</td>
<td>Chronic Myeloid Leukaemia, gastrointestinal</td>
</tr>
<tr>
<td>HER2 (ErbB)</td>
<td>Lapatinib(35)</td>
<td>Breast</td>
</tr>
<tr>
<td></td>
<td>Trastuzumab(36)</td>
<td>Breast, Stomach</td>
</tr>
<tr>
<td>Oestrogen receptor</td>
<td>Tamoxifen(37)</td>
<td>Breast (hormone-receptor-positive)</td>
</tr>
<tr>
<td>BRAF</td>
<td>Vemurafenib(38)</td>
<td>Melanoma</td>
</tr>
<tr>
<td>ALK</td>
<td>Crizotinib(39)</td>
<td>Lung</td>
</tr>
</tbody>
</table>

Note: New/off-label use for many targeted therapies is also common.(40, 41)
Monoclonal antibodies are also used in some cases and are even being explored as antivirals (42) but these too may come with adverse effects (see (43) for review). Anti-hormonals (e.g. selective estrogen-receptor modulators) and aromatase inhibitors that affect specialized but unessential normal tissues have also seen success as treatment for certain cancer types; however, their use is limited to hormone responsive cancers and still come with some side effects (44-46).

**Specific drug delivery systems for localized release**

Another approach towards selective therapy is through the use of targeted drug delivery systems. There are currently at least five clinically approved nanoparticle chemotherapeutics (47), with numerous others being studied now (48). Nanocarriers can help deliver hydrophobic- and multi-drug payloads, provide a more sustained drug release in a stimuli-sensitive manner, reduce drug clearance, and may even hinder development of multidrug resistance (MDR) (49). A variety of nanocarriers exist, including nanoparticles, polymeric micelles, dendrimers, magnetic nanoparticles, minicells, carbon nanotubes, nanogolds, liposomes, polymeric nanoparticle-drug conjugates, quantum dots, and viral-based nanoparticles (50). Some of the most promising molecular carriers currently are cucurbituril-based nanoparticles that offer improvements in drug stability (chemical and physical), solubility, and targeted release (51, 52). In addition to providing a hydrophobic cavity, cucurbiturils also act as cation-receptor hosts at their two portals, favoring protonated states of drugs and thus allowing for pK\textsubscript{a} shifts of up to 5 units (53). This ability to shift the pK\textsubscript{a} of molecules can be utilized, especially to improve bioavailability, in several ways by
increasing solubility or degree of ionization, activating prodrugs, or stabilizing the active form of the drug, for example. Chemical stability within these hosts can also be achieved via protection of drugs against nucleophiles and electrophiles or enhanced photostability and thermal stability in the solid state (see (53) for further review and discussion). These favorable drug delivery systems are poised to improve the efficiency and tolerability of targeted chemotherapeutics (54, 55), help overcome resistance to anticancer drugs such as cisplatin (56), and serve as valuable and intriguing host models of molecular recognition (Chapter 2).

**Structure-based, computer-aided modeling and discovery of cancer or viral and parasitic-infection drug leads**

When 3D structural information is known for a well-established drug target, computational or *in silico* approaches for small molecule design can be quite rewarding (57-59). Although several limitations still exist in virtual screening, we have seen much progress in drug lead identification, making even high throughput screening (HTS) seem less overwhelming when done first by docking the virtual library in order to narrow down the number of potentially interesting hits (60). Computer-aided drug discovery has already shown success in identifying hits against cancer (61-63), viral (64), and parasitic drug targets (65-67) that may help guide development of improved treatment options for patients in need of chemotherapy.

Mapping programs such as FTMAP (68) can be used to identify other “druggable” sites and possible allosteric pockets amenable to drug design, which has already been applied to cancer targets (69) and allowed for selective inhibitors of pathogenic enzymes instead of human enzymes (70). Glide (71), which was used in
work described in Chapter 4 to identify protein-protein inhibitors (PPI), was also valuable in a recently reported study to identify low affinity inhibitors of a Grb2 SH3C domain interaction (72). Additionally, one might imagine several ways to utilize computational modeling of mutated targets and subsequent virtual screening or lead optimization to identify inhibitors more selective for the mutated form. With developing resistance of viruses and cancer or parasitic cells, a faster method to identify new leads against the mutated targets could start with computational approaches to narrow down the search for such leads. These hits, although often not great drug candidates themselves, can be computationally optimized for improved potency, selectivity, and ADME properties before synthesis once the number of leads has been minimized (58). This could be especially important in the future if NGS becomes standard in patient diagnosis and treatment, as the wealth of available genomic data will accumulate continually.

**Accurate computational free energy calculations**

Part of lead optimization may include computational estimation of relative binding free energies to investigate how particular modifications might affect the affinity of the small molecule (58). Two rigorous methods include free energy perturbation (FEP)(73, 74) and thermodynamic integration (TI)(75), which use a coupling parameter to transition from one molecule to another via a thermodynamic cycle. Similar alchemical approaches have been used to estimate absolute binding free energy where a ligand goes from a fully coupled, interacting molecule to a fully decoupled, ideal gas-like molecule (76, 77) (Figure 1.2), but additional considerations
and precautions, such as extended molecular dynamics (MD) sampling and proper handling of restraining potentials, are necessary (See Chapter 2). Although more rigorous and often accurate approaches like these are computationally expensive, continued improvements in computing power, speed, and efficiency will help make these less time-consuming.

Figure 1.2. Thermodynamic Integration (TI) cycle using “computational alchemy” to estimate absolute binding free energy.

**Concluding remarks**

There have been tremendous advancements in DNA sequencing technology including enhanced speed, reduced costs, and improvement in ease of operation, making the potential of NGS for clinical impact in personalized, genomics-driven medicine imminent (78). With rapid identification of cancer mutations and drug-
resistant parasitic and viral infection, we will be able to put precision oncology (79, 80) or infection treatment in practice and limit unnecessary patient exposure to ineffective chemotherapeutics and radiation. These NGS advances will also propel identification of new oncogenes and candidate viral or parasitic genes essential to infection, both of which may also further confer drug resistance when mutated. However, without a suitable drug or method of delivery, a good target is not of much use. Rational, structure-guided high throughput virtual screening is an appropriate way to enrich a library for experimental validation and reduce the burden of HTS. Protocols exist to calculate free energy (see (81-85) for recent overviews) during lead optimizations that also help direct necessary modifications of a small molecule toward increased affinity before the daunting task of chemical synthesis. The following chapters will describe select investigations toward improved chemotherapy by identifying possible target-directed drug-leads and studying molecular recognition within a promising drug delivery system.
Chapter 2: On the Role of Dewetting Transitions in Host-Guest Binding Free Energy Calculations

Abstract

We use Thermodynamic Integration (TI) and explicit solvent molecular dynamics (MD) simulation to estimate the absolute free energy of host-guest binding. In the unbound state, water molecules visit all of the internally accessible volume of the host, which is fully hydrated on all sides. Upon binding of an apolar guest, the toroidal host cavity is fully dehydrated; thus, during the intermediate $\lambda$ stages along the integration, the hydration of the host fluctuates between hydrated and dehydrated states. Estimating free energies by TI can be especially challenging when there is a considerable difference in hydration between the two states of interest. We investigate these aspects using the popular TIP3P and TIP4P water models. TI free energy estimates through MD largely depend on water-related interactions and water dynamics significantly affect the convergence of binding free energy calculations. Our results indicate that wetting/dewetting transitions play a major role to slow the convergence of free energy estimation. We employ two alternative approaches – one analytical and the other empirically based on actual MD sampling – to correct for the standard state free energy. This correction is sizeable (up to 4 kcal/mol) and the two approaches provide corrections that differ by about 1 kcal/mol. For the system considered here, the TIP4P water model combined with an analytical correction for the
standard state free energy provides higher overall accuracy. This observation might be transferable to other systems in which water-related contributions dominate the binding process.

**Abbreviations**

Cucurbit[7]uril (CB[7]); 1,4-dimethyl alcohol bicyclo[2.2.2]octane (B2); thermodynamic integration (TI); molecular dynamics (MD); electrostatics (ele); van der Waals (vdW); host (H); guest (G); host-guest (HG); particle-mesh Ewald (PME); simulation (sim); center of mass (COM); number (No., Nr.).

**Introduction**

Cucurbit[7]uril (CB[7]; Figure 2.1) is a synthetic host that is attracting increasing interest for its ability to selectively bind various neutral or positively charged aromatic guests and metal complexes in aqueous solution.(86, 87) CB[7] can achieve affinities with some cationic guests that are greater than those typically measured for protein-ligand complexes.(88, 89) With their recognition properties and ease of synthesis, members of the CB[n] family have a wide range of potential applications including catalysis, gas purification and waste-stream remediation, crystal engineering, self-assembling and self-sorting systems, molecular machines, supramolecular polymers, self-assembled monolayers, and gene transfection.(90) As a host for metal complexes, CB[7] has also shown promise as a drug carrier for platinum chemotherapeutics such as Oxaliplatin by improving stability and decreasing negative side effects of the drug.(55) This auspicious molecular carrier demonstrates low
toxicity, efficient cellular internalization, and delivered drug bioactivity nearly as effective as that of the unbound drug in some cases (53, 91, 92), suggesting that it could be employed as a sophisticated drug delivery system. (93) In order to better utilize CB[7] in these applications, it is important to understand its behavior within a water environment and its molecular recognition of possible guests. Recent work on CB[7] describes its molecular dynamics (94) and affinity for several synthetic ferrocene and bicyclo[2.2.2]octane based guest molecules (e.g. B2 shown in Figure 2.1c) both experimentally and computationally. (95, 96) Even in seemingly simple, small host-guest systems with relatively rigid and symmetrical structures such as these (see Figure 2.1d), estimating the binding free energy can be quite challenging. However, to our knowledge no study to date investigated the wetting/dewetting events of the CB[7] system and the relevance of hydration in host-guest binding.

The importance of wetting/dewetting transitions in receptor or model host cavities has recently received much attention already. (97, 98) One of the most important consequences of these studies was the demonstration that non-covalent binding can largely depend on water-related contributions, not primarily on direct host-guest interaction. In particular, while the dehydration of the ligand molecules was shown to be the driving factor for binding, removal of solvent fluctuations can result in entropic penalties that are sizeable compared with the binding free energy. (97, 99) However, our understanding of solvation effects and their thermodynamic role in molecular recognition of more realistic systems remains quite poor. (98, 100, 101)
Here we study CB[7] hydration dynamics and its role upon binding of B2, a recently designed bicyclo[2.2.2]octane apolar guest. We investigate the dependence of hydration properties using the popular TIP3P and TIP4P water models with Thermodynamic Integration (TI) and explicit solvent molecular dynamics (MD) simulation. Overall, the closest match of estimated host-guest absolute binding free energy against experiment (within 0.3 kcal/mol) is achieved using the TIP4P water model. However, both TIP3P and TIP4P water models lead to qualitatively similar host hydration in the bound and unbound states, and along the
unphysical $\lambda$ state intermediates. For the unphysical $\lambda$ points at which significant host wetting/dewetting transitions occur, we observe remarkably large $\frac{\partial V}{\partial \lambda}$ fluctuations along the simulation time, ranging up to 140 kcal/mol and 130 kcal/mol for TIP3P and TIP4P water models, with standard deviations up to 22 kcal/mol and 19 kcal/mol, respectively. Furthermore, these fluctuations directly correlate with changes in host hydration, indicating that wetting/dewetting effects play a major role in the host-guest binding processes and, as a result, in determining free energy estimates. The work presented herein underscores the importance of water-related contributions and their convergence in free energy calculations of non-covalent binding.

**Computational Details**

**Molecular models**

Initial coordinates of the host-guest complexes (102) were previously generated with the Vdock (104) program from the CB[7] crystal structure.(55, 105) The general Amber force field (GAFF) was used to describe all intra-molecular energy contributions.(106, 107) The partial charges of CB[7] and the bicyclo[2.2.2]octane (B2) were calculated using the RESP program.(108, 109) To ensure symmetry, charge equivalence was enforced on each one of the seven units of the cucurbituril host (CB[7]). The molecular electrostatic potential was calculated at the HF/6-31G* level. Lennard-Jones parameters for B2 and CB[7] were assigned as previously described by Moghaddam et al.(110). These host-guest complexes were solvated in cubic boxes with a buffer region of 12 Å.
**Molecular Dynamics Simulations.**

All simulations were performed using a modified version of AMBER (111, 112) that enables calculation of absolute binding free energies with restraining and soft-core potentials. For both B2 and B2-CB[7], decoupling of the guest partial charges (ele) occurred first at 11 λ values (0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0). Decoupling of the van der Waals (vdW) energy terms utilized 19 different λ values (0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.65 0.7 0.75 0.775 0.8 0.825 0.85 0.875 0.9 0.925 0.95 1.0). All simulations were performed using rectangular periodic boundary conditions with isotropic position scaling in the isothermal-isobaric (NPT) ensemble, with a pressure reference of 1 atm and relaxation time of 0.5 ps. The system was kept at the reference temperature of 300 K using the Langevin thermostat (113) (collision frequency of 20 ps$^{-1}$) and Newton’s equations of motion were integrated using the leapfrog algorithm (114) with a MD time step of 0.001 ps. Using the sander Amber module, long-range electrostatics interactions were handled with the particle-mesh Ewald (PME) procedure and long-range van der Waals interactions approximated by a continuum model.(111) We refer the reader to Ref. (111) for additional computational details.

**Thermodynamic Integration.**

The change in free-energy between two states, $A$ and $B$, can be estimated using standard thermodynamic integration (TI)(115) as

$$\Delta G_{A \rightarrow B} = \int_{\lambda_A}^{\lambda_B} d\lambda \left( \frac{\partial V(\lambda)}{\partial \lambda} \right)_\lambda$$  \hspace{1cm} \text{(2.1)}
with $V(\lambda)$ representing the potential energy from a single trajectory of the system as a function of the coupling parameter $\lambda$, and with $\langle \cdots \rangle$ denoting the cumulative backward average at the given $\lambda$ point. The thermodynamic perturbation in this work went from state A ($\lambda=0$), in which the guest (G) experiences full interactions with the host (H) and the water environment, to state B ($\lambda=1$), in which the guest-related interaction energy vanishes. In order to improve phase-space sampling and avoid free energy singularities, the soft-core potential by Zacharias et al. (112, 116) was employed for all guest atoms ($\text{scalpha}=0.5$) (111, 117). Eq. 1 was integrated numerically using the trapezoidal rule.

**Statistical Analysis of Uncertainties**

A simulation standard error $\sigma_{\text{sim}}(t)$ of the time-varying potential energy derivative at a given $\lambda$, can be calculated as

$$
\sigma_{\text{sim}}(t) = \sqrt{\frac{1}{T-1} \sum_{i=1}^{T-1} \left[ \frac{\partial V_i(\lambda)}{\partial \lambda} - \langle \frac{\partial V_T(\lambda)}{\partial \lambda} \rangle_\lambda \right]^2} \sqrt{T}
$$

with $T$ being the total number of block-averages (118) throughout the single $i$th trajectory or all $N$ concatenated independent trajectories. $\left( \frac{\partial V_i(\lambda)}{\partial \lambda} \right)_\lambda$ denotes the potential energy derivative, block-averaged at time $t$, and $\langle \frac{\partial V_T(\lambda)}{\partial \lambda} \rangle_\lambda$ is the ensemble average over the entire simulation time at a given $\lambda$. As an example, $\sigma_{\text{sim}}(t)$
uncertainties are reported as error bars for $\frac{\partial V_T(\lambda)}{\partial \lambda}$ vs. $\lambda$ in Figure 2.2 (vertical bars). Then, a corresponding free-energy uncertainty can be obtained as

$$\sigma = \left( \int_{\lambda_1}^{\lambda_A} \sigma^2_{\text{sim}}(t) d\lambda \right)^{1/2}$$

(2.3)

We define a criterion, $\tau_{\text{sim}}$, for convergence monitoring of all TI simulations and automated decisions for termination of individual $\lambda$ runs. This criterion is based on the backward cumulative averages of the $\left( \frac{\partial V_T(\lambda)}{\partial \lambda} \right)_\lambda$ terms. This reads

$$\tau_{\text{sim}}(T-t) = \sqrt{\frac{1}{T-t-1} \sum_{t=T}^{T-1} \left( \frac{\partial V_T'(\lambda)}{\partial \lambda} \right)_{\lambda} - \left( \frac{\partial V_{T-t}'(\lambda)}{\partial \lambda} \right)_{\lambda}} / \sqrt{T-t}$$

(2.4)

Examples of the variation of this quantity backward in simulation time are given in Figure 2.3 and Table SI-2.1. A criterion of $\tau_{\text{sim}} < 0.1$ was employed to enforce proper convergence and automatically decide when to terminate TI runs. The term proposed in Eq. 2.4 is a well-behaved quantity that tends to zero for increasingly long trajectories.
Figure 2.2. Thermodynamic Integration along the $\lambda$ coupling parameter. vdW-decoupling TI curves of (a) guest and (b) host-guest system simulations in TIP3P (red) and TIP4P (blue). $<\partial V(\lambda)/\partial \lambda>_{\lambda}$ values and corresponding uncertainties, $\sigma_{\text{sim}}(t)$, are shown at each $\lambda$ sampled.

Figure 2.3. Convergence monitoring of $<\partial V(\lambda)/\partial \lambda>_{\lambda}$, the potential energy derivative backward average, across time with different criteria, $\tau_{\text{sim}}$, during B2-CB[7] vdW decoupling simulations where the highest $\partial V/\partial \lambda$ fluctuations occur using either TIP3P (left) or TIP4P (right) water models ($\lambda=0.8$ and $\lambda=0.825$, respectively).

**Separation of Thermodynamic States**

Corrections for the standard state free energy (119, 120) and separation of the thermodynamic states of interest were obtained using a harmonic restraining potential between the guest center of mass, $r_G$, and the host center of mass, $r_H$, to confine the guest to the region within the host cavity $V_{\text{cavity}}$, with a force constant $k_h$ of 2.5
kcal/mol. This $V_{\text{cavity}}$ was either estimated empirically from a 10 ns MD trajectory, according to

$$V_{\text{cavity}} = \frac{4}{3} \pi (r_G - r_H)^3 \quad (2.5)$$

or defined analytically as (121)

$$V_{\text{cavity}} = \left( \frac{2 \pi R T}{k_b} \right)^{3} \quad (2.6)$$

for comparison. Eqs. 2.5 or 2.6 were then utilized to find the standard-state free energy

$$\Delta G^{\circ}_{\text{Bind}} = \int_{\lambda_A}^{\lambda_B} d\lambda \left( \frac{\partial V(\lambda)}{\partial \lambda} \right) + R T \ln \left( \frac{V_{\text{cavity}}(\lambda)}{V^{\circ}} \right) \quad (2.7)$$

for guest transfer from this restricted $V_{\text{cavity}}$ volume to the bulk volume $V^{\circ}$ (as described in Ref. (122)).

**Results and Discussion**

In Figure 2.2, we include the following $\left( \frac{\partial V(\lambda)}{\partial \lambda} \right)$ vs. $\lambda$ TI curves for the van der Waals decoupling steps: i) solvated guest vdW decoupled along $\lambda$ in the unbound state (Figure 2.2a) and ii) the solvated guest vdW decoupled along $\lambda$ in the bound state.
(Figure 2.2b). Although \( \left\langle \frac{\partial V}{\partial \lambda} \right\rangle \) at most \( \lambda \)-values show very good convergence (see also supporting information Figure SI-2.1), we observe a region of \( \lambda \)-space from 0.7 to 0.875 with noticeably higher uncertainty and where convergence is considerably hindered within the host-guest vdW decoupling simulations. This narrow region is also where the \( \left\langle \frac{\partial V_T(\lambda)}{\partial \lambda} \right\rangle \) vs. \( \lambda \) TI curve changes most dramatically in the B2-CB[7] vdW transformation (Figure 2.2b).

In order to observe the dependence of \( \left\langle \frac{\partial V_T(\lambda)}{\partial \lambda} \right\rangle \) on total simulation time, we use a convergence criterion based on backward block-averaging. Figure 2.3 depicts the simulation convergence monitoring used in this work (see Computational Details) at the \( \lambda \) points with the largest statistical uncertainty for either TIP3P or TIP4P. These results further confirm that the convergence of \( \left\langle \frac{\partial V_T(\lambda)}{\partial \lambda} \right\rangle \) is considerably impeded at \( \lambda \) values within this region of higher uncertainty (e.g. \( \lambda=0.8 \) for TIP3P and \( \lambda=0.825 \) for TIP4P). Interestingly, the shape of the TI curves for the vdW decoupling simulations display remarkable dependence on the water model employed (Figure 2.2). The large dip in \( \left\langle \frac{\partial V_T(\lambda)}{\partial \lambda} \right\rangle \) is noticeably dissimilar in both width and location for the B2-CB[7] trajectories (Figure 2.2b).

Nevertheless, integration of each CB[7]-B2 bound vdW decoupling curve produces comparable differences in free energy contribution for TIP3P and TIP4P (\( \Delta G_{vdW} \) of 11.2 kcal/mol and 12.3 kcal/mol, respectively). Similar results were also
obtained for the unbound B2 guest, with $\Delta G_{\text{vdW}}$ of -4.87 kcal/mol and -6.61 kcal/mol for TIP3P and TIP4P respectively (Table 2.1). Figure SI-2.1 displays the TI curves calculated from the electrostatic decoupling steps. Not surprisingly, the electrostatic contribution of free energy change between the two water models is even smaller ($\Delta G_{\text{ele}}$ estimates for the unbound state in TIP3P and TIP4P differ by 0.8 kcal/mol and those for bound state by only 0.4 kcal/mol).

Table 2.1. Absolute Host-Guest Binding Free Energy Estimates.
a
<table>
<thead>
<tr>
<th>Water model</th>
<th>System</th>
<th>$\Delta G_{\text{ele}}$ [kcal/mol]</th>
<th>$\Delta G_{\text{vdW}}$ [kcal/mol]</th>
<th>Correction Term for Potential Bias from Host-Guest Restraint</th>
<th>$\Delta G_{\text{ele}}^a [\text{kcal/mol}]$</th>
<th>$\Delta G_{\text{exp}}^{(110)} [\text{kcal/mol}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIP3P</td>
<td>B2</td>
<td>-21.6 $\pm$ 0.22 $\pm$ 0.22</td>
<td>-4.87 $\pm$ 0.44</td>
<td>$\text{Volume} [\AA^3]$ $\Delta G_{\text{ele}}^b [\text{kcal/mol}]$ $\Delta G_{\text{vdW}}^b [\text{kcal/mol}]$ $\text{Correction Term for Potential Bias from Host-Guest Restraint}$</td>
<td>-6.4 (0.08) $\Delta G_{\text{ele}}^b [\text{kcal/mol}]$</td>
<td>-12.0 $\pm$ 0.9 $\Delta G_{\text{exp}}^{(110)} [\text{kcal/mol}]$</td>
</tr>
<tr>
<td></td>
<td>B2-CB[7]</td>
<td>-22.4 $\pm$ 0.23 $\pm$ 0.23</td>
<td>11.2 $\pm$ 0.77</td>
<td>$\Delta G_{\text{ele}}^b [\text{kcal/mol}]$ $\Delta G_{\text{vdW}}^b [\text{kcal/mol}]$ $\text{Correction Term for Potential Bias from Host-Guest Restraint}$</td>
<td>-3.3 (-6.0) $\Delta G_{\text{ele}}^b [\text{kcal/mol}]$</td>
<td>-11.2 $\pm$ 0.9 $\Delta G_{\text{exp}}^{(110)} [\text{kcal/mol}]$</td>
</tr>
<tr>
<td>TIP4P</td>
<td>B2</td>
<td>-20.8 $\pm$ 0.33 $\pm$ 0.34</td>
<td>-6.61 $\pm$ 0.60</td>
<td>$\text{Volume} [\AA^3]$ $\Delta G_{\text{ele}}^b [\text{kcal/mol}]$ $\Delta G_{\text{vdW}}^b [\text{kcal/mol}]$ $\text{Correction Term for Potential Bias from Host-Guest Restraint}$</td>
<td>-6.2 (0.07) $\Delta G_{\text{ele}}^b [\text{kcal/mol}]$</td>
<td>-13.4 $\pm$ 0.1 $\Delta G_{\text{exp}}^{(110)} [\text{kcal/mol}]$</td>
</tr>
<tr>
<td></td>
<td>B2-CB[7]</td>
<td>-22.0 $\pm$ 0.34 $\pm$ 0.34</td>
<td>12.3 $\pm$ 0.82</td>
<td>$\Delta G_{\text{ele}}^b [\text{kcal/mol}]$ $\Delta G_{\text{vdW}}^b [\text{kcal/mol}]$ $\text{Correction Term for Potential Bias from Host-Guest Restraint}$</td>
<td>-3.3 (-6.0) $\Delta G_{\text{ele}}^b [\text{kcal/mol}]$</td>
<td>-13.7 $\pm$ 1.0 $\Delta G_{\text{exp}}^{(110)} [\text{kcal/mol}]$</td>
</tr>
</tbody>
</table>

a. Uncertainties were calculated as in Eq. 2.3. See also Figure SI-2.1 and Figure 2.2 for raw $\Delta G_{\text{ele}}$ and $\Delta G_{\text{vdW}}$ data, respectively.
b. From Eq. 2.5, using maximum $r_G$-$r_H$ Center of Mass (COM)-COM distance (n/a for B2 alone)
c. From Eq. 2.5 using average $r_G$-$r_H$ Center of Mass (COM)-COM distance (n/a for B2 alone)
d. From Eq. 2.6; no restraint used for B2
e. Absolute binding free energy calculation derived from Thermodynamic Integration cycle (decoupled partial charges and decoupled van der Waals terms) with correction for standard-state:

\[
\Delta G_{\text{B2}}^{\text{ele}} + \Delta G_{\text{B2,vdW}}^{\text{ele}} - (\Delta G_{\text{B2,CB[7]}^{\text{ele}}} + \Delta G_{\text{B2,CB[7],vdW}}^{\text{ele}} + \text{B2-CB[7] Correction Term for Potential Bias from Restraint})
\]

To calculate the final free energy values according to Eq. 2.7, either an empirical or an analytical correction were used to take into account the standard-state for $\Delta G$. The empirical volumes were estimated from the total spherical space of radius ($r_G$-$r_H$) sampled by the guest in the bound state simulations. By using the maximum
distance between the guest center of mass (COM) and the host COM \( (r_{G-H}) \) in Eq. 2.5, the empirical correction energy was estimated to be -3.3 kcal/mol for both TIP3P and TIP4P (Table 2.1). If instead we use the average \( r_{G-H} \) during sampling, we estimate a correction of -6.0 kcal/mol using either water model (Table 2.1). Using Eq. 2.6, the calculated analytical restraining bias correction term for both water models was -4.1 kcal/mol. We hope that the average \( r_{G-H} \) correction term would converge (with further sampling in the host-guest MD simulation) to the same value as the analytical correction term but we believe that the difference in free energy values here might be attributable to underestimating the volume sampled by the guest using this average value and overestimating it using the maximum \( r_{G-H} \) sampled.

The absolute binding free energy estimate derived from simulations in TIP4P (best estimate: -13.7 kcal/mol) is slightly closer to experiment (-13.4 kcal/mol) than from those in TIP3P (best estimate: -12.0 kcal/mol). Taken as a whole, our calculations with TIP4P in combination with either correction term and the parameters used for the CB[7]-B2 host-guest system provide a more accurate estimate of absolute binding free energy. Our results do suggest that the TIP4P water model may describe the binding thermodynamics more accurately compared with the TIP3P model by bringing our calculated value significantly closer to the experimental value but further studies and comparisons would be necessary to confirm this.

These results are in good agreement with a previous study that showed that the TIP4P model describes water properties significantly more accurately compared with the TIP3P water model.(123) Due to the important role of the solvation/desolvation of the CB[7] and B2 during the binding event, we decided to further investigate the
relationship between slow simulation convergence and host hydration with both water models.

In Figures 2.4 and 2.5, we examine the average water density surfaces throughout selected trajectories with TIP3P and TIP4P. Because both host-guest water model systems reach their minimum on the vdw TI curve at $\lambda=0.85$ (Figure 2.2b), in Figure 2.4 we include the average water density surfaces at this point in addition to those for the unbound ($\lambda=1.0$) and bound ($\lambda=0.0$) cases, with densities shown at $\sim0.95$ times bulk TIP3P and $\sim1.1$ times bulk TIP4P. In Figure 2.5, we display higher density water surfaces of the B2-bound case, with densities shown at $\sim1.3$ times TIP3P bulk water density and $\sim1.5$ times TIP4P bulk water density.

At $\lambda=1$, the average water density from our 50 ns simulation was highest inside of the host. The average volume surfaces of the TIP4P water oxygen atoms are shown here as a few small clusters, around and between the nitrogen atoms of the host, and 3 toroidal shapes with the largest ring in the center of the host and two others near the top and bottom (Figure 2.4), which is consistent with other models of the host alone (124). Interestingly, at this same density value TIP3P waters display average surfaces that appear as more of a cylinder of ordered water oxygen atoms inside and more of a seven-petal flower shape outside the host but show some similar density clusters near the host nitrogen atoms.

Both water models still show similar ordering near the host nitrogen atoms at $\lambda=0.85$ but we do observe some different volume density patterns (Figure 2.4). TIP4P waters still form ring-like average density surfaces but the center toroid is smaller and
those on the outside of the host become more concave than in the unbound ($\lambda=1$) case. Here we also observe an additional concave flower-shaped surface of TIP3P waters at this density (as at $\lambda=1$) appearing on both ends of the host. However, while TIP3P water molecules do occupy the host at this $\lambda$, the surfaces do not appear inside the host at this high-density value, unlike the TIP4P and unbound TIP3P cases.
Figure 2.4. Average hydration surfaces of 50 ns vdW host-guest simulations for unbound (λ=1.0; top) and bound (λ=0.0; bottom) thermodynamic states. Corresponding maps are also shown for the λ=0.85 intermediate unphysical state (middle). Hydration maps for TIP3P (red wireframe, left of vertical slice) and TIP4P (blue wireframe, right of vertical slice) water models were generated using the average density of the water oxygen atom on a grid (0.3 Å resolution) after superimposing MD snapshots using all atoms of the host (shown as licorice representation here). The isosurfaces represent water density that are ~1.1 times TIP4P bulk water density and ~0.95 times TIP3P bulk water density. Images generated with VMD(125).
In the B2-bound host case ($\lambda=0.0$), again we see unique surface patterns, even between the two water models. The TIP4P surface almost completely covers the outside of the host, but TIP3P surfaces only appear as large clusters on the outside of the host and as fuller flower shapes near the host oxygen atoms (Figure 2.4). As expected though, no water densities are seen within the host cavity because the guest vdW parameters remain fully coupled in this case. Higher TIP4P density surfaces appear just as small clusters circling the host and as two toroids, one at each end of the host, while TIP3P surfaces form only small circular clusters along these outside rings, but do not form solid toroids there yet (Figure 2.5). As expected here, where the guest occupies the host, still no water densities appear within its cavity. The distinctive hydration maps observed for TIP4P and TIP3P models at the same specific isovalue are expected, owing to both the different mixed interactions of these water models with the same host molecule as well as the well-known structural disparities between the water models themselves.

Figure 2.6 displays both the $\frac{\partial V}{\partial \lambda}$ values (Figure 2.6a,c) and the number of waters inside the host (Figure 2.6b,d) along simulation time for several $\lambda$ points. All waters within 4 Å of both the very top and bottom atoms of the B2 core were measured at each frame of the 50 ns vdW host-guest complex trajectories and compared to the $\frac{\partial V}{\partial \lambda}$ measured at that time point. Clearly, two distinctive energy extremes are represented by hydrated and dehydrated states (Figure 2.6). Frequent fluctuations can occur between the two hydration states when the guest vdW
decoupling is sufficiently high to allow for increased host hydration, but still low enough that the host cannot remain stably hydrated throughout the rest of the simulation. The $\frac{\partial V}{\partial \lambda}$ of the host-guest system at several of the higher $\lambda$’s near 0.825 fluctuate between two very different hydration extremes with periods of several nanoseconds during vdw TI simulations (Figure 2.6). The frequency of wetting/dewetting transitions varies depending on the $\lambda$ point considered, with highest frequencies of up to 10-15 events occurring in under 10 ns of simulation. This behavior is the major factor slowing free energy convergence. The slowest simulations to converge were at $\lambda=0.8$ for TIP3P and at $\lambda=0.825$ in TIP4P as these experienced the most frequent and extreme host hydration fluctuations, and therefore energy fluctuations, throughout their trajectories (Figure 2.6). We note that apparent convergence would be misleadingly attributed to the simulations considering MD periods shorter than the wetting/dewetting transition timescales (see Figure 2.3 and Supporting Information Table SI-2.1 for an example).

Because host hydration plays such a major role in the B2-CB[7] vdw decoupling simulations and free energy estimates, we include Table 2.2, which reveals some additional key differences in the TIP3P and TIP4P host wetting events during our simulations at $\lambda$’s where fluctuations occurred most. There is a striking yet plausible difference in both the percent of time that the host is hydrated as well as the mean number of waters inside the host from $\lambda=0.775$ to 0.825 between the simulations using the two different water models, with more TIP3P molecules occupying the host cavity than TIP4P during these simulations. However, at the
unbound ($\lambda=1.0$) and bound ($\lambda=0.0$) states, the two models behave quite similarly with the host stably hydrated or dehydrated throughout most if not all of the simulation, respectively. Additionally, at $\lambda=0.8$ our TIP3P simulation shows up to seven waters inside the host, while that of TIP4P fits only a maximum of six inside. The time to reach the first maximum hydration event and the mean number of waters inside the host are also included for each case. Overall, this analysis does confirm the opposite behavior at the two $\lambda$ endpoints though, with a hydrated cavity at $\lambda=1$ where the guest is fully decoupled (Figures 2.4 and 2.6b), representing the unbound state, and a dehydrated host cavity at $\lambda=0$ where the guest is fully coupled, representing the bound state (Figures 2.4, 2.5 and 2.6b).

![Figure 2.5](image.png)

**Figure 2.5.** Average hydration surfaces (as in Figure 2.4) for host-guest vdW simulation bound ($\lambda=0.0$) thermodynamic state with TIP3P (red wireframe, left side) and TIP4P (blue wireframe, right side) water oxygen atom densities shown in and around the host structure (licorice representation) from topview (left) and sideview (right) for each at ~1.5 times TIP4P bulk water density and ~1.3 times TIP3P bulk water density.
Figure 2.6. Fluctuations of (a,c) $\langle \partial V/\partial \lambda \rangle$ values and (b,d) number of water molecules in the CB[7] host system over time for the TIP3P (red, a-b) and TIP4P (blue, c-d) water models during 50 ns host-guest vdW simulations. The physical process of host-guest binding goes from the unbound state ($\lambda = 1$) to the bound state ($\lambda = 0$), while absolute binding free energies were effectively calculated conveniently in the opposite direction by annihilation of ligand interactions (from $\lambda = 0$ to $\lambda = 1$).
Figure 2.7 further confirms the correlation of system free energy and host hydration seen with both water models (Figure 2.6, see also Supporting Information). Here we look at the $\left( \frac{\partial V}{\partial \lambda} \right)$ vs. the number of TIP4P or TIP3P waters in the host cavity at $\lambda=0.85$, a point where the host is usually hydrated with up to seven waters but on average holds about 3 waters (Table 2.2). Although the standard deviations of $\frac{\partial V}{\partial \lambda}$ are large, there is a clear trend toward lower energy with increasing number of water molecules using both models, which is consistent across all $\lambda$’s where the host cavity is capable of hydration (see Figure SI-2.2). At the $\lambda$ endpoint states representing the unbound and bound host, we observe very different, stable behavior and no such correlation (Figure 2.6 and Figure SI-2.2). The range of $\frac{\partial V}{\partial \lambda}$ values is smaller (~20 kcal/mol) at both endpoints, with the unbound case showing a steadier $\left( \frac{\partial V}{\partial \lambda} \right)$ at or near 1.2 kcal/mol, regardless of the number of waters inside the host, and the bound case (where no waters ever fit inside the host) averaging around 0.2 kcal/mol (Figure SI-2.2). Clearly, this system presents a fascinating yet challenging behavior for TI absolute binding free energy calculations using explicit water and demonstrates how important the role of wetting and dewetting events is in the accuracy and convergence of free energy calculations.
Table 2.2. Hydration During the Endpoint and Unphysical States Employed for Thermodynamic Integration *

<table>
<thead>
<tr>
<th>Host Hydration Features</th>
<th>Water Model</th>
<th>(\lambda=0.0)</th>
<th>(\lambda=0.775)</th>
<th>(\lambda=0.8)</th>
<th>(\lambda=0.825)</th>
<th>(\lambda=0.85)</th>
<th>(\lambda=0.875)</th>
<th>(\lambda=1.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time hydrated (%)</td>
<td>TIP3P</td>
<td>0</td>
<td>29.5</td>
<td>59.0</td>
<td>87.8</td>
<td>93.8</td>
<td>96.7</td>
<td>97.0</td>
</tr>
<tr>
<td></td>
<td>TIP4P</td>
<td>0</td>
<td>8.02</td>
<td>18.3</td>
<td>62.1</td>
<td>86.2</td>
<td>88.8</td>
<td>97.7</td>
</tr>
<tr>
<td>Time to 1st max. hydration event (ns)</td>
<td>TIP3P</td>
<td>n/a</td>
<td>6.47</td>
<td>35.0</td>
<td>45.7</td>
<td>10.2</td>
<td>6.88</td>
<td>12.7</td>
</tr>
<tr>
<td></td>
<td>TIP4P</td>
<td>n/a</td>
<td>21.0</td>
<td>3.40</td>
<td>2.26</td>
<td>2.02</td>
<td>2.00</td>
<td>1.44</td>
</tr>
<tr>
<td>Mean Nr. of water molecules inside the host</td>
<td>TIP3P</td>
<td>0</td>
<td>0.6</td>
<td>1.4</td>
<td>2.4</td>
<td>2.6</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>TIP4P</td>
<td>0</td>
<td>0.1</td>
<td>0.4</td>
<td>1.8</td>
<td>3.0</td>
<td>3.3</td>
<td>3.0</td>
</tr>
<tr>
<td>Maximum Nr. of water molecules inside the host</td>
<td>TIP3P</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>TIP4P</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Minimum Nr. of water molecules inside the host</td>
<td>TIP3P</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>TIP4P</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* Results are summarized for the \(\lambda\) states at which largest wetting/dewetting transitions occur and compared to endpoint states.

Figure 2.7. Correlation between \(\langle \partial V/\partial \lambda \rangle\) values and number of waters within host for \(\lambda=0.85\) vdW simulation with TIP3P and TIP4P water models. Vertical error bars are the standard deviation values.
Conclusions

Thermodynamic Integration was used to calculate the absolute binding free energy of a high affinity host-guest system in explicit solvent. Using both TIP3P and TIP4P water models, our estimations correspond remarkably to experimental calculations. Although both models proved to be adequate for our studies of this host-guest system, our simulations in TIP4P did lead us to an absolute binding free energy estimate gratifyingly closer to the experimental value. We observe a toroidal average water density surface first appearing inside the host in the unbound state ($\lambda=1.0$) or ordered around the top and bottom host oxygen molecules in the guest-bound state ($\lambda=0.0$), with fascinating patterns of intermediate water ordering seen both inside and out of the host at varying $\lambda$-values in between. Significant changes in $\frac{\partial V}{\partial \lambda}$ occur along molecular dynamics simulations as the guest van der Waals interactions are decoupled along $\lambda$ space using both water models. The dynamic behavior of host wetting and dewetting events is directly correlated with these large energy fluctuations of the system. For both water models, we see a decrease in $\frac{\partial V}{\partial \lambda}$ with increasing water molecules inside the host during van der Waals simulations with sufficient guest decoupling. The work described here emphasizes the importance of wetting/dewetting transitions and their influence on free energy estimation for the host-guest system considered in this study. We believe that the wetting transitions observed may be relevant in more complex biological scenarios as well, including some protein-ligand
binding scenarios (98). However, it is also expected that a protein cavity, which is accessible from the bulk from one side only, might have quite largely different hydration properties than a host guest system, such as CB[7], as the latter is open to bulk from two symmetric openings. Our study is, to our knowledge, the first that addressed these slow water transitions in a host-guest system.

Acknowledgments

This work was supported by NSF, NIH, HHMI, the National Biomedical Computation Resource, and the Center for Theoretical Biological Physics. The authors thank Michael Gilson and his group for providing original system coordinates and helpful discussions. R. B. acknowledges the University of Utah, department of Medicinal Chemistry for startup funding to support the last stage of this project. J. M. Ortiz-Sánchez acknowledges the Fulbright Commission/Generalitat de Catalunya Program and the Generalitat de Catalunya for a Fulbright and a Beatriu de Pinos postdoctoral grants, respectively.

Chapter 2 is a minimally modified reprint of the material as it appears in Rogers, Kathleen E.; Ortiz-Sanchez, Juan M.; Baron, Riccardo; Fajer, Mikolai; de Oliveira, César A.F., McCammon, J. Andrew. “On the Role of Dewetting Transitions in Host–Guest Binding Free Energy Calculations.” Journal of Chemical Theory and Computation 2013, 9 (1), 46-53. I was the first author and principal investigator on this paper. Dr. Ortiz-Sanchez provided useful discussions, Professor Baron was instrumental in much of the design of this work, Dr. Fajer made essential modifications to the AMBER code used, Dr. Oliveira completed all system
parameterizations, and Professor McCammon suggested and further guided the project.

Supporting Information

TI curves of partial charge (ele) decoupling steps and correlation graphs of \( \left\langle \frac{\partial V}{\partial \lambda} \right\rangle \) vs. number of waters inside the host at additional \( \lambda \)'s are included. Examples of the variation of the convergence monitoring criterion, \( \tau_{sim} \), are also provided to show changes in statistical uncertainty with increasing simulation time.

<table>
<thead>
<tr>
<th>Water Model</th>
<th>( \tau_{sim} )</th>
<th>( \Delta G_{vdW} ) [kcal/mol]</th>
<th>Longest time, ( t ), to convergence for an individual ( \lambda ) [ns]</th>
<th>Total time to convergence [ns]</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIP3P</td>
<td>0.5</td>
<td>11.4 +/- 1.6</td>
<td>0.90</td>
<td>7.09</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>11.2 +/- 1.2</td>
<td>8.68</td>
<td>53.6</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>11.3 +/- 1.0</td>
<td>15.6</td>
<td>96.6</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>11.2 +/- 0.77</td>
<td>35.5</td>
<td>228</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>12.2 +/- 1.5</td>
<td>0.54</td>
<td>5.64</td>
</tr>
<tr>
<td>TIP4P</td>
<td>0.2</td>
<td>12.1 +/- 1.1</td>
<td>11.2</td>
<td>52.3</td>
</tr>
<tr>
<td></td>
<td>0.15</td>
<td>12.4 +/- 1.0</td>
<td>20.8</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>12.3 +/- 0.82</td>
<td>46.1</td>
<td>225</td>
</tr>
</tbody>
</table>
Figure SI-2.1. Figure SI-2.1. 

\[ \left( \frac{\partial V_{f}(\lambda)}{\partial \lambda} \right)_{\lambda} \] 

vs. \( \lambda \) TI curves of partial charge (ele) decoupling steps for both systems and water models. \( \sigma_{\text{sim}}(t) \) uncertainties are reported as vertical error bars.
Figure SI- 2.2. Correlation between the $\frac{\partial V}{\partial \lambda}$ and the number of waters within the host for vDW TI simulations at endpoint $\lambda$ values and those with slow convergence. Vertical error bars are the standard deviation values.
Chapter 3: Novel cruzain inhibitors for the improved treatment of Chagas disease

Abstract

The protozoan parasite Trypanosoma cruzi, the etiological agent of Chagas disease, affects millions of individuals and continues to be an important global health concern. The poor efficacy and unfavorable side effects of current treatments necessitate novel therapeutics. Cruzain, the major cysteine protease of T. cruzi, is one potential novel target. Recent advances in a class of vinyl-sulfone inhibitors are encouraging; however, as most potential therapeutics fail in clinical trials and both disease progression and resistance call for combination therapy with several drugs, the identification of additional classes of inhibitory molecules is essential. Using an exhaustive virtual-screening and experimental-validation approach, we identify several additional small-molecule cruzain inhibitors. Further optimization of these chemical scaffolds could lead to the development of novel drugs useful in the treatment of Chagas disease.

Abbreviations

Trypanosoma cruzi (T. cruzi); relaxed complex scheme (RCS); root mean square deviation (RMSD); dynamic light scattering (DLS); extra precision (XP); Induced Fit Docking (IFD)
Introduction

The World Health Organization estimates that over 10 million people are infected by the protozoan parasite *Trypanosoma cruzi*, with another 25 million at risk. The associated illness, called Chagas disease, is spread through a triatomine vector or through blood transfusion. The acute phase lasts at most a few months and is characterized by mild symptoms such as fever, malaise, facial edema, generalized lymphadenopathy, and hepatosplenomegaly. In approximately 30% of infected patients, parasite multiplication via asynchronous cycles contributes to the chronic stage of the disease, with the associated cell destruction, reinfection within the reticuloendothelial system, and organ infection. Infection of the heart, digestive tract, and central nervous system can lead to fatal heart-rhythm abnormalities, megacolon, and dementia, respectively.

*T. cruzi* is not susceptible to many of the drugs used to treat closely related parasites like *Trypanosoma brucei*. Benznidazole and nifurtimox are the only available therapies for acute-phase Chagas disease. These nitroheterocyclics are highly toxic and have poor efficacy in long-lasting chronic infections. No extensive studies of the long-term sequellae of these therapeutics have been conducted in humans, but several reports of neuropathy and tumorigenic or carcinogenic effects have been described. Efforts to develop a vaccine against *T. cruzi* have also failed thus far, likely because the disease pathology has an autoimmune component.
The major *T. cruzi* cysteine proteinase cruzain (also referred to as cruzipain, the full-length native enzyme) has been shown to be crucial for all stages of the parasite life cycle. This papain-like cysteine protease is thought to play an important role in differentiation, cell invasion, intracellular multiplication, and immune evasion (135, 136). Furthermore, studies have demonstrated that cysteine-proteinase inhibitors have trypanocidal activity with negligible mammalian toxicity (137).

Previous efforts have identified vinyl sulfones, sulfonates, and sulfonamides as high-affinity cruzain inhibitors (138, 139); one of these vinyl sulfones, K11777, is currently undergoing Investigational New Drug (IND) enabling studies (140, 141). α-ketoamide-, α-ketoacid, α-ketoester-, aldehyde-, and ketone-based inhibitors have also been described (142–144). While these successes are encouraging, many potential drugs, including those that enter clinical trials, ultimately fail to gain approval (145), and those that are approved are subject to growing parasitic resistance. Consequently, a diverse set of inhibitory scaffolds that can be optimized into distinct therapeutic candidates is urgently needed.

Hoping to contribute to this ever-growing diverse set of compounds, we here use an advanced virtual-screening methodology that accounts for receptor flexibility to identify three promising non-covalent inhibitors of *T. cruzi* cruzain.

**Experimental Methods**

**Ligand Preparation**

A small-molecule library was prepared from the ligands of the NCI Diversity
Set II using the Schrödinger LigPrep program (146). Protonation states were assigned at pH 5.5 to mimic the natural acidic environment of the *T. cruzi* digestive vacuole. Multiple tautomers and stereoisomers were generated. One ligand could not be processed by LigPrep; instead, Discovery Studio (147) was used to add hydrogen atoms to this ligand and to optimize its geometry.

**Initial Screen Against the Crystal Structure**

The prepared ligand models of this small-molecule library were docked into a 1.20 Å crystal structure of cruzain (PDB ID: 1ME4)(143), with hydrogen atoms included using PDB2PQR (148, 149) at pH 5.5. Residues CYS25 and H159 (called H162 by some) formed the thiolate-imidazolium pair required for the catalytic mechanism (150) of the proteinase at this pH. This initial virtual screen was performed using the CDOCKER docking software (147) with a docking sphere 15 Å in diameter centered on the coordinates of the crystallographic ligand.

**Rescoring Protocol**

The CDOCKER-predicted pose of each ligand model was rescored using six additional scoring functions: LigScore1, LigScore2 (151), PLP1, PLP2 (152), PMF (153), and PMF04 (154). The best scoring models as evaluated using each of these seven scoring functions were compiled into a new small-molecule library of 302 models (182 unique ligands) enriched for predicted *cruzain* inhibitors.
**Molecular Dynamics Simulations**

The molecular dynamics simulations used in the current study have been described previously (155). In brief, the simulations were based on a 1.20 Å cruzain crystal structure (PDB ID: 1ME4) (143) protonated at pH 5.5 to mimic the natural acidic environment of the *T. cruzi* digestive vacuole. Following appropriate minimization and equilibration, five distinct 20-ns simulations of the cruzain protein bound to a hydroxymethyl ketone inhibitor, [1-(1-BENZYL-3-HYDROXY-2-OXO-PROPYLCARBAMOYL)-2-PHENYL-ETHYL]-CARBAMIC ACID BENZYL ESTER, were performed. The *gromos* clustering algorithm (156) was used to cluster 4,002 conformations extracted from the simulations every 50 fs. We found that decreasing the cutoff below 0.95 Å resulted in a precipitous rise in the number of clusters; consequently, we chose an RMSD cutoff of 0.95 Å, which yielded 24 clusters. The central member of each cluster, considered most representative, was selected for subsequent analysis; this set of central members is said to constitute an ensemble.

**Relaxed-Complex Screen**

The 302 compound models of the enriched small-molecule library were docked into the 24 clusters of the ensemble using CDOCKER (Accelrys). Each of these docked small-molecule models was rescored with the following scoring functions: LigScore2 (151), PLP1, PLP2 (152), PMF (153), and PMF04 (154). For each ligand/scoring-function pair, an ensemble-average score was calculated according to the following equation:
where $\bar{E}$ is the weighted ensemble-average score, $w_i$ is the size of cluster $i$, and $E_i$ is the best score of each unique ligand, independent of tautomeric or stereoisomeric form, docked into the centroid of cluster $i$.

Two methods were used to select compounds for subsequent experimental validation. First, for each of the five scoring functions, the compounds were ranked from best to worst by the ensemble-average score. The top seven compounds were selected from each of these five ordered lists and merged into a single list of potential binders. Second, the average rank of each compound across all five scoring functions was calculated. The compounds were then reordered by this average rank, and the top thirty were likewise identified as potential binders. Any compound indicated by either of these two protocols was subsequently recommended for preliminary experimental validation.

**Enzymatic Assays**

Each compound was obtained from the National Cancer Institute’s Development Therapeutics Program, which guaranteed 90% purity. Compounds were tested for cruzain enzymatic inhibition using a protocol that has been described previously (157). The eight compounds with the lowest $IC_{50}$ values were assessed for aggregation by observing enzymatic activity under varying experimental conditions.
As detergent is known to disrupt colloidal aggregation (158), inhibition in the absence of detergent was compared to inhibition in the presence of Triton X-100 (0.02%) and Tween (0.002%). The reducing agent was also varied (10mM DTT or 10mM Beta-Mercaptoethanol) to control for redox cycling compounds (RCC) (159). Each experimental condition was tested in at least two separate experiments. Finally, a dynamic light scattering (DLS) technique, described in detail elsewhere (160), was applied to the top four compounds to further confirm the presence or absence of aggregation. Additional experimental details can be found in the Supporting Information.

**Final Pose Predictions**

All compounds submitted for experimental validation were subsequently docked a final time into the binding pocket of the Protein Preparation Wizard prepared (161, 162) crystal structure (PDB ID: 1ME4) using the Induced Fit Docking (IFD) module of the Maestro 9.2 (Schrödinger, LLC) computer package with Glide XP precision (163, 164). For each of the top three non-aggregating ligands, the best-scoring pose that positioned the inhibitor near the crucial catalytic triad was selected and visualized using PyMOL (165). Although the top poses using this IFD protocol were generally similar to those from the RCS CDOCKER work, we choose to show them here in the commonly represented crystal structure conformation for ease of recognition to the reader of subsites within the well-characterized active site of cruzain.
Results and Discussion

Discovered by Carlos Chagas in 1909 (166, 167), *T. cruzi* is one of only two known pathogenic *Trypanosoma* species. Current trypanocidal therapeutics like nifurtimox and benznidazole are inadequate because they are toxic (131-133), subject to growing resistance (168), and ineffective at eradicating the parasite and preventing cardiomyopathy over the long term (169). Given the dire need for novel therapies, we here use virtual-screening methods to identify three promising inhibitors of cruzain, a critical cysteine protease required for *T. cruzi* survival.

Weaknesses of Virtual Screening

Virtual-screening techniques have been used to identify a number of inhibitors in recent years (see, for example, references (64, 170-175)). Though widely used, these screens are often characterized by many false positives and negatives. Two principal weaknesses explain these inaccuracies. First, there are errors intrinsic to the scoring functions themselves. Because virtual-screening efforts often attempt to identify true binders from among the many thousands of molecules in a compound library, they are generally optimized for speed at the expense of accuracy. Second, current docking programs account for, at best, only limited receptor flexibility. When a small-molecule binder encounters its receptor *in vivo*, that receptor often undergoes conformational rearrangements or an “induced fit” to better accommodate the ligand. These *holo* conformations can differ significantly from those of x-ray crystallographic structures. Even if the perfect docking scoring function did exist, it could not accurately predict binding affinity if receptor flexibility were ignored. In the current
work, we use a scheme designed to minimize both these sources of error.

**Overcoming Inaccuracies Inherent to the Scoring Functions Themselves**

The ligands of the NCI Diversity Set II (176) were initially docked into a cruzain crystal structure using CDOCKER (147) because this program was able to capture the crystallographic poses of two positive controls. Docking with AutoDock (177) was initially performed, but docked and crystallographic poses differed significantly. The CDOCKER-docked poses were subsequently rescored using several different scoring functions, and potential binders were selected by consensus rather than by the score of any single function. Consensus scoring has two advantages. First, when multiple scoring functions are combined, the errors of each may in part cancel out (178). Second, different scoring functions likely have different intrinsic weaknesses and strengths. Some, for example, may better account for hydrophobic contacts, while others better capture electrostatic interactions. When combined, accuracy may be improved if the weaknesses and strengths of the constituent functions are complementary (179).

The scoring functions used in the current work come from different classes and thus provide independent assessments of ligand binding that may be complementary. Scoring functions fall into one of three categories: force field, empirical, and knowledge based. Force-field scoring functions like that used by CDOCKER, based on the CHARMm force field, evaluate ligand binding by accounting for bonded and non-bonded atomic interactions explicitly. Empirical scoring functions like LigScore1, LigScore2, PLP1, and PLP2 are based on counting the number of different types of
receptor-ligand interactions (e.g., hydrogen-bond and hydrophobic interactions) as well as countable changes in molecular properties (e.g., the number of rotatable bonds immobilized upon ligand binding). Finally, knowledge-based scoring functions like PMF and PMF04 are based on statistical analyses of large structural databases like subsets of the Protein Data Bank (180). Intermolecular interactions that occur more often than expected by pure chance are assumed to be energetically favorable.

The top candidate binders from this initial crystal-structure screen, as judged by consensus scoring, were compiled into a single list of 302 small-molecule models that were subsequently subjected to further study.

**Overcoming Inaccuracies Caused by Ignoring Full Receptor Flexibility**

While consensus scoring may have helped overcome in part the errors intrinsic to each individual scoring function, those caused by ignoring receptor flexibility remained. In order to overcome this second challenge, we first studied receptor dynamics by performing five distinct 20-ns molecular dynamics simulations of cruzain, described elsewhere (155). The protein conformations sampled during these simulations were clustered into 24 groups using an RMSD-based clustering algorithm (156). Each of the centroid members of each cluster, considered to be the most representative, was identified, and the group of all centroids is said to constitute an ensemble.

Having identified multiple cruzain conformations, we re-docked the 302 small-molecule models identified in the initial crystal-structure screen into each of the 24 members of the ensemble, again using CDOCKER. The ligands were then re-ranked
by an ensemble-average score that was not dependent on a single crystal structure but rather accounted for receptor flexibility. This multi-receptor docking protocol, called the relaxed complex scheme (RCS), has been used successfully in the past to identify inhibitors of FKBP (170), HIV integrase (64), and *T. brucei* RNA editing ligase 1 (171), among others (172-174).

**Final Rescoring**

As each of the 302 small-molecule models was docked into 24 different cruzain conformations, there were 7,248 docked poses in all. Each of these 7,248 poses was rescored, again using multiple scoring functions. Ensemble-average scores were calculated for each scoring function, and the 302 small-molecule models were appropriately ranked.

Two criteria were used to identify likely inhibitors from among these ranked compounds. First, we selected the 7 best inhibitors as predicted by each of the individual scoring functions and merged them into a single list of 22 candidates. Second, the top 30 predicted ligands as judged by the ensemble-average rank were likewise identified. In all, 37 unique candidate inhibitors were selected; of these, the best 30 were tested experimentally.

**Predicted Binding Modes**

Of the 30 compounds tested, eight inhibited cruzain at 100 μM (Figure 3.1, all structures below the first arrow). Using standard conditions, the best-scoring compound had an initial IC$_{50}$ value of 471 nM (see Figure 3.1, asterisk). Subsequent experiments suggested that this compound inhibited cruzain nonspecifically *via*
aggregation (Supporting Information). Fortunately, three other compounds, although less potent, did in fact appear to inhibit cruzain specifically (Supporting Information). These three compounds, NSC 227186, NSC 67436, and NSC 260594, have IC$_{50}$ values of 16 µM, 63 µM, and 66 µM, respectively (see Figure 3.1, bottom row). While these compounds lack the nanomolar affinity characteristic of approved drugs, they do possess the low micromolar affinity typical of lead compounds. With proper optimization, including fragment addition, moiety swapping, and similarity searching, these compounds could be transformed into viable drug candidates. We are hopeful that these new leads will be helpful to those in the drug-discovery community targeting Chagas disease.

Figure 3.2 shows the predicted binding poses (a,c,e) and important interactions (b,d,f) associated with each of the three most promising compounds using the IFD protocol and visualized in PyMOL (described in Methods). Figure 3.2b shows the predicted polar contacts of NSC 227186 with GLN19, THR185, and TRP184; TRP184 is also oriented for possible aromatic stacking with one of the ligand aromatic moieties. GLU208 (called GLU205 by some) is predicted to swing even further away from the cruzain S2 subsite than is seen in the 1ME4 crystal structure, allowing a ligand methylpyrrolidine moiety to occupy the site, reminiscent of the Phe and Tyr rings of several known inhibitors (138-143, 181, 182). An oxane moiety is also predicted to occupy the S3’ subsite (138). NSC 227186, a commercially available aminocoumarin antibiotic also referred to as Clorobiocin, is derived from several *Streptomyces* species and shows additional activity against *Trypanosoma brucei* Uridine Diphosphate Galactose 4’-Epimerase (TbGalE) (174) as well as the
Escherichia coli DNA gyrase (183). It was also previously shown to inhibit growth of *T. cruzi* but the target was unknown. (184) This work may have identified a more specific mechanism for Clorobiocin’s inhibitory activity on *T. cruzi* by finding a notable IC$_{50}$ with cruzain.

**Figure 3.1.** The structures of all experimentally validated compounds. Associated IC50 values are given for the top three non-aggregating inhibitors. Means and standard deviations were calculated by considering all IC50 values associated with each compound, measured under varying experimental conditions as described in the Methods ($n = 8$). *NSC 61610 appears to inhibit cruzain nonspecifically *via* aggregation (see the Supporting Information).

NSC 67436 (Figure3.2d) is predicted to form even more interactions with the
cruzain receptor, including various hydrogen-bond interactions with residues GLN19, GLY66, ALA141, ASP161, and ASN182, as well as possible aromatic stacking with HIS162. Two distinct cyclohexanamine moieties are predicted to occupy the S2 and S1’ subsites, and a chlorocyclohexane (4-chlorocyclohexane-1,3-dicarbaldehyde) moiety is predicted to bind near S1. S4 and S3’ are also occupied, both with distinct imidazoline rings (138). Interestingly, NSC 67436 has previously shown anti-
*Trypanosoma brucei* activity as well (185).

Figure 3.2. The Predicted Binding Poses and Receptor-Ligand Interactions of the Experimentally Validated Inhibitors. a, c, and e: Compounds and cruzain active sites are shown in stick and surface representation, respectively. Active site domains are also labeled (S1-S3, S1’, S3’). The poses shown were obtained using the Induced Fit Docking module of the Schrödinger Maestro computer package. b, d, and f: The predicted receptor-ligand interactions, with hydrogen bonds represented as black dotted lines. Residues predicted to participate in receptor-ligand interactions, as well as the residues of the catalytic triad, are visualized.
Figure 3.2f shows the hydrogen bonds predicted to form between NSC 260594 and MET68, ASN69, and ASP161. Again reminiscent of several known ligands, the S2 subsite is occupied by an aromatic ring system, and a slightly rotated GLU208 accommodates the larger aromatic moiety (1-methyl-6-nitro-decahydroquinolin-4-amine).

In an attempt to identify predicted binding characteristics that might aid the identification of future inhibitors, we used the Induced Fit Docking (IFD) module of the Schrödinger Maestro computer package to re-dock the 30 experimentally tested compounds into a crystallographic active-site model (PBD ID: 1ME4) at pH 5.5. Unlike CDOCKER docking, the IFD protocol allows for local active-site conformational changes and so is judged to better predict ligand binding, albeit at the cost of speed. The most favorable IFD poses of the validated inhibitors (Figure 3.2) consistently placed ligand atoms near the catalytic triad in a position that could conceivably compromise cruzain enzymatic activity. Additionally, the IFD poses of these compounds also place at least one aromatic ring in the S2 subsite, a site known to be favorable for the binding of Phe and Tyr aromatic side chains (138-143, 181, 182).

These predicted binding poses are promising because they represent unique scaffolds that differ significantly from those of previously identified small-molecule inhibitors. While the poses included here are mere predictions, we are hopeful that they will guide future optimization of these experimentally validated ligands.
Conclusions

Chagas disease is the leading cause of heart failure in Latin America, affecting over 10 million individuals worldwide. Although progress has been made in the treatment of the disease, especially given the recent success of K11777, multiple cruzain inhibitors are needed given the difficulty of obtaining FDA approval and ever-progressing drug resistance. The work presented here provides chemical scaffolds that, with further optimization, could be developed into promising therapeutics for Chagas disease. The exhaustive virtual-screening approach and thorough experimental validation used to identify these leads also represents a promising and useful method for inhibitor identification.

Acknowledgments

This work was carried out with funding from the National Institutes of Health (GM31749 to J.A.M.; and GM71630 to Brian Shoichet, UCSF) and the National Science Foundation (MCB-1020765 and MCA93S013 to J.A.M.). Additional support from the Howard Hughes Medical Institute, the NSF Supercomputer Centers, the San Diego Supercomputer Center, the W. M. Keck Foundation, the National Biomedical Computational Resource, and the Center for Theoretical Biological Physics is gratefully acknowledged. Compounds were obtained free of charge from the National Cancer Institute’s Development Therapeutics Program. We also thank Rafaela Ferreira for the expression and purification of the cruzain protein, and Palmer Taylor and Ákos Nimenecz for the use of their FlexStation. All compounds were provided by the NCI/DTP Open Chemical Repository (http://dtp.cancer.gov).
Chapter 3 is a minimally modified reprint of the material as it appears in Rogers, Kathleen E.; Keränen, Henrik; Durrant, Jacob D.; Ratnam, Joseline; Doak, Allison; Arkin, Michelle R.; McCammon, J. Andrew. “Novel cruzain inhibitors for the treatment of Chagas' disease.” *Chemical Biology & Drug Design*. 2012, Sep;80(3):398-405. I was the first author and principal investigator on this paper. Mr. Keränen and Dr. Durrant began the Relaxed Complex Scheme work with molecular dynamics and initial virtual screens. Dr. Ratnam, Ms. Doak, and Professor Arkin provided essential guidance for the enzymatic inhibition assays. Professor McCammon supported this work and advised during the project.

**Supporting Information**

Included in the available supporting information are the chemical structures of the 8 NCI compounds with preliminary IC_{50} values ≤100 µM (Table S1-3.1) and the IC_{50} curves of each under various experimental conditions (Figure S1-3.1). Further description of experimental validation methods and assessment of non-specific inhibition *via* aggregation is included as well.

**Varying Conditions in Enzymatic Assays to Validate Cruzain Inhibition**

The *Trypanosoma cruzi* cysteine protease cruzain was expressed and purified according to reference (157). All compounds were obtained from the NCI/DTP Open Chemical Repository (http://dtp.cancer.gov) and serially diluted in DMSO to create final assay concentrations (FAC) of 1000 µM to 0.01 µM. Using the standard
conditions, the best and second-best scoring compounds had initial IC$_{50}$ values of 471 nM and 15 µM, respectively (Table S1-3.1).

The top 8 compounds that inhibited cruzain at 100 µM were selected for further validation. In all assays, a buffer of 100 mM sodium acetate (NaAc) at pH 5.5 was used, and the initiation of cruzain (0.4 nM FAC) enzymatic activity was measured by adding fluorogenic substrate (2.5 µM FAC) Z-Phe-Arg-amino-methyl-coumarin (Z-FR-AMC) following a five-minute incubation with the novel compound. To investigate detergent-dependent sensitivity, the IC$_{50}$ values of each compound in the presence of two different detergents were determined. First, results using Triton X-100 (0.02%) as the detergent were compared to those obtained without detergent (Figure S1-3.1, left panel) to assess the general effect of detergent on cruzain inhibition. Next, assay results using Triton X-100 (0.02%) and Tween (0.002%) were likewise obtained for those compounds that did not appear to be aggregators (Figure S1-3.1, on the right, last three rows). In order to assess inhibition dependence on the reducing agent, β-mercaptoethanol (BME) was subsequently used in place of DTT in the standard protocol for the four most potent compounds (Figure S1-3.1, right panel).

Cruzain inhibition by each compound was evaluated by measuring the increase in fluorescence (excitation wavelength = 355 nm, emission wavelength of 460 nm) for five minutes in a microtiter plate spectrofluorimeter (Molecular Devices, FlexStation). Percent inhibition was determined from the initial velocities with SoftMax Pro 5.4, and dose response curves created in Prism 4 (GraphPad, San Diego, USA) were used to calculate IC$_{50}$ values. Initial curves (detergent vs. no detergent) were generated by measuring inhibition at seven to eight concentrations of the predicted inhibitor, and
final curves (best of at least two independent experiments of the detergent and reducing agent tests) were derived by measuring inhibition at fourteen to fifteen different concentrations, with the exception of NSC 61610, which included only 8 concentration measurements.

**Dynamic Light Scattering**

DLS experiments, performed according to reference (157), suggested that NSC 61610 was an aggregating compound even at low concentrations, with particles detected at 0.5 µM in both potassium phosphate (KPi) and sodium acetate (NaAc) buffers. AmpC β-lactamase (AmpC) was used to test for compound inhibition via aggregation (as described in reference (186)). NSC 61610 showed 52% AmpC inhibition at 5 µM, 30% inhibition at 1.5 µM, and 14% inhibition at 0.5 µM. Inhibition was reversible by adding 0.01% Triton X. NSC 67436 and NSC 260594 form particles at 15 µM and 40 µM KPi, respectively, but do not significantly scatter light in NaAc (up to 50 µM and 75 µM, respectively). There was no inhibition of AmpC up to 50 µM for either NSC 67436 or NSC 260594. NCS 227186 forms particles at 100 µM in KPi and 20 µM in NaAc.

Table SI-3.1. Preliminary experimental validation results for compounds with IC50 values of 100 µM or lower.

<table>
<thead>
<tr>
<th>Structure</th>
<th>ID</th>
<th>IC50 (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image_url" alt="Structure Image" /></td>
<td>NSC 61610</td>
<td>0.471</td>
</tr>
</tbody>
</table>
Table SI-3.1, Continued

<table>
<thead>
<tr>
<th>Structure</th>
<th>ID</th>
<th>IC&lt;sub&gt;50&lt;/sub&gt; (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Structure Image" /></td>
<td>NSC 67436</td>
<td>15</td>
</tr>
<tr>
<td><img src="image2.png" alt="Structure Image" /></td>
<td>NSC 227186</td>
<td>94.9</td>
</tr>
<tr>
<td><img src="image3.png" alt="Structure Image" /></td>
<td>NSC 19803</td>
<td>98.2</td>
</tr>
<tr>
<td><img src="image4.png" alt="Structure Image" /></td>
<td>NSC 57670</td>
<td>98.3</td>
</tr>
<tr>
<td><img src="image5.png" alt="Structure Image" /></td>
<td>NSC 91529</td>
<td>100</td>
</tr>
</tbody>
</table>
Figure SI-3.1. IC50 curves for selected NSC compounds under various experimental conditions, used to investigate the mechanisms of inhibition and determine aggregation-based activity.
Chapter 4: Computational modeling of TEAD and discovery of YAP-TEAD complex inhibitors

Abstract

Yes-associated protein (YAP) is an integral regulator of the Hippo signaling pathway, mediating cell proliferation and survival. Several reports demonstrate its involvement in the response to cell contact, epithelial to mesenchymal transition (EMT), anchorage-independent growth, inhibition of apoptosis, and growth factor-independent proliferation and survival. As a transcription co-activator, much of YAP’s activity depends on its association with another important transcription factor, TEAD. YAP contains an activation domain and TEAD contains a DNA-binding domain. Together, their normal downstream effectors vary depending on the cellular environment but abnormal YAP activity has been found in several human diseases. Not only is yap a candidate oncogene but both the expression and localization of YAP protein can be found at higher levels in several forms of human cancers. YAP association with TEAD is essential for much of its behavior, including proliferation and EMT, making this complex an interesting therapeutic target. The structure of the regions involved in YAP-TEAD binding along with characterization of the most crucial interface between the two proteins provides an opportunity for structure-based, computer-aided drug discovery (CADD). Although targeting a protein-protein interaction is often challenging, initial studies suggest that this system may be
amenable to small-molecule inhibition uncovered using CADD. This work describes the use of a complemented Relaxed Complex Scheme (RCS) CADD approach that led to a discovery of a YAP-TEAD inhibitor with previously recognized antiproliferative activity.

**Abbreviations**

Yes-associated protein (YAP); epithelial to mesenchymal transition (EMT); Molecular dynamics (MD); Relaxed Complex Scheme (RCS); root mean square deviation (RMSD); root mean square fluctuation (RMSF); Induced Fit Docking (IFD); YAP-binding domain (YBD); Computer-Aided Drug Discovery (CADD).

**Introduction**

Drug resistance and metastasis remain two of the major obstacles in cancer treatment and lethality.\(^{(187, 188)}\) A promising new cancer drug target is the oncogenic Yes-associated protein (YAP), a transcription co-activator essential to organ size regulation and the primary effector of the mammalian Hippo pathway.\(^{(189-192)}\) Under normal conditions, levels of YAP protein are regulated in a cell density-dependent manner.\(^{(193)}\) When there is a low density of cells, YAP affects gene expression in the nucleus to promote proliferation and inhibit apoptosis. Under high cell density conditions, YAP is mostly inactive in the cytoplasm where it gets phosphorylated at several sites and can be ubiquitinated for degradation. In a high cell density yet abnormal environment such as a tumor, YAP often shows increased protein expression and nuclear localization, indicating that it is behaving as if it were in a low cell density environment and turning on various inappropriately timed and
potentially dangerous signals. This elevated expression and nuclear localization of YAP has been reported in various types of human cancers, including colon, lung, ovarian, liver, and prostate cancers. (194-196) The roles of YAP in the response to cell contact, epithelial to mesenchymal transition (EMT), anchorage-independent growth, inhibition of apoptosis, and growth factor-independent proliferation and survival make it an interesting player not only in tumorigenesis but in metastasis as well. (189, 196, 197)

YAP is made up of an N-terminal proline-rich domain, a TEAD-binding region, one or two WW domains depending on the splicing variant (YAP1 or YAP2), an SH3-binding motif, a coiled-coil domain, a transcription activation domain, and a C-terminal PDZ-binding motif. Without a DNA-binding domain, much of YAP’s activity depends on interaction with a key transcription factor target within the nucleus, TEAD. The YAP-TEAD association is important for both physiological and pathological cellular processes, with one particular contact between S94 in YAP and Y406 in TEAD being especially essential for this complex to form effectively. (198) In fact, the missense TEAD1 Y406H mutation is associated with Sveinsson's chorioretinal atrophy, a human genetic autosomal dominant eye disease also referred to as helicoid peripapillary chorioretinal degeneration. (199) This further demonstrates the importance of the TEAD Y406 residue for association with YAP and the significance of this complex for normal development.

Without the YAP-TEAD interaction, most of YAP’s gene targets and roles promoting cell proliferation, oncogenic transformation, and EMT are unaffected. (198) Pharmacological disruption of this protein complex is an exciting concept for future
cancer treatment, especially because of TEAD is largely dispensable for normal tissue growth and homeostasis (200). Already, at least one small molecule has been identified as a proof of principle, suggesting that inhibiting the YAP-TEAD association is a feasible approach against YAP oncogenic activity (200). However, like the design of many protein-protein interaction inhibitors, these efforts will not likely come without challenges. Fortunately, two 3-dimensional structures of the YAP-TEAD interfaces are available, confirming the essential YAP S94 contact with TEAD1 Y406 via a hydrogen bond and providing valuable data for structure-based drug design (201, 202). Although there appears to be a small pocket nearby residue Y406 in these structures, initially the surface does not seem to be particularly promising for small molecule binding.

In order to explore and better characterize the target interface where YAP binds for future structure-based drug discovery, the computational approach referred to as the Relaxed Complex Scheme (RCS) (170) was implemented on a truncated human TEAD1 monomer model based on the incomplete crystal structure [PDB ID: 3KYS] (202) of the YAP-TEAD complex with missing loops filled in by SWISS-MODEL (203). Using molecular dynamics to uncover several other possible TEAD conformations to include in virtual screening efforts against a diverse compound library further suggests that small molecule YAP-TEAD inhibition may be viable. Additionally, the work presented herein investigates interference of YAP-TEAD association by a previously recognized anti-proliferative agent.
**Results and Discussion**

The YAP-TEAD interaction is an interesting therapeutic target for inhibition, yet the solved structure of the interfaces in this complex makes structure-based drug design near the most crucial contact, TEAD residue Y406, appear only slightly encouraging due to the lack of a well-defined, deep pocket there. In hopes of revealing other pockets on this TEAD surface that did not form in the solved structure but may be more promising for small molecule inhibition of YAP, the molecular dynamics of a truncated TEAD model structure alone were simulated computationally. Although the YAP-binding domain (YBD) of the solved unbound TEAD structure [PDB ID: 3L15](204) is similar to that bound to YAP [PDB IDs: 3KYS(202), 3JUA(201)] and therefore is believed to be somewhat rigid, there are some unsolved, disordered regions nearby suggesting that there may be some intriguing flexibility in solution.

**Protein Modeling and Virtual Screening**

The SWISS-MODEL Workspace (205) was used to create a more complete model of human TEAD1 by filling in flexible loops missing from the solved YAP (residues 50-171)-TEAD1 (residues 194-411) three-dimensional structure [PDB ID: 3KYS] (Figure 4.1a). One of these loops was near the YAP-TEAD interface 3, which contains the contacts most crucial to complex recognition, and formed an attractive pocket (Figure 4.1b). Without the bound YAP segment, the TEAD protein is allowed to simulate more freely in solution. To explore this flexibility, 60 ns of molecular dynamics (MD) were performed with NAMD 2.7 (206) on the solvated TEAD model. Not surprisingly, four of the residues filling in the loop for missing residues 230-235
were among the top 12% of residues with the highest RMSF (Supporting Information, Figure SI-4.1). There were another four residues within or nearby that cavity formed with this loop that were in the top 7% (Figure SI-4.2). The smaller pocket near Y406 was also noticeably flexible and took on more distinct shapes during some of the MD simulation (Figure 4.2), suggesting that the surface there may be able to accommodate small molecule docking better than the solved structure had hinted.

Figure 4.1. New pocket formed in TEAD model near edge of YAP segment. (A) The TEAD solved structure (surface representation) is missing several residues, including 230-235, which are near the edge of an important interface where a YAP segment (cartoon representation) binds. (B) Filling in the loop here (violet surface) for the TEAD model helps enclose a cavity, left of the Y406 residue (yellow surface) that is key to YAP binding.

The trajectory was then clustered into 39 different conformations (average RMSD of 0.78Å) based on the uniqueness of the YAP-binding domain (YBD) surface configurations of interface 3 near the Y406 residue crucial for YAP binding. This work helped to identify some flexibility and potential new opportunities for drug design within the YAP-binding area. However, it was not surprising to find a relatively stable surface here because previous work shows that the YBD TEAD structure is largely unchanged if bound or unbound to YAP. (202) Nonetheless, the
TEAD YBD surface near Y406 did adopt a variety of pocket, nook, or tunnel-like forms during the 60 ns simulation (Figure 4.2) that were not apparent in the original structure (Figure 4.1). This hints that some small molecules may be able to bind there more favorably and block YAP binding more efficiently than initially thought, especially if an induced-fit type behavior is possible.

Figure 4.2 Changes in TEAD YBD surface during MD suggest flexibility that may allow for and adapt to small molecule docking. Conformations adopted by the TEAD YBD surface near Y406 during the simulation show a variety of more promising forms. (A-B) Central member from the 9th, (C-D) 19th, and (E-F) 36th largest cluster shown in surface representation as examples. Y406 is yellow surface in all and labeled in left column. YAP shown as yellow cartoon (B, D, F only).
In addition to providing snapshots representing some of the protein’s flexibility, the ensemble of clustered conformations from the MD simulations were also used to dock compound libraries to the YBD of TEAD computationally. Initially, the docking was centered around residue TEAD Y406 using AutoDock Vina (207) to find compounds that may bind closest to the area essential for YAP binding. A small segment of the crystal structure of YAP was docked as a control to assess the ability of AutoDock Vina to recreate the solved complex structure. The top pose of this pseudo ligand produced with AutoDock Vina captured YAP F95 close to its PDB position, with other residues (including S94) only slightly shifted (Figure 4.3). Although small molecules in the NCI Diversity Set II library were able to dock to the surface here (Figure SI-3a-b), many of their predicted binding energies were weak enough to also consider additional pockets on the TEAD YBD that may be more favorable for small molecule docking.

Following the Autodock Vina screen, the Schrödinger Glide (208) Virtual Screening Workflow protocol was used with a slightly less conservative docking box, which included the side pocket closer to the flexible loop at TEAD residues 230-235 but was still centered around TEAD Y406. A few of the top scoring compounds showed a minor preference for this rather deep pocket represented in the TEAD model structure (Figure 4.1b) or for smaller pockets formed in some of the trajectory conformations (e.g. see Figure 4.2a showing two: a more modest cavity left of Y406 and another near the end of YAP, included in Figure 4.2b). Intriguingly, because there is a TEA (Transcription Enhancer Activation) DNA-binding domain before the YBD,
the flexibility here may be important to the transcriptional role of TEAD. However, this pocket is also close enough to YAP within interface 3 that small molecules binding there may allow for a disruption in YAP-TEAD formation as well.

Figure 4.3. Reproduced pose of pseudo-ligand YAP segment bound to TEAD. A small segment of YAP (pink sticks), residues L91-F95, docked similarly compared to the crystal structure of the YAP TEAD binding domain (yellow sticks). TEAD model shown as green surface except for yellow Y406.

A small set of compounds was chosen for experimental testing from both screens based on predicted binding energy scores and docking poses, with YAP aligned and superimposed to show proximity. Many of the top predicted-binding scores came from poses of compounds docked partially or completely within a pocket in the flexible region near TEAD residues 230-235 (FigureSI-3, c-d). Hence, some of the compounds that were first ordered and tested for experimental validation were predicted to dock near the edge of the YAP segment from the solved YAP-TEAD complex structure. However, there were also some other high scores where the compound bound near or over TEAD Y406 (Figure SI-3, a-b).
Experimental Activity of Virtual Screen Hits

Initially, the top ten compounds were investigated for their effect on YAP and its association with TEAD. A luciferase reporter assay was used to assess the dose-dependent inhibitory activity of the compounds selected. Three of these first ten compounds appeared to inhibit YAP-TEAD4 reporter activity in a dose-dependent manner. (Figure 4.4) Next, these three compounds were further tested for their effects on YAP and phosphorylation of YAP using co-transfection of TEAD and YAP, followed by immunoprecipitation and immunoblotting. In Figure 4.5, we see that NSC 105827 inhibits YAP-TEAD4 interaction both by treating cells directly and in the post-lysis lysate only (i.e. not before IP), which could suggest a possible mechanism for inhibition by effectively blocking YAP binding. Interestingly though, this effect is not nearly as pronounced with the TEAD1 isoform in preliminary results (data not shown).

Figure 4.4. Inhibition of YAP-TEAD4 luciferase activity by three NCI Diversity Set II compounds indicate effect on YAP activity. A dose-dependent decrease in YAP reporter activity can be seen by three of the first ten compounds tested. Luciferase activity normalized to Renilla.

Phosphorylation of YAP by Lats upstream via the Hippo Pathway, especially at S127 (although four serine residues in other HXRXXS motifs may also be
phosphorylated: S61, S109, S164, and S381, (193)), results in 14-3-3 cytoplasmic retention and hence inhibits YAP activity in a reversible way (196, 209). Interestingly, phosphorylation of S381 is believed to lead to YAP degradation in an irreversible way, suggesting that YAP activity is regulated both spatially and temporally (193) (See Figure 5.4). Because YAP is implicated in many types of cancer, standard dose treatment and immunoblotting studies were performed on YAP as well as YAP phosphorylated (pYAP) at the major phosphorylation site, S127, in several cancer cell lines (Figure 4.6).

Figure 4.5. NSC 105827 decreases YAP-TEAD4 interaction both by treatment of cells directly or in the lysate. NSC 81750 and NSC 134159 do not appear to inhibit YAP binding in Co-IP experiments but NSC 105827 looks to decrease YAP in a dose-dependent manner. Experimental design left, results of western blots right.

We include pYAP in the western blots to investigate (I) whether these small molecules might inhibit YAP via targets upstream of Hippo pathway somewhere by
increasing pYAP and (II) whether both YAP and pYAP decrease as the total amount of YAP is degraded (e.g. via apoptosis, perhaps due to inhibited YAP activity). These results demonstrate potentially cell- or context-specific effects on YAP and phosphorylated YAP (pYAP) but, because the stability (half-life) of pYAP127 vs. YAP proteins and the sensitivity of their antibodies are very different, M14 Melanoma cells may just be showing a result of a longer YAP half-life after NSC 105827 treatment compared to others. However, again, it appears that NSC 105827 has the most pronounced effect on YAP in these cell lines, particularly in SF268 Glioma cells and SK-MEL-28 cells. Although we cannot draw too many conclusions from these results, NSC 105827 does appear to be decreasing the amount of phosphorylated YAP in all cell lines and not increasing this amount in any case, indicating that this compound does not inhibit YAP upstream by increasing pYAP and cytoplasmic sequestration.

![Figure 4.6](image)

Figure 4.6. Effect of NCI compounds on protein level of YAP and phosphorylated YAP in various cells lines. NSC 81750 and NSC 134159 both appear to slightly inhibit pYAP in at least one cell type but NSC 105827 appears to be the most effective overall. α-Tubulin shown as loading control.

Although YAP-binding appears to be inhibited even when treatment was done to the lysate (Figure 4.5) and there does not seem to be an increase in pYAP (Figure 4.6) after NSC 105827 treatment, both of which might otherwise indicate inhibition of
YAP upstream in the Hippo Pathway or via some other mechanism, it cannot be confirmed that this compound is inhibiting YAP directly by binding to TEAD from the data presented here. More work is especially necessary to determine the specific mechanism of YAP inhibition. However, using Induced Fit Docking (IFD)(210), a more computationally expensive docking protocol, with knowledge from the faster Autodock Vina RCS screen provided encouraging predicted poses as described below.

The IFD protocol was used in an ensemble-enriched screen, based on the outcome of prior docking efforts. Using Schrödinger IFD to dock thiosangivamycin to a few ensemble members shown in prior screens to accommodate compounds near Y406 suggested that this small molecule could bind in ways that may sufficiently disrupt YAP binding. For example, NSC 105827 is able to (I) form a hydrogen bond directly with Y406 that normally forms with the essential YAP S94 or (II) dock right on top of Y406 with the pyrrole poised for aromatic stacking and form numerous hydrogen bonds with surrounding TEAD residues (Figure 4.7). This method, using IFD with MD ensemble members already known to bind small molecules preferably, combines receptor flexibility sampled during MD with smaller adjustments needed to accommodate small molecule docking without being overly computationally expensive.
Figure 4.7. Induced Fit Docking Pose of NSC105827 with an MD Ensemble Member. 

NSC 105827 screened with an ensemble member (that performed well overall in the Vina 
screen) using IFD shown here lays right on top of Y406. (A) TEAD residues shown as 
green lines, except Y406 in yellow. NSC 105827 (green sticks) appears poised for 
aromatic stacking, possibly blocking YAP binding at this critical residue. (B) TEAD 
surface near Y406 (yellow) seems to fit NSC 105827 nicely, with the ligand docked to 
potentially interfere with YAP (yellow cartoon) binding. Possible hydrogen bonds shown 
as black dotted lines.

Conclusions

There are still many unanswered questions about the precise process and 
consequences of YAP-TEAD co-activation but disruption of this complex is an 
established area of therapeutic interest for many reasons.\textsuperscript{(192, 194, 196)} The 
combined Autodock Vina and Glide XP RCS Computer Aided Drug Discovery 
approach described above proved useful in identifying compounds with inhibitory 
effects on YAP-TEAD. The induced-fit docking pose and the results presented here 
examining the effect of Thio-sangivamycin (NSC105827) on the YAP-TEAD 
interaction are encouraging but cannot sufficiently confirm that its inhibitory activity 
is not a result of off-target, metabolic effects. This previously discovered small 
molecule should be investigated further, especially within the context of the Hippo
Pathway, to determine the mechanism. Additional computational drug discovery efforts using much larger compound libraries to uncover more specific and therapeutically favorable TEAD binders are on-going and preliminary work so far is promising (Figure SI-4.4).

Materials and Methods

TEAD Computational Model System Preparation

The SWISS-MODEL (203) workspace was used to fill in and build loops missing from a truncated TEAD monomer (TEAD1 residues 194-411) of the 3KYS PDB dimeric YAP-TEAD structure. This TEAD model, representing the YAP-Binding Domain (YBD), was further prepared at pH 7 using the AMBER forcefield with the PDB2PQR software package, which includes protonation and pKa calculations performed by PROPKA (148, 149). To prepare the system for MD, the crystallographic waters were then added back before further parameterizing and solvating the protein in a TIP3P (103) water box (extending 10 Å out from protein on all ends) with neutralizing sodium counter ions using the LEAP module of AMBER 10 (211).

Molecular Dynamics (MD) Simulations of the TEAD YBD

A four-stage, gradually more inclusive minimization process was implemented on the system using NAMD 2.7 (206). The first stage allows only the hydrogen atoms to move for 5000 steps. The next allows the hydrogen atoms, waters, and counter ions to minimize for the same number of steps. The third stage also allows the protein
backbone to move, in addition to the previously included atoms, for 1000 steps. The final stage allows all atoms to move in a 2500 step energy minimization.

Next, the system was equilibrated in a five-stage process starting from the final minimization step. A series of four 250-ps simulations with progressively lower harmonic restraint forces (constraint scaling factors of 1, 0.75, 0.5, 0.25, respectively) were completed using NAMD 2.7, each at 310K. Then a 4-ns equilibration without any constraints was performed to allow the system to move freely before continuing to the production MD simulations.

The equilibrated system was then used as the input for 60 ns of production MD, which was maintained at 310K with Langevin dynamics and at 1 atm (1.01325 bar) with Nosé-Hoover Langevin piston pressure control in NAMD 2.7 (206). The time constants for the barostat oscillation and damping (i.e. period and decay) used were 100-fs and 50-fs, respectively. The Particle Mesh Ewald electrostatics method was implemented and full electrostatics were evaluated every two time steps. The same frequency was used for calculating short-range non-bonded interactions, with a cutoff distance defined at 14 Å, but bonded interactions were evaluated at each 2-fs time step. The ShakeH algorithm (212) was implemented on all waters, keeping the bond lengths and angles rigid with a tolerance of 0.0005Å.

TEAD Conformational Clustering and Docking

To investigate any major structural changes seen along the YBD during the simulation, the 30,000 TEAD MD trajectory “frames” (500 frames representing each ns) were aligned and clustered based on the RMSD of residues in the region near the
Y406 TEAD residue (i.e. YAP-TEAD interface 3). Using an RMSD cutoff of 0.7Å, gromos clustering with Gromacs 4.0.4 (213) identified 39 representative structures from the MD simulation. The central member of each cluster was used for individual visualization of any interesting changes near Y406 and collectively as an ensemble for subsequent docking. The Root Mean Square Fluctuation (RMSF) was calculated for each residue in the VMD (125) tcl console.

The virtual screens presented in this work were performed using Autodock Vina (207) and the modules within the Schrödinger Suite. First, a short segment of YAP (residues L91-F95) was docked to the TEAD model using Autodock Vina as a “pseudo ligand” control. The Phe95 residue of this YAP segment docked very close to its crystallographic position (Figure 4.3) and the other residues, including Pro92, were only slightly shifted from their solved structure position. LigPrep (214) was used with the OPLS_2005 forcefield (215) to generate 3D molecular models of the compounds within the NCI Diversity set II. For the RCS (170) Autodock Vina screen of this library, all 39 ensemble members were used to include a representative conformation from each cluster and account for more TEAD flexibility. The docking box was centered on TEAD residue Y406 with X,Y, Z dimensions 21.73 Å, 14.85 Å, 17.37 Å, respectively. The docking scores were ranked by an ensemble-average and an ensemble-best scoring scheme for each compound as previously described (174). Additionally, the average overall library score of each ensemble member was calculated to estimate the adaptability and potential “druggability” of the conformation represented by each cluster near Y406. This was used to prioritize docking pose refinement later on in the study using Schrödinger Induced Fit Docking (IFD) (163).
For the Schrödinger Glide screen, protein grids for each of the ensemble MD structures were prepared with the Protein Preparation Wizard (161). An inner docking box $14\AA^3$ in volume was generated for each, within which the center of the ligand must bind. The box was centered on a short YAP segment bound near TEAD Y406 from the solved complex structure, which was used as a pseudo-ligand here as well. The optimized compound library was then docked to each RCS ensemble member using the Schrödinger virtual screening workflow protocol with the Glide (208) module. In the progressively more advanced binding energy calculation and exhaustive compound conformational sampling steps of Glide, the filters used were relaxed to allow further characterization of more potential hits. The top 50% of compounds were allowed to pass through both the Glide HTVS (high-throughput virtual screening) and the SP (standard precision) stages, and 100% of the compounds were scored in the XP (extra precision) stage so that there would be both a score and pose to reference even the weaker bound ligands that made it through to that final phase. Otherwise, the default parameters were used for this screen.

Compounds were selected for experimental validation based on predicted poses along the YBD, ensemble average scores in the Autodock Vina screen, and those that scored in the top 5% of Glide XP scores for at least 2 ensemble members.

All computational modeling figures were generated using Pymol (165).

**Cell Culture**

All cell lines were acquired from American Tissue Culture Collection (ATCC) at 37C with 5% CO2 and maintained at medium composition according ATCC.
Luciferase Reporter Assay

The luciferase reporter assay was performed using HEK293 BOCs cells. 5× UAS-luciferase reporter, Renilla, YAP2, and TEAD4 plasmids were co-transfected as described previously. (198) After 48 hours of transfection, concentrations of 100nM to 10μM were used and luciferase reporter activity was normalized to Renilla for the dose response curve.

YAP/TEAD co-transfection, immunoprecipitation and immunoblotting

Cells were transfected with YAP2 and TEAD4 DNA plasmids according to procedure described previously (12) and lysed using mild lysis buffer. Cell lysates were centrifuged for 10 min at 4°C, and supernatants were used for 16 hour treatment with compounds and immunoprecipitation. Immunoprecipitates were washed four times with lysis buffer, and proteins were eluted with SDS-PAGE sample buffer. Immunoblotting with Flag, Myc, Parp, and pYAP127 antibodies from Cell Signaling or YAP and α-Tubulin from Santa Cruz Biotechnology were performed using standard protocol according to manufacturers’ instruction.

All experiments described in Chapter 4 and data presented in Figures 4.4-4.6 come from collaboration with the Guan Lab, specifically Dr. Ian Lian.

Acknowledgments

This work was carried out with funding from the National Institutes of Health and the National Science Foundation. Additional support from the Howard Hughes Medical Institute, the NSF Supercomputer Centers, the San Diego Supercomputer Center, the W. M. Keck Foundation, the National Biomedical Computational
Resource, and the Center for Theoretical Biological Physics is gratefully acknowledged. Compounds were obtained free of charge from the National Cancer Institute’s Development Therapeutics Program.

Chapter 4 contains text from the article in preparation Rogers, Kathleen E.; Lian, Ian; Guan, Kun-Liang; McCammon, J. Andrew. “Computational Drug Discovery of TEAD-YAP inhibitors” (In preparation for Chemical Biology & Drug Design). I am the first author and principal investigator on this paper. Dr. Lian provided the experimental results while Professor Guan and Professor McCammon supported the collaborative work.

Supporting Information

Figure SI-4.1. Root-mean-square fluctuations (RMSF) of α-C atoms during TEAD Trajectory. Some of the highest fluctuations of α-C atoms are seen near residues 230-235. This region became an interesting focus in the dynamics and virtual screening studies due to pocket-forming surface flexibility there.
Figure SI-4.2. Dynamic Residues Near Modeled Pocket. (A) Residues 230-235 (purple), many of which also have high RMSF, are near the end of the YAP (yellow) PDB segment at interface 3 and form part of the pocket filled in for the TEAD model (green). (B) Rotated view of (A).

Figure SI-4.3. NCI Compounds Appear to Dock in Various Ways to TEAD. Many of the top hits from the Autodock Vina (A-B) and Glide (C-D) RCS screens of the NCI Diversity Set II library to the TEAD model (C) and ensemble members (A, B, D) are noticeably unique. TEAD Y406 shown as yellow surface. YAP (yellow cartoon) is included and aligned to show proximity to ligands.
Figure SI-4.4. Additional NCI Compounds Show Preliminary Inhibitory Effects on TEAD-YAP Association. Compounds labeled here as A and H appear to inhibit TEAD-YAP binding in both TEAD and YAP pull-down experiments. The Guan lab continues to further validate these and many other CADD-identified potential hits.
Chapter 5: Conclusions

Recent advances in DNA sequencing technology and its applications are making personalized medicine a reality (216). Soon, Next Generation Sequencing (NGS) may become routine in patient diagnosis of diseases, disorders, and infections. NGS will uncover new candidate disease-causing genetic mutations, human, viral or parasitic gene modifications that allow for drug resistance, genetic profiles that help dictate variations in patient pharmacokinetics and pharmacodynamics, predicted side effects and toxicity, and many other pharmacogenomic or metabolite markers. These discoveries will not only help identify new drug targets and drug metabolism enzymes, they will also help diagnose patients and inform clinicians to prescribe the best method of prevention or course of treatment for each individual based on predicted benefits and expected toxicities (217-219). Although many challenges and cautions remain, stratified medicine holds the power to innovate healthcare, making it more effective, cost-efficient, and safe. (220-222)

One of the fields that will benefit most from genomic-medicine is oncology (223). The chances of developing invasive cancer(s) between birth and death is 1 in 3 for females and 1 in 2 for males (224). Additionally, infectious disease control has come a long way in the last century yet it remains a significant public health concern as well because of globalization, evolutionary dynamics and progression of drug-resistant parasites or viruses, and developing countries still facing a huge burden caused by these communicable diseases (225, 226). Even in the United States, though
the number of people living with HIV now has steadied (Figure 5.1), advances in HIV treatment and an expanded arsenal of available antivirals are needed because of the ongoing threat of drug-resistance and the likelihood that this will be one of the more prevalent chronic infectious diseases for at least the next few decades (227). Cytotoxic chemotherapy is still used today for many of these infectious diseases and cancers but translational genomics and targeted therapy have the potential to decrease the necessity and use of these therapeutics over more tolerable and effective choices. The myriad side effects (7-14) and sequelae of radiation and current chemotherapeutic options (228-231) require innovative approaches, both in diagnostics and treatment, to alleviate the burden of these diseases.

The work presented here addresses the necessary expansion of therapeutic alternatives for patients whose only option currently is cytotoxic chemotherapy and/or may suffer from a drug-resistant infection or form of cancer. Several avenues are being explored, some already showing initial or clinically validated success. For example, a promising molecular carrier for chemotherapeutic delivery, cucurbit[7]uril (CB[7]), helps improve drug stability, solubility, and selectivity (198) and even overcomes development of drug resistance when used to carry chemotherapy drugs such as cisplatin via a purely pharmacokinetic effect (52-55). We used this CB[7] and a neutral, bicyclo[2.2.2]octane derivative (B2) as a host-guest model of molecular recognition for free energy estimation. With explicit solvent molecular dynamics (MD) simulations and Thermodynamic Integration (TI), we were able to estimate the absolute free energy of host-guest binding rather well (Figure 5.2). We further investigated host wetting and dewetting events as well as average water density surfaces using two different water models, TIP3P and TIP4P. During the van der Waals decoupling step of the bound-guest, certain λ values (around 0.8) experience significant changes in $\frac{\partial V}{\partial \lambda}$ during the MD, which correspond to rapid fluctuations in the number of TIP3P or TIP4P waters that are able to fill the host in place of the disappearing guest. We report a decrease in $\frac{\partial V}{\partial \lambda}$ values with increasing water molecules inside the host during van der Waals simulations with sufficient guest decoupling in both water models. However, there were some differences in statistical quantities and average water density surfaces (Figure 5.2) between the two during the 50 ns of MD run at each lambda. These results highlight the large influence of
wetting/dewetting transitions on absolute binding free energy estimation for this host, which is exposed to bulk water at two large symmetric openings, and how they might also impede convergence in similar studies of protein-ligand systems, but likely to a lesser extent. Additionally, this method proved suitable for estimating absolute binding free energy of a small molecule to this intriguing drug delivery system and hence may be appropriate to use for other drugs or compounds bound to this host.

![CB[7] - B2](image)

Figure 5.2. Absolute Free Energy of Host-Guest Binding. Pictured are the host and host-guest systems with the average water density surfaces (TIP3P and TIP4P water models) shown for both states. The wetting and dewetting events during Double Decoupling/Thermodynamic Integration MD played an important role in simulation convergence.

Although worth the effort for lead optimization, these types of accurate and rigorous free energy calculations are usually too time-consuming and computationally expensive to use for drug design before higher-throughput virtual screens, experimental validation of top hits, and some analysis of structure–activity relationships (SAR) are performed. The computer-aided drug discovery work and validation described herein help identify starting points for future design of chemotherapy alternatives. First, we identified three new inhibitors (see Figure 5.3 for
pose of top hit) of *Trypanosoma cruzi* cruzain using a complex virtual-screening methodology that takes advantage of consensus scoring and accounts for receptor flexibility. *T. cruzi*, the etiological agent of Chagas disease, affects millions of individuals, especially in Latin American where it is the leading cause of cardiovascular death among patients between the ages of 30 and 50 years (232). Cruzain is a major cysteine proteinase essential for all stages of the parasitic life cycle and is thought to help the parasite evade an immune response from the host. The exhaustive virtual-screening approach and thorough experimental validation used to identify and verify these cruzain hits at the bench as non-aggregating, non-redox cycling compounds also represent a promising and useful method for inhibitor identification. Lead optimization guided by results from this work could contribute to development of acceptable treatment of Chagas disease by targeting *T. cruzi* cruzain specifically.

![Figure 5.3. Induced Fit Docking pose of Cruzain with top lead, NSC 227186, which had an IC\textsubscript{50} value of 16.6 \textmu M in enzymatic inhibition assays described in Chapter 3.](attachment:image)
We also approached computer-aided drug discovery of TEAD-YAP inhibitors. The Yes-associated protein (YAP) is an encouraging new cancer drug target as it is involved in tumorigenesis and even metastasis. As a transcriptional co-activator, most of YAP’s roles are dependent on TEAD, a key transcription factor (See Figure 5.4 for model and (233) for further review of Hippo Pathway). These two proteins share three interfaces, with the third being most important. Although designing protein-protein inhibitors is the “road less traveled” (234) for good reason, this particular system seemed more amenable to small molecule drug design due to the importance of one specific residue interaction for complex formation. The TEAD surface within interface 3, especially near a critical contact between TEAD Y406 and YAP S94, seems reasonable enough for small molecule drug design. Hence, we performed a series of virtual screening and pose predictions on the TEAD structure there with some experimental success as well. Hit validation from related screens using significantly larger libraries is ongoing and we hope to uncover several additional leads for TEAD-YAP inhibitors.

We hope the work presented here will encourage further studies of these systems and drug targets to improve upon current therapeutic options, particularly for cancer and Chagas disease. Drug delivery systems such as CB[7] may be used to decrease side effects and increase bioavailability of otherwise inefficient drugs and can help overcome drug resistance to chemotherapeutics agents like cisplatin (56). New targets allow us to uncover specific drug leads for development of acceptable
treatment for Chagas disease and tolerable, potentially more effective, cancer therapy. Insights into the computational approaches taken and results reported from them are also of general interest to guide future related research.

Figure 5.4. A model of the Hippo pathway and spatial/temporal regulation of YAP activity via phosphorylation by Lats.
References


8. Wardill HR, Bowen JM, Gibson RJ (2012) Chemotherapy-induced gut toxicity: are alterations to intestinal tight junctions pivotal? *Cancer Chemother Pharmacol*; **70**: 627-35.


112. Fajer M, Baron R, de Oliveira CAF, Rogers K, Ortiz-Sanchez JM, Wereszczynski J, et al. (2013) Absolute binding free energies from thermodynamic
integration in AMBER: Incorporating restraints into the double decoupling method. *in preparation:*


135. Tomas AM, Miles MA, Kelly JM (1997) Overexpression of cruzipain, the major cysteine proteinase of Trypanosoma cruzi, is associated with enhanced metacyclogenesis. Eur J Biochem;244: 596-603.


146. LigPrep, version 2.5, Schrödinger, LLC, New York, NY, 2011.


165. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.


