UNIVERSITY OF CALIFORNIA, SAN DIEGO

Leveraging evolutionary information to guide structure-based drug design

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

in

Biomedical Sciences

by

Aaron Jacob Friedman

Committee in charge:

       Professor J. Andrew McCammon, Chair
       Professor Philip E. Bourne
       Professor Michael K. Gilson
       Professor Vivian Hook
       Professor Christina Sigurdson

2013

The Dissertation of Aaron Jacob Friedman is approved, and it is acceptable in quality

and form for publication on microfilm and electronically:

_____

_____

_____

_____
                                                            Chair

University of California, San Diego

2013

Dedication

This work is dedicated to my grandfather, Poppa Ben, who was the first person to truly inspire me to pursue medical research. We had many fascinating conversations when I was younger, and I still remember them to this day. He shared his love of learning with me and taught me so much about science, and I will always cherish that.

Table of Contents

List of Figures

List of Tables

List of Schemes

Acknowledgements

I would like to first thank Professor Andy McCammon for excellent support and guidance during my graduate studies. I also wish to extend extreme thanks to Dr. Jacob Durrant, who mentored me in my early graduate years and taught me the elegance of simplicity. He has been an incredible resource in both the *Tb*GalE and CrystalDock projects. Additionally, I also must thank Dr. Bill Sinko, Dr. Levi Pierce, Dr. Kathleen Rogers, Dr. Paul Gasper, Patrick Blachly, and the rest of the McCammon lab for all of their help and advice during the past several years. I wish to thank Drs. David Timson and Thomas McCorvie for their collaboration on the *Tb*GalE project. Bailey Miller, Dr. Hyukjae Choi, Professor Bill Gerwick, and Professor Vivian Hook were critical to the cathepsin natural product inhibitors project. I would like to thank Professors Philip Bourne, Michael Gilson, and Christina Sigurdson for many helpful discussions along the way.

Chapter 2 is a minimally modified reprint of the material appearing in Friedman AJ, Durrant JD, Pierce LCT, McCorvie TJ, Timson DJ, McCammon JA. The Molecular Dynamics of *Trypanosoma brucei* UDP-Galactose 4'-Epimerase: A Drug Target for African Sleeping Sickness. *Chem Biol Drug Des* 2012, 80, 173-181. I was the first author and co-principal investigator of this paper.

Chapter 3 is a minimally modified reprint of the material appearing in Durrant JD, Friedman AJ, and McCammon JA. CrystalDock: A Novel Approach to Fragment-Based Drug Design. *J Chem Inf Model* 2011, 51(10), 2573-2580. I was the second author of this paper. Additional work in this chapter will be included in an upcoming paper in which I will be first author.

Chapter 4 is a minimally modified reprint of the material appearing in Miller B, Friedman AJ, Choi H, Hogan J, McCammon JA, Hook V, and Gerwick WH. The Marine Cyanobacterial Metabolite Gallinamide A is a Potent and Selective Inhibitor of Human Cathepsin L. Submitted. I was the second author, and principal computational author, on this paper.

Vita

2008   Bachelor of Science, Washington University in St. Louis

2011   Teaching Assistant, Department of Biology, University of California, San Diego

2013   Doctor of Philosophy, University of California, San Diego

1. Klingenberg JM, McFarland KL, **Friedman AJ**, Boyce ST, Aronow BJ, Supp DM. Engineered human skin substitutes undergo large-scale genomic reprogramming and normal skin-like maturation after transplantation to athymic mice. J Invest Dermatol. **2010** Feb;130(2):587-601
2. **Friedman AJ**, Torkamani A, Verkhivker G, Schork NJ. From Coding Variant to Structure and Function Insight. Protein Structure. **2011**: 93-103
3. Durrant JD, **Friedman AJ**, Rogers KE, McCammon JA. CrystalDock: a novel approach to fragment-based drug design. J Chem Inf Model. **2011** Oct 24;51(10): 2573-80.
4. **Friedman AJ**, Durrant JD, Pierce LC, McCorvie TJ, Timson DJ, McCammon JA. The Molecular Dynamics of Trypanosoma brucei UDP-galactose 4'-epimerase: a drug target for African Sleeping Sickness. Chem Biol Drug Des. **2012** Aug;80(2):173-81.
5. Durrant JD, **Friedman AJ**, Rogers KE, McCammon JA. Comparing Neural-Network Scoring Functions and the State of the Art: Applications to Common Library Screening. J Chem Inf Model. **2013** Jul 22;53(7):1726-35
6. Miller B, **Friedman AJ**, Choi H, Hogan J, McCammon JA, Hook V, Gerwick WH. The Marine Cyanobacterial Metabolite A is a Potent and Selective Inhibitor of Human Cathepsin L. Submitted.
7. Durrant JD, **Friedman AJ**, McCammon JA. PyMolecule: A Python Framework for Manipulating Biomolecules. In preparation.
8. **Friedman AJ**, Durrant JD, McCammon JA. CrystalDock2: Leveraging Known Structural Information for Fragment-Based Drug Design. In preparation.

ABSTRACT OF THE DISSERTATION

Leveraging evolutionary information to guide structure-based drug design

by

Aaron Jacob Friedman

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2013

Professor J. Andrew McCammon, Chair

The accurate prediction of biomolecular recognition is fundamental to modern drug discovery. Computational chemists seek to optimize interactions between proteins and drug candidates. As nature has optimized these interactions over billions of years, fundamental understanding of evolutionarily conserved interactions is important for the design of selective, high affinity treatments.

Herein, several projects are described that leverage such understanding. Identification of conserved residues in drug targets can suggest residues more resistant to mutation, and therefore more attractive for drug discovery. Similarly, analysis of slight differences in protein subpockets can be leveraged to identify regions to target to improve selectivity between related proteins. Examining interactions between functional moieties on the ligand and receptor microenvironments they interact with can identify evolutionarily conserved enthalpic contributions.

Chapter 1: Leveraging evolution can enhance computer-aided drug discovery efforts.

Abstract

Ever since Darwin published his theory of natural selection, it has been repeatedly validated that selective pressures on a species can dictate dominant phenotypes. Mutations resulting in increased fitness for organisms are more likely to persist and spread over time. Natural selection and evolution pervades all areas of biology, from the creation of novel genes and proteins to the evasion of drugs by pathogens. Selective pressures exerted by therapies cause resistant phenotypes to become dominant. Consequently, understanding how and why organisms evolve is fundamental to drug discovery efforts.

Introduction

Over the past several billion years, Earth has accumulated an incredible amount of genetic diversity. Through a diverse array of mechanisms, genomes diversify, leading to the emergence of new genes, and eventually species (Scheme 1.1). Based on selective pressures exerted by the surrounding environment, those organisms best equipped to survive and proliferate have their specific alleles become dominant. If no dominant gene subtype produces protein more apt to survive, multiple alleles for a single gene can arise. These genes may duplicate or evolve separately, which can increase the rate of evolution.

a Exon or domain shuffling

b Gene duplication

c Retrotransposition

d TE domestication

e Lateral gene transfer

f Gene fission or fusion

g *De novo* origination

h Reading-frame shift

i Alternative splicing

j Non-coding RNA

k Pseudogene as RNA regulator

Nature Reviews | Genetics
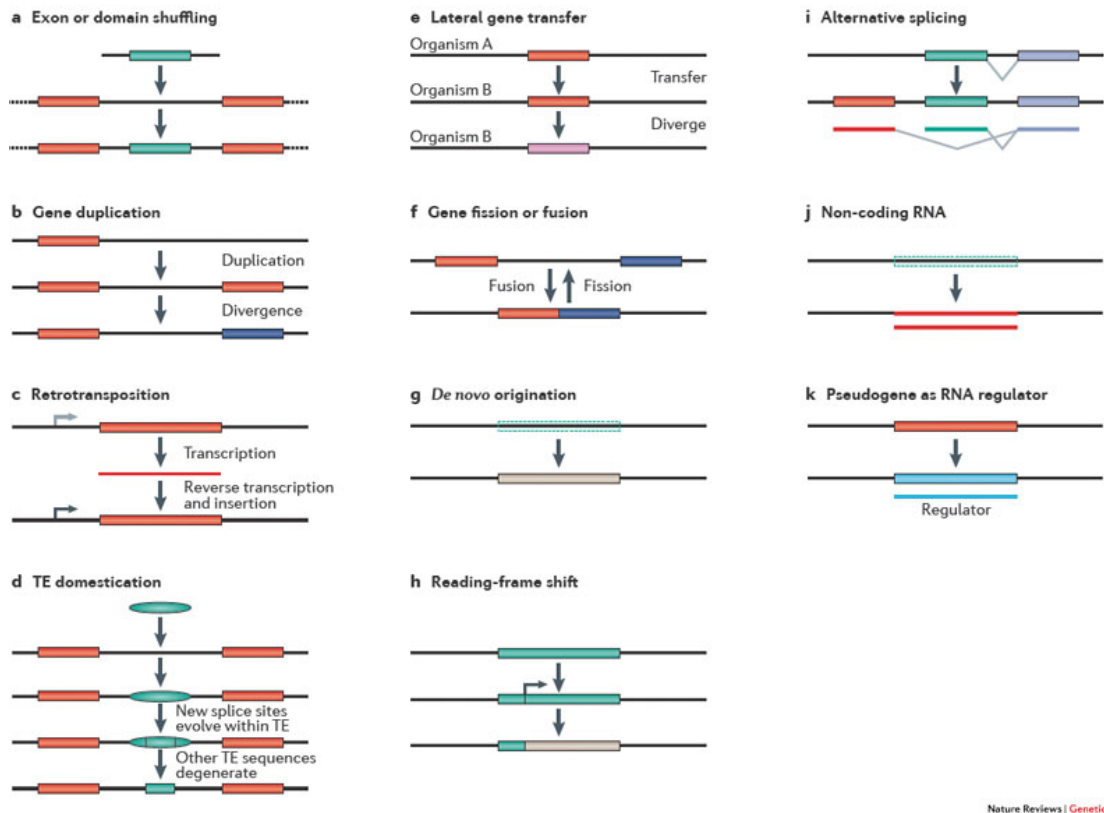
Scheme 1.1. Eleven known molecular processes that can generate new genes. (1)

It is often the case that a mutation in a DNA coding region does not produce a measurable effect. Due to the redundancy of the genetic code, a single nucleotide polymorphism does not always result in a new protein sequence. In the event that the variant translates to a change in primary structure, the magnitude of biological change depends on several different variables, including residue similarity and structural location of the variant.

Significant previous work has quantified the effects of mutating one residue to another. One such approach is to analyze sequences to generate substitution matrices of sequence conservation. One of the most commonly used matrices, BLOSUM, is based on of methods that bases substitution rates off of highly conserved regions in the

genome (2, 3). By counting substitution probabilities of these regions, Henikoff and Henikoff generated log odds scores for each pairing, corresponding to the substitution likelihood for each pairing. Alternatively, one can perturb the structure of a protein *in silico* to measure the effect of such mutations on the dynamics and function of different proteins. Such methodologies vary in nature and are restricted to such proteins with known or modeled structures, but often combine conformational sampling with free energy estimations (4). By simulating the wild type and mutant versions of the same protein under the same conditions, one might derive the functional consequences of a specific variant. These consequences are diverse in nature and vary based on the location and type of the mutation.

As receptors bind specific substrates, one can imagine that a perturbation, whether changes in metabolites in the environment or a novel variant in a protein, in one would necessitate changes in the other. This purported coevolution of receptors and their substrates suggests that due to relative fitness, the interactions between proteins and substrates tend to become more optimal over time (5-7). This does not always result stronger, or more selective binding, of which the millimolar binding of ATP to myriad receptors is a prime example. Understanding how these specific perturbations drive coevolution can provide structural and functional insights to inform rational drug discovery.

How can we estimate the affinity of protein-ligand interactions?

While computer-aided drug design has existed for several decades, many problems still persist. As computational chemists, we are primarily concerned with

how strongly and selectively an inhibitor binds to a given target. This binding free energy can be decomposed into two components, enthalpy, which measures the energy of a system, and entropy, which measures system disorder. These components can be further decomposed and estimated using various methods (8-10). While calculation of the absolute binding free energy is used to identify strong binders, evaluation of the relative binding free energy between two targets can identify potential off target side effects.

As many of these methods approximate a combination of full conformational search and binding free energy calculations, error persists in these computations. Much research is currently underway to improve the accuracy and precision of these calculations, but increasing the complexity and accuracy of these calculations almost invariably increases the time it takes to evaluate each bound complex. These computational approaches have been used to facilitate the design of highly specific inhibitors (11). However, standard approaches, such as virtual screening, have associated errors in the scoring functions that are often large enough to complicate the identification of selective inhibitors (12-15). Nevertheless, such computational docking approaches are valuable when screening a large number of compounds against a target. Even though precise energies and rankings are not often obtained, certain molecular interactions, such as hydrogen bonding and electrostatic interactions can still be observed.

Other approaches focus on characterizing the protein binding site (16-21). While these methods do not necessarily allow for the automated placement of a ligand or fragment into a binding pocket of interest, they are useful for comparing binding

pockets between proteins. Such approaches have been leveraged to both repurpose drugs to target new diseases and understand their polypharmacology (22-24). Other methods employ physics- and heuristic-based approaches to predict where certain atoms may lie in a binding pocket, and use that information to construct molecules (16).

The importance of selectivity

One of the greatest challenges for computational chemists is designing small molecules that are not only high affinity compounds, but also selective for their target. Inability to develop selective compounds results in aberrant side effects, which can be deleterious for both patients and the pharmaceutical industry. Although these proteins are often unrelated on a global scale by sequence, their binding sites have evolved to bind similar ligands. Identification of off-target effects earlier in the drug discovery process by cataloguing similarities in protein structures could drastically reduce costs associated with failures in late-stage clinical trials.

One recent example of this was the withdrawal of Pfizer's CETP inhibitor, Torcetrapib, from phase III clinical trials. Before this withdrawal, Pfizer had invested approximately $800 million in the development of this compound to serve as a complement to its lucrative drug, Lipitor (25). Earlier identification of its deadly hypertensive side effects would have saved millions in development, and more importantly, those who succumbed to these side effects. Later computational analysis by Xie et al. found that several proteins that did not share a global structural similarity to CETP still possessed binding sites of similar geometry and composition to that of

CETP(26). In particular, they postulated that Torcetrapib bound to several nuclear hormone receptors and up-regulated the Renin-Angiotension-Aldosterone System, leading to potentially fatal hypertension, which was the cause of the initial withdrawal.

Evolving away from drugs: The development of resistance

Perhaps the most sobering effect that evolution has is the complication of disease treatment through the development of resistance. The evolutionary arms race between diseased cells/organisms and corresponding therapies is an incredible medical burden and necessitates continual development of novel drugs for many diseases. Complicating efforts in treating these diseases is that drug therapies accelerate the rate of natural selection, and therefore resistance. Similar to founder effects, these remaining cells proliferate with the now-dominant mutations, shifting the tumor allele frequency toward the resistant genotype.

Two diseases of note that have received particular interest from pharmaceutical companies as a result of evolved resistance are HIV and cancer. In HIV, virions rapidly evolve resistance to standard treatments. Consequently, multiple drugs are often used at once under the assumption that fewer virions will be resistant to all drugs. However, those that survive such treatments, allow for the multidrug resistance to spread more rapidly. Similarly, cancer cells that survive a round of intense chemotherapy often have accumulated somatic mutations that have led to their survival.

Evolution occurs on a molecular scale: Cysteine cathepsins as an example

During the course of evolution, genes accumulate novel mutations that, over time, can lead to novel functions. Such events occur throughout the genome, and give rise to gene families with similar, albeit different, functions. One such example is the C1A papain-like class of cysteine proteases, more commonly known as cathepsins. These proteins are involved in the processing of proneuropeptides and other proteolytic processes, and have been implicated in a variety of diseases (Table 1.1). For example, cathepsin B has been shown to be involved in the pathogenesis of Alzheimer's disease (27), cathepsin L is over expressed in many cancers and is involved in remodeling of the extracellular matrix (28), and cathepsin V has been implicated in several autoimmune diseases (29). As these proteins have high structural homology, small differences in their subpockets determine selectivity. Since these proteases carry out important functions in healthy cells, designing inhibitors to block the function of all cathepsins is not desirable. Therefore, the development of selective cathepsin inhibitors is necessitated.

Table 1.1. Cathepsins and disease. Adapted from (27, 30)

Disease	Cysteine cathepsins
Alzheimer's Disease	B
Atherosclerosis	K, L, and S
Cancer and metastasis	B, F, H, K, L, V, S, and Z
Metabolic Syndromes	K, L, and S
Lung diseases	B, H, K, L, and S
Immune Defects	C, L, V, and S
Rheumatoid Arthritis	B, L, K, and S

The high homology of these proteins complicates computer-aided drug discovery efforts (Figure 1.1A). In fact, in their mature forms, cathepsins L and V share 79% of their sequences as well as structurally indistinguishable protein backbones. Interestingly, a significant portion of the differences between these proteins involves charged residues. One possible explanation for this is that these two proteins likely resulted from a gene duplication of an ancestral cathepsin, with cathepsin V evolving to become more positively charged in the active site, and cathepsin L becoming more negatively charged (Figure 1.1B). For example, these electrostatic differences can help result in differences in selectivity as demonstrated by higher binding affinity of the Z-Phe-Arg-AMC fluorogenic substrate for cathepsin L as compared to cathepsin V. Specifically, the greater negative charge in the S2 pocket of cathepsin L drives this selectivity as it better accommodates a positively charged substrate. Thus, despite the high sequence and structural homology, there is hope for the design of inhibitors that selectively target each cathepsin. Certain groups have taken advantage of these electrostatic differences, and have had particular success in designing inhibitors selective for cathepsin L over cathepsin V (31).

A

B
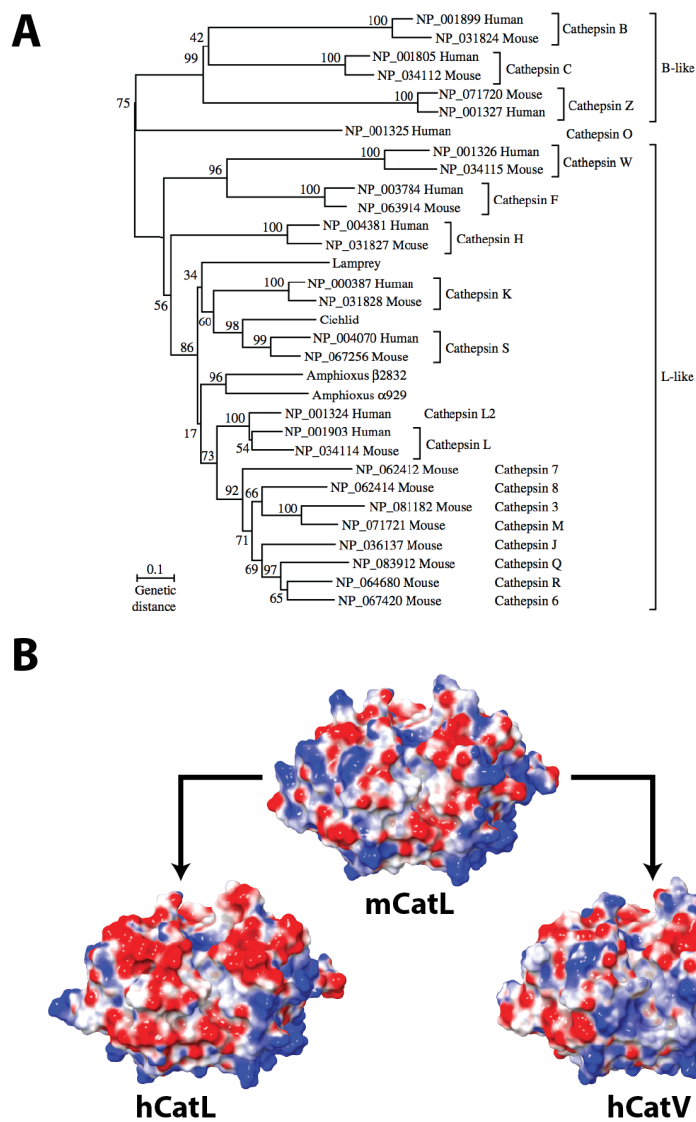
Figure 1.1. Cysteine cathepsins have a high homology. (A) High sequence similarity of different human and murine cathepsins, from (32). (B) Cathepsins V and L evolved different electrostatic potentials from a common cathepsin ancestor.

Chapter 2: The molecular dynamics of *Trypanosoma brucei* UDP-galactose 4'-epimerase: a drug target for African Sleeping Sickness

Abstract

During the past century, several epidemics of human African trypanosomiasis, a deadly disease caused by the protist *Trypanosoma brucei*, have afflicted sub-Saharan Africa. Over 10,000 new victims are reported each year, with hundreds of thousands more at risk. As current drug treatments are either highly toxic or ineffective, novel trypanocides are urgently needed. The *T. brucei* galactose synthesis pathway is one potential therapeutic target. Although galactose is essential for *T. brucei* survival, the parasite lacks the transporters required to intake galactose from the environment. UDP-galactose 4'-epimerase (*Tb*GalE) is responsible for the epimerization of UDP-glucose to UDP-galactose and is therefore of great interest to medicinal chemists. Using molecular dynamics simulations, we investigate the atomistic motions of *Tb*GalE in both the apo and holo states. The sampled conformations and protein dynamics depend not only on the presence of a UDP-sugar ligand, but also on the chirality of the UDP-sugar C4 atom. This dependence provides important insights into *Tb*GalE function and may help guide future computer-aided drug discovery efforts targeting this protein.

Abbreviations

UDP, uridine diphosphate; *Tb*GalE, *Trypanosoma brucei* UDP-galactose 4'-epimerase; HAT, Human African Trypanosomiasis; MD, molecular dynamics; RMSD, root-mean-square deviation; PCA, principal component analysis; *Hs*UGDH, human UDP-glucose dehydrogenase; *Ec*GalE, *E. coli* UDP-galactose 4'-epimerase; *Hs*GalE, human UDP-galactose 4'-epimerase; NAD, nicotinamide adenine dinucleotide; VMD, visual molecular dynamics

Introduction

Human African trypanosomiasis (HAT), a disease caused by the protist *Trypanosoma brucei* (*T. brucei*), directly affects thousands of sub-Saharan Africans and indirectly places hundreds of thousands more at risk. Current treatments are often ineffective or dangerous, necessitating the development of a new generation of HAT therapeutics.

Disease pathology occurs in two stages. The early hemolymphatic stage is characterized by fever, cephalgia, arthralgia, and pruritus. Current treatments for early-stage HAT include pentamine and suramin. However, pentamine, often used to treat the *T. brucei gambiense* strain, is associated with hypoglycemia and hypotension, and suramin, effective against *T. brucei rhodesiense*, is associated with severe cutaneous reactions, renal failure, anaphylactic shock, and neurotoxicity (33-36). Once the parasite crosses the blood–brain barrier, the more serious and often irreversible symptoms of the neurological phase are manifest, including disrupted cognition, coordination, and sleep (37). Since 1949, melarsoprol has been the standard

treatment for late-stage HAT. Despite its efficacy, a significant number of patients relapse; additionally, 5–10% of those who receive treatment develop severe encephalopathy, often leading to death (38). In the late 20th century, Aventis developed eflornithine, a drug that targets *T. brucei* ornithine decarboxylase. While eflornithine is significantly less dangerous than melarsoprol (39), it is ineffective against the *T. brucei rhodesiense* subspecies (40).

As current therapeutics are problematic, medicinal chemists are actively seeking to identify novel *T. brucei* drug targets. The proteins of the biochemical pathway involved in galactose synthesis are excellent candidates. Although *T. brucei* requires galactose for the synthesis of vital glycoproteins (41), it is unable to intake galactose from the environment. Instead, glucose is acquired via hexose transporters (42) and is subsequently converted to galactose. One of the proteins in the pathway that effectuates this conversion, UDP-galactose 4′-epimerase (*Tb*GalE), inverts the stereochemistry of the UDP-glucose C4 carbon atom to produce UDP-galactose. The two-step reaction proceeds via a transient ketose intermediate and requires rotation of the sugar portion of the ligand for epimerization (Scheme 2.1) (43). Initial drug discovery efforts targeting this protein have identified several promising inhibitors (44). Unexpectedly, virtual screening and subsequent experimental validation identified several agonists as well.
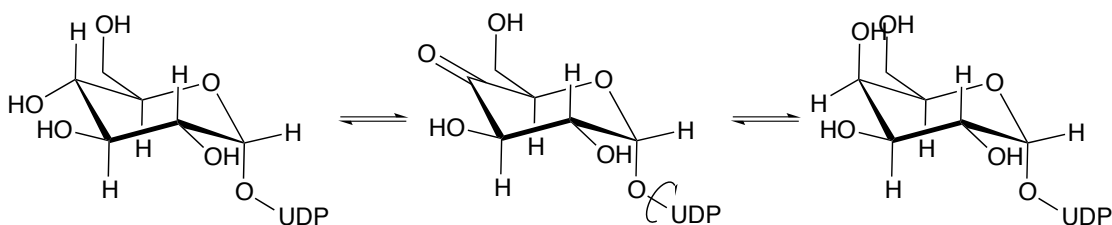
Scheme 2.1. UDP-sugar epimerization. The epimerization reaction proceeds via a UDP-ketose intermediate. Epimerization requires a 180° rotation ('flip') of the sugar portion of the ligand.

This study provides insight into the general mechanisms of *Tb*GalE ligand binding. The static lock-and-key model of binding, first proposed by Emil Fischer in 1894 (45), has been largely abandoned in favor of theories that account for receptor flexibility. Specifically, the induced-fit and population-shift models have been much debated over the course of the past 50 years (46). Induced fit suggests that *apo* and *holo* receptors sample distinct regions of conformational space; ligand binding itself induces conformational changes in the receptor (47). Population shift, on the other hand, suggests that an *apo* protein samples many conformations in solution, a subset of which are amenable to ligand binding. Ligands bind to rare but amenable receptor conformations and energetically stabilize them, causing the population of all conformations to shift toward those that can accommodate the ligand (48-51).

Here, we explore the atomistic dynamics of the *Tb*GalE homodimer by investigating the major steps along its reaction coordinate. Using molecular dynamics (MD) simulations, we study the motions of *Tb*GalE homodimer in both the *apo* and *holo* forms, bound to UDP-galactose, UDP-glucose, and the UDP-ketose intermediate. The conformations sampled by the protein depend not only on the presence of a UDP-sugar ligand, but also on the chirality of the UDP-sugar C4 atom.

This dependence provides important insight into *Tb*GalE function and may help guide future computer-aided drug discovery efforts targeting this protein.

Results and Discussion

Evidence of simulation equilibration

Alpha carbon RMSD plots of each homodimer trajectory are shown in Figure 2.1. As the initial protein conformation of each system was that of the crystallographic UDP-galactose-bound state, the first 9 ns of each homodimer simulation were discarded to account for system equilibration. The remaining 50 ns of the dimeric simulation were used for subsequent analysis. In total, 400 ns of productive *Tb*GalE monomer simulation were generated.
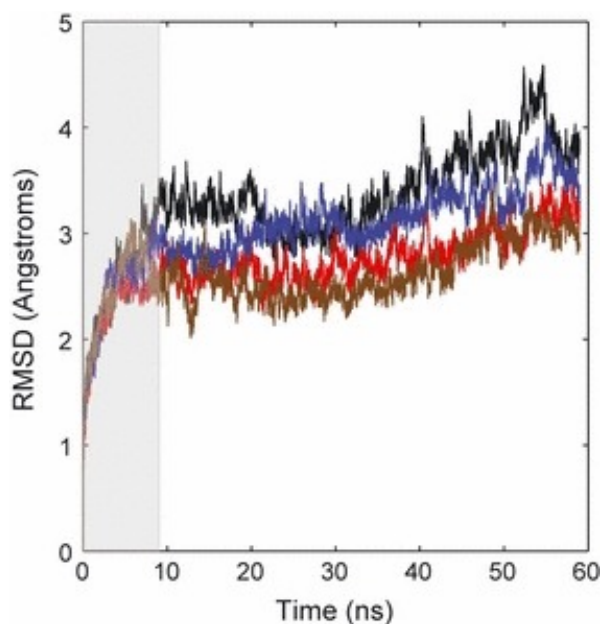
Figure 2.1. Trajectory RMSD. Each dimer simulation was aligned to the first frame by minimizing the root-mean-square deviation (RMSD) of the Cα's. RMSD was calculated using the first frame as a reference. For this plot, and in subsequent figures, *apo* is depicted in black, UDP-galactose in blue, UDP-glucose in red, and the UDP-ketose intermediate in brown. As each substrate was derived from UDP-galactose, the systems needed 9 ns to equilibrate (gray box). Trajectory analysis was performed on the subsequent 50 ns of each dimer simulation.

A gating mechanism may mediate ligand binding

To better understand how receptor dynamics might impact ligand binding, principal component analysis (PCA) was used to identify the most important molecular motions of each of the four simulations. Principal components, or eigenvectors, were first calculated for the backbone atoms of each trajectory, and each trajectory was subsequently projected onto the *apo* eigenvectors for reference. A majority of the variance in the molecular motions could be explained by the first two principal components, with almost 40% of the variance explained by the first eigenvector alone (Figure 2.2A). Consequently, only the first two principal components were used for subsequent analysis (Figure 2.2B).

The Interactive Essential Dynamics computer package (52) was used to identify the extreme structures of each simulation (Figure 2.2C,D), as judged by projection onto the first *apo* eigenvector. The largest differences in the motions defined by the first principal component were present in the loop region containing residues 230-255. This loop was already thought to be highly dynamic, given that residues 236–248 were unresolved in the 2CNB crystal structure (53).
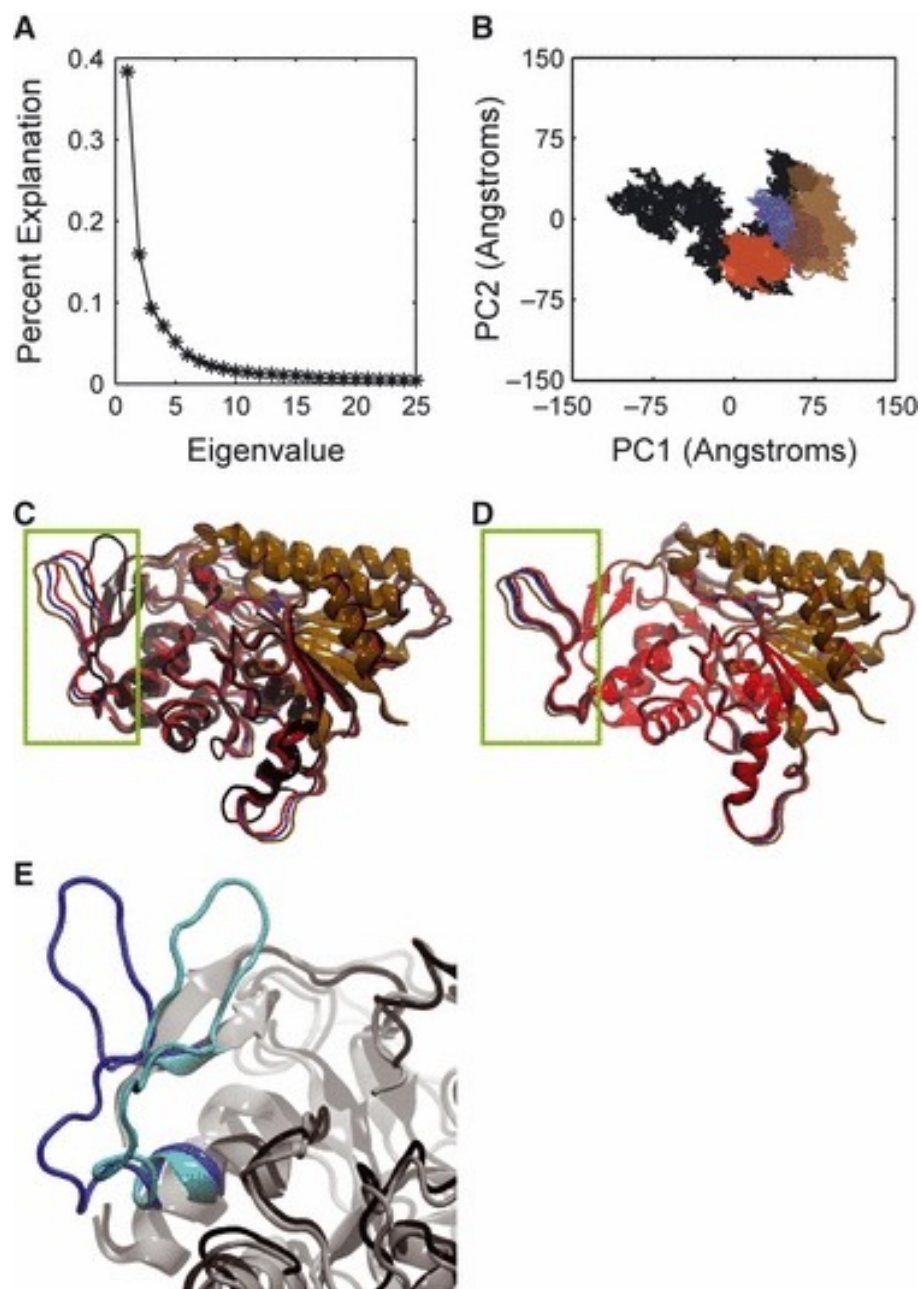
Figure 2.2. Principal component analysis (PCA). To facilitate comparison, all trajectories were projected onto the principal components of the *apo*simulation. *Apo* is shown in black, UDP-galactose in blue, UDP-glucose in red, and the UDP-ketose intermediate in brown. (A) Percent of the trajectory dynamics attributable to each of the top 25 *apo* eigenvectors. A majority of these motions can be explained by the first two principal components (PC). (B) Projection of the trajectories onto the first two *apo* eigenvectors. Each *holo*simulation explores its own unique set of motions. (C) The PC1 minimum extreme structures for each simulation, color-coded as above. The location of gate is boxed in green. (D) The PC1 maximum extreme structures, color-coded as above. The location of gate is boxed in green. (E) The closed conformation evident in the PC1 minimum *apo* structure results from a kink in the highlighted α-helix. The PC1 minimum structure is shown in cyan/gray, and the maximum is shown in blue/black for reference.

Much of the difference in the minimum extreme structures can be explained by a kink that forms at the C-terminal end of the α-helix preceding the loop containing residues 230–255 in the *apo* simulation (Figure 2.2E). This helical bend causes the flexible loop to move toward the active site, as evidenced by the change in the angle defined by the alpha carbon atoms of D245, G228, and R109, which ranges from 67.69° to 36.01° as PC1 decreases, compacting the protein. Additionally, the corresponding distance between D245 and R109 similarly changes from 29.81 Å to 16.55 Å. The loop may therefore serve as a gating mechanism likely relevant to ligand binding. As the gating loop containing residues 230–255 is near the active site, it is possible that the movement of this loop helps control the UDP-sugar rate of entry.

UDP-sugar binding: induced fit vs. population shift

The principal components suggest that both induced fit and conformational shift play a role in the binding of *Tb*GalE to its natural substrates. The projection of the *apo* simulation onto the first two principal components demonstrates that the *apo* protein explores a large region of conformational space (Figure 2.2B). Similar projections of the remaining simulations onto the same *apo* principal components reveal that the UDP-glucose and UDP-galactose simulations sample distinct regions within this larger, *apo* conformational space, suggesting a population shift mechanism of binding. On the other hand, the UDP-ketose intermediate samples a region of conformational space largely unexplored by the *apo* protein; the intermediate itself may therefore provoke receptor conformational changes via an induced-fit

mechanism. Limited conformational overlap was expected as the ligands only differ at the UDP-sugar C4 atom; these overlaps likely represent the transitions between states during the actual reaction.

These results suggest a general binding mechanism that may be relevant to the study of other receptor–ligand systems as well. The *apo* protein likely samples a large region of conformational space, occasionally assuming conformations in which the gating loop is open, permitting access to the main, UDP-sugar binding site. When UDP-glucose binds, it stabilizes certain conformations, causing the region of the sampled conformational space to constrict via a population shift mechanism. Next, the bound ligand induces changes in the receptor conformation uncharacteristic of the *apo* protein via an induced-fit mechanism. These conformational changes are likely caused by the transformation of the sugar and are required to accommodate the UDP-ketose intermediate. As the epimerization progresses, the conformational space sampled by the protein returns to that sampled in the *apo* state, again indicative of a population shift mechanism of binding.

Remarkably, the regions of the *apo* conformational space sampled by the UDP-galactose- and UDP-glucose-bound proteins are distinct, despite the fact that these two ligands differ by only a chiral inversion at a single sugar carbon atom, demonstrating that even small differences in bound ligands can drastically change the region of conformational space sampled. This finding has significant implications for computer-aided drug design, as it suggests that the crystallographic and MD-derived structures used in virtual screening can vary remarkably depending on the bound ligand. As previous computer-aided drug discovery efforts directed toward *Tb*GalE

only used *holo* simulations to generate receptor structures for virtual screening (44), additional screening against *apo* conformations may reveal new, effective *Tb*GalE inhibitors.

Hydrogen bonding of the UDP-sugar

To examine how the hydrogen bond network of the ligand changes over the course of the reaction, we first identified persistent hydrogen bonds present in over 75% of the trajectory frames analyzed. As expected, the hydrogen bond network that mediates the binding of the UDP-ketose intermediate differs from the networks associated with UDP-galactose and UDP-glucose (Figure 2.3A–C). Two of the identified hydrogen bonds mediate interactions with Y173 and S142, conserved residues of the catalytic triad that are critical for the two-step GalE epimerization reaction (Scheme 2.1) (43, 54).

The hydrogen bond with S142 is particularly interesting, as earlier studies of *Ec*GalE suggested that S142 shuttles a proton to Y173 (55). However, crystallographic studies of several GalE proteins have called into question this theory, as they reveal that the distance between the S142 and Y173 hydroxyl groups is generally large. Our results suggest that the specific role of S142 is to stabilize the UDP-ketose intermediate via hydrogen bonding with the C4 hydroxyl group during the conformational flip required for epimerization (Figure 2.3B).
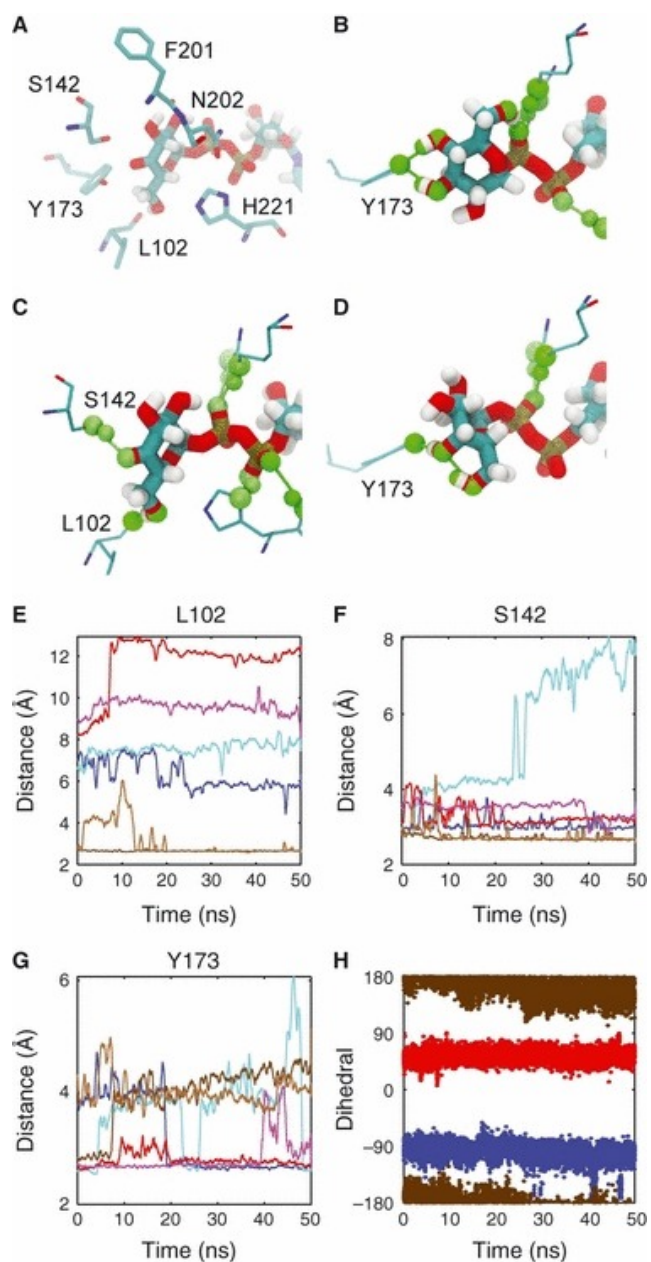
Figure 2.3. Ligand hydrogen bonding. Hydrogen bonds present in over 75% of the 1000 equidistant frames extracted from each MD simulation are shown in green. In all charts, UDP-galactose is shown in blue, UDP-glucose is shown in red, and the UDP-ketose intermediate is shown in brown. For D–F, chain A is represented by the darker color, and chain B by the lighter color. (A) UDP-glucose. The glucose is held in place by two hydrogen bond interactions with Y173. (B) UDP-ketose intermediate. The orientation of the sugar is flipped from that of UDP-glucose. The sugar C6 hydroxyl group interacts with L102, and the intermediate carbonyl oxygen is tethered to S142. (C) UDP-galactose. The now epimerized sugar, still flipped, forms a single hydrogen bond with Y173. (D–F) Moving average using 500 ps windows. (D) Distance from the L102 carbonyl oxygen to the C6 hydroxyl on the UDP-sugar. (E) Distance from the S142 side chain hydroxyl to the C4 hydroxyl on the UDP-sugar. (F) Distance from the Y173 side chain hydroxyl to the C4 hydroxyl on the UDP-sugar. (G) [C1-O-P-O] torsion angle of the UDP-sugar. C1 belongs to the sugar.

The UDP-sugar hydrogen bond networks were generally similar regardless of which monomer was used for analysis, with the exception of UDP-galactose. When acceptor–donor heavy-atom distances were calculated over the course of the trajectories (Figure 2.3D–F), it was noted that the chain B active site of UDP-galactose underwent a conformational shift approximately 5 ns into its trajectory, causing a break in the hydrogen bonds between the UDP-galactose and S142 and Y173, respectively. However, this shift was not observed in chain A, which had a UDP-sugar hydrogen bond network more consistent with the *Tb*GalE crystal structure, 2CNB. Additionally, despite the conformational shift evident in chain B, the key sugar contacts with S142 and Y173 continued to transiently reform during the course of the simulation, confirming that these residues play an important role in stabilizing the UDP-sugar. Other residues, such as N202 and H221, also formed hydrogen bonds with the UDP-sugar, but more transiently (data not shown).

Like S142 and Y173, L102 may also be important for catalysis, warranting further pharmacological study. Our simulations suggest that L102 may be fundamental to the conformational flip required for epimerization (Scheme 2.1). The carbonyl oxygen atom of L102 is predicted to form a persistent hydrogen bond with the C6 hydroxyl group of the UDP-ketose intermediate. Previous crystallographic studies of a UDP-galactose-like ligand suggested that L102 might mediate ligand binding through interactions with the UDP-sugar C6 hydroxyl group (53). Our simulations suggest that an additional function of the L102 backbone carbonyl group is to maintain the UDP-ketose-intermediate in a conformation approximately halfway between that of UDP-galactose and UDP-glucose via a persistent hydrogen bond. The average UDP-sugar

[C1-O-P-O] dihedral angle of the UDP-ketose intermediate is approximately halfway between that of UDP-galactose and UDP-glucose (Figure 2.3G), supporting this notion.

The same hydrogen bond networks that mediate UDP-sugar binding may be germane to the design and optimization of potential small-molecule therapeutics. For example, a common practice in designing inhibitors is to create transition-state analogs. Our simulation of the UDP-ketose intermediate bound to *Tb*GalE has elucidated the important role L102 may play in stabilizing this intermediate. These results suggest that L102 may be a good residue to target in future drug discovery efforts, in addition to the previously identified Y173 and S142 (53).

Identification of conserved residues

As current trypanocidal compounds are subject to growing resistance (56-60), future drug design strategies should also attempt to anticipate mutations that may reduce therapeutic efficacy. One strategy to avoid resistance is to pharmacologically target protein residues that are conserved across related members of the same protein family. Conserved residues are likely to be essential for catalysis and/or the binding of natural substrates and thus are less likely to undergo resistance-conferring mutations. After considering the 392 reviewed members of the sugar epimerase family listed in UniProt, five representative proteins spanning both the eukaryotic and bacterial domains of life were selected: *Trypanosoma brucei* GalE (PDB ID: 2CNB), a chloroplastic protein from *Arabidopsis thaliana* (UniParc: Q8H124), Gal10 from *Schizosaccharomyces pombe* (UniParc: Q9HDU3), bifunctional polymyxin

resistance protein ArnA from *Yersinia enterocolitica* (UniParc: A1JPN5), and

probable rhamnose biosynthetic enzyme 1 from *Arabidopsis thaliana* (UniParc:

Q9SYM5). The sequences of these five proteins were aligned, and the

following *Tb*GalE active site residues were found to be highly conserved: G7, G10,

I12, D75, A100, N117, S141, S142, A143, Y173, K177, and R268. Of these, S142,

A143, Y173, K177, and R268 appear to associate with the UDP-sugar.

A143 is particularly noteworthy. While A143 did not form a persistent

hydrogen bond with the UDP-sugar C3 hydroxyl group in our simulations (i.e., the

bond was present in fewer than 75% of all simulation frames analyzed), this hydrogen

bond did form transiently, consistent with the 2CNB crystal structure. As S142 and

A143 form hydrogen bonds with adjacent UDP-sugar hydroxyl groups, concurrent

binding likely serves to stabilize the conformation of the UDP-sugar in a conformation

amenable to catalysis. Thus, small-molecule inhibitors that have adjacent hydroxyl

groups or similar hydrogen bonding moieties might be good candidates targeting these

two conserved residues.

Allosteric binding

In a recent study, Durrant *et al.* (44) used virtual screening to identify *Tb*GalE

ligands from among the compounds of the NCI Diversity Set II. Fourteen low-

micromolar inhibitors were ultimately reported. However, in addition to these

antagonists, seven agonists, unreported at the time, were also identified (Table S2.1).

We found this result interesting, as the agonists were identified by docking small-

molecule models into the *Tb*GalE active site and therefore should compete with the UDP-sugar substrate rather than enhance its catalysis.

Although *Tb*GalE agonists are ill suited as HAT therapeutics, a human condition known as type III galactosemia arises from a deficiency in *Hs*GalE. Understanding GalE agonism is therefore of great therapeutic interest. To determine whether agonism might arise from allosteric binding, the *holo* and *apo* monomer trajectories were clustered into five groups using RMSD cutoffs of 0.7 and 0.75 Å, respectively (Figure 2.4A). A representative protein conformation was then taken from each cluster; together, these representative conformations are said to constitute an ensemble.

The FT-MAP server (21) was then used to computationally flood the entire surface of the ensemble conformations with models of small organic probes in an attempt to identify potential binding pockets. FT-MAP identified the NAD(H) and UDP-sugar binding pockets in every structure; however, no other potential allosteric pockets were consistently recognized (Figure 2.4B), suggesting that agonism does not likely occur by allosteric modulation.
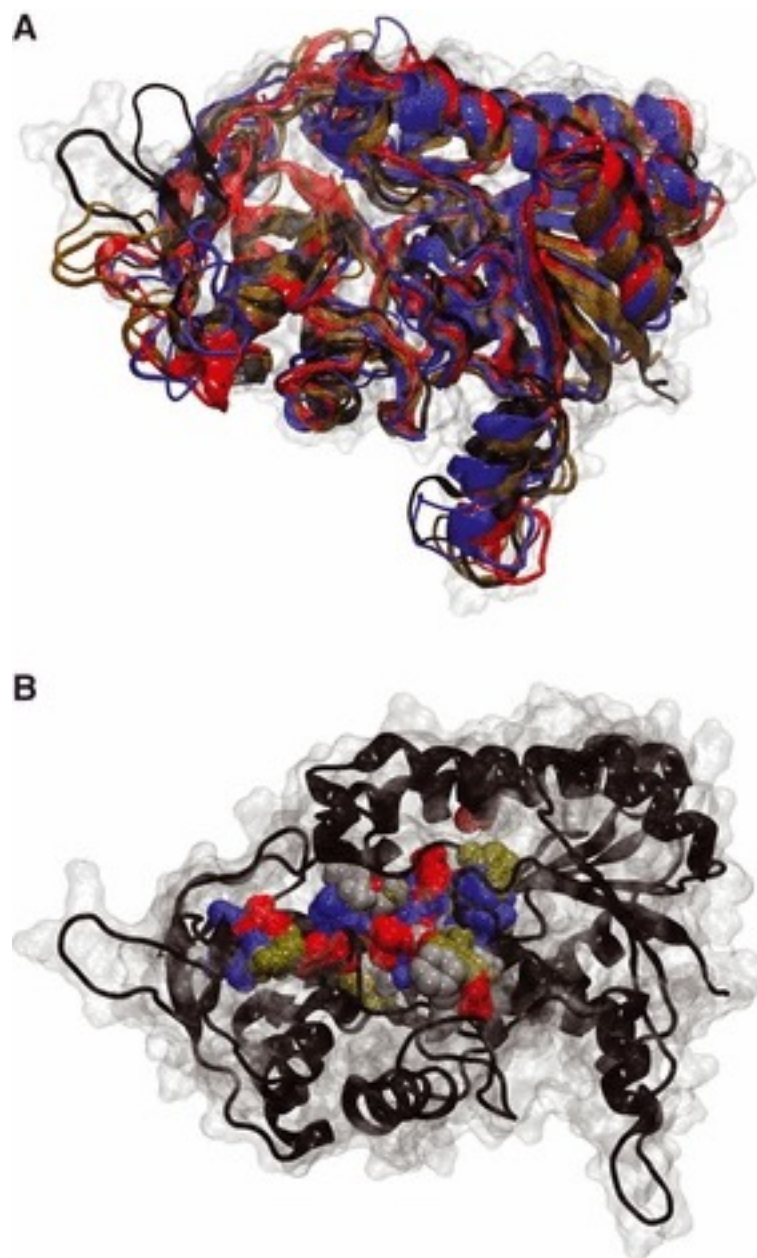
Figure 2.4. Binding pocket identification. (A) Active site clustering of the*Tb*GalE monomers. The frames of each simulation were clustered by the active site Cα's using 0.70 and 0.75 Å RMSD cutoffs for the *holo* and *apo*simulations, respectively. The top two representative structures of chain A from each simulation are shown; *apo* is depicted in black, UDP-galactose in blue, UDP-glucose in red, and the UDP-ketose intermediate in brown. Darker and lighter colors correspond to the most populated and the second most populated cluster, respectively. (B) FT-MAP analysis. The top five clusters from each chain were submitted to the FT-MAP server. Shown are the results for the top chain A cluster of each simulation. These results suggest that*Tb*GalE contains no allosteric sites; the observed agonism likely results from ligand-induced dimerization and/or cooperativity between the two monomers.

While the possibility of allosteric binding at an uncharacterized secondary site could not be ruled out, the fact that FT-MAP revealed no such site suggests cooperative agonism mediated by binding to the principal site. It may be that agonist binding to one *Tb*GalE monomer alters the affinity of the binding site on the other dimeric monomer, thus facilitating additional binding. The steep Hill coefficients associated with several of the inhibitors, even in the presence of a detergent that disrupts colloidal aggregates, support this theory (61, 62). Hemoglobin is the classic example of cooperative binding, but many other examples exist in nature as well. As an alternate explanation, agonist binding to the principal site of one *Tb*GalE monomer might drive dimerization by stabilizing the four-helix bundle at the dimer interface. This theory is supported by previous evidence in related proteins suggesting that, while the monomer is functional, full functionality is achieved only through dimerization (63, 64).

Unfortunately, when the identified *Tb*GalE agonists were tested against *Hs*GalE, no agonism was noted (Figure S2.1, Tables S2.1 and S2.2). However, it is possible that the lack of activity arose from differences in the human and *T. brucei* versions of the GalE protein rather than fundamental differences in the agonistic mechanism. For example, the NAD^+ adenine-binding pocket is more closed in *Tb*GalE, and*Hs*GalE G237 in the UDP-sugar binding domain is replaced by C266 in *Tb*GalE, a potentially reactive residue that may prove useful in future drug design efforts (43). Further efforts to identify *Hs*GalE agonists may therefore be justified.

Conclusions:

In this work, we have used molecular simulations to probe the dynamics of *Tb*GalE and to specifically analyze the mechanisms governing ligand-binding and enzymatic conversion. Our MD simulations suggest that the conformations sampled by *Tb*GalE are highly dependent on the composition of the ligand, as even the chirality of the UDP-sugar C4 atom greatly affected the conformations explored.

Additionally, we have identified an active site residue, L102, that may be important in the stabilization of the UDP-ketose intermediate. While this residue has been previously identified in crystallographic studies as a potential mediator of UDP-galactose binding (53), to our knowledge, its possible role in stabilizing the UDP-ketose intermediate has not been previously recognized.

Finally, as none of our simulations revealed any *Tb*GalE allosteric sites, we anticipate that dimeric agonism likely results from either cooperative binding or dimeric stabilization. We hope that the results presented here will not only provide insight into the function of this and related enzymes, but also assist future computer-aided drug discovery efforts targeting *Tb*GalE and *Hs*GalE.

Materials and Methods:

System preparation

A crystal structure of *Tb*GalE homodimer [PDB ID: 2CNB (53)] was obtained from the Protein Data Bank (65). To generate missing loops, the structure was aligned to a model of *Tb*GalE that had been created previously (44) based on the 1GY8 *Tb*GalE structure (43). Following alignment, the coordinates of the missing

loop atoms were copied from the 1GY8 model. All cocrystallized waters were retained; PDB2PQR (66) and PROPKA (67, 68) were used to assign residue protonation states at pH 7.0. Histidine protonation states were visually inspected in VMD (69) to ensure optimal hydrogen bonding. The NAD(H) and Y173 protonation states were manually assigned to properly mimic the active site configuration needed for catalysis (43).

NAD+/NADH parameters were obtained from Walker *et al.* (70, 71). Accelrys Discovery Studio 2.5 was used to model the structure of UDP-galactose by changing the fluorine atom of the 2CNB UDP-4-deoxy-4-fluoro-alpha-D-galactose ligand to a hydroxyl group. UDP-glucose and the UDP-ketose intermediate were built by altering the stereochemistry and hybridization of the UDP-galactose C4 carbon atom. Hydrogen atoms were added to the three UDP-sugars using Discovery Studio. All ligand partial charges were generated using GAUSSIAN03, and ligand atoms were parameterized according to the GAFF force field (72).

Receptor atomic parameters and partial charges were assigned according to the Amber ff99SB force field (73) using the AMBER 10 *xleap* module. Sodium ions were subsequently added to bring the system to electric neutrality. The protein was then solvated in a TIP3P (74) water box that extended 10 Å beyond the protein in each direction, and additional sodium and chloride ions were added to bring the total salt concentration to 20 mM.

Molecular dynamics simulations

NAMD 2.6 (75, 76) was used for all minimizations, equilibrations, and free-dynamics runs. Minimization and equilibration steps were performed as described previously by Wang *et al.* (72). In brief, each system was minimized in four phases totaling 45 000 minimization steps. Hydrogen atoms were relaxed in the first 5000 steps; hydrogen atoms and water molecules were relaxed in the next 5000 minimization steps; hydrogen atoms, water molecules, and the atoms of the protein backbone were relaxed in the next 10 000 minimization steps; and all atoms were relaxed for the last 25 000 minimization steps.

For equilibration, 1-ns simulations were performed at 310K using the final minimized structures as the initial coordinates. Harmonic constraints were placed on the atoms of the protein backbones and relaxed in a series of four 250-ps steps. The harmonic restraining force was weakened from 4.0 kcal/mol/Å^2 in the first 250-ps segment to 3.0, 2.0, and 1.0 kcal/mol/Å^2 in the following steps, respectively. Before beginning the productive dynamics simulations, each system was checked to ensure that the root-mean-square deviation (RMSD) between the equilibrated and pre-minimization structures was <1 Å.

For each of the four systems, a 59-ns MD simulation was then performed with a 2-fs time step. Bonds with hydrogen atoms were constrained using the RATTLE algorithm (77), and water geometries were maintained using SETTLE (78), with a bond length error of 0.0005 Å. The temperature bath was kept at 310K with Langevin dynamics. The pressure was maintained at 1 atm using the Nose–Hoover–Langevin piston method (79) with period and decay times set at 100 and 50 fs, respectively.

Long-range electrostatics were calculated using Particle mesh Ewald (80). The free-dynamics runs were performed on the TACC Ranger supercomputer. A typical benchmark on the 102 911 and 102 884 atom systems was 4.35 ns/day of simulation on 96 processors. The system was sampled every 1 ps, generating a total of 59 000 coordinate snapshots. For analysis, every 5th frame was used. Each frame was aligned to the first frame of the trajectory by minimizing the alpha carbon root-mean-square (RMS) deviation using the RMSD Trajectory Tool in VMD (69).

Trajectory clustering

The monomers of each homodimer simulation were isolated, and the two resulting trajectories were concatenated to form a single monomeric trajectory. These monomeric trajectories were subsequently clustered using the gromos algorithm as implemented in the GROMOS++ analysis software (81). The alpha carbon atoms in the active site, defined as all alpha carbon atoms belonging to a residue that was within 5 Å of the NAD or UDP-sugar in the first frame of the trajectory, were used for the mass-weighted RMSD clustering. The RMSD cutoff was increased by 0.05 Å until the trajectory clustering yielded fewer than 30 clusters, with over 90% of all frames contained in the seven largest clusters.

Hydrogen bond analysis

Frames extracted from the simulation every 50 ps were used for hydrogen bond analysis. The program HBonanza (82) was set to identify all persistent hydrogen bonds present in at least 75% of the frames. Angle and distance cutoffs of 30° and 3.5 Å were used, respectively.

Principal component analysis

Principal component analysis (PCA) was performed using the *ptraj* module in AMBER 10[b]. Normalizing each eigenvalue of the covariance matrix to its total sum yields the percent of all *Tb*GalE movements attributable to the corresponding eigenvector. That is, the largest eigenvalues correspond to the PC modes that best explain the molecular motions sampled by the system trajectories. The principal component projections were visualized using a modified version of the Interactive Essential Dynamics module in VMD (52).

Identifying highly conserved active site residues

UniProt (83) was used to identify 7142 members of the sugar epimerase family. Only reviewed sequences were considered for subsequent analysis. These sequences were loaded into the MultiSeq (84) module of the VMD molecular graphics computer program (69). MultiSeq was used to generate a non-redundant set of four representative sequences from this input. ClustalW (85) was then used to align these four sequences to the sequence of *Tb*GalE obtained from the PDB structure 2CNB (53). A set of active site residues was obtained by identifying all receptor residues from the 2CNB structure (chain A) that came within 3.5 Å of the cocrystallized NAD$^+$ cofactor and the UDP-sugar substrate.

HsGalE inhibition and thermal scanning fluorimetry assay

Recombinant wild-type *Hs*GalE and human UDP-glucose dehydrogenase (*Hs*UGDH) proteins were expressed in *E. coli* and purified as described previously

(86, 87). DTP compounds were obtained from the NCI/DTP Open Chemical

Repository (http://dtp.cancer.gov).

The rate of the HsGalE-catalyzed reaction was determined by coupling it to the

oxidation of UDP-glucose by the action of HsUGDH, essentially as previously

described (86). Reaction mixtures (20 nM HsGalE, 1.2 μM HsUGDH, 100 μM UDP-

Gal, 10 mM NAD$^+$, 10 mM HEPES-NaOH, pH 8.8, 1% (v/v) DMSO) were set up in

triplicate with and without DTP compounds (100 μM) in a total volume of 150 μL.

The rate of NADH formation was measured at 340 nm for 20 min at 37 °C using a

Multiskan Spectrum plate-reader spectrophotometer (Thermo Scientific, Hemel

Hempstead, UK) and was used to calculate the rate of production of UDP-glucose

by HsGalE. Initial rates were calculated from the linear section of each progress curve

and the activity expressed as a percentage of the activity in the absence of the

compound.

The binding of the DTP compounds was measured using a thermal scanning

fluorimetry assay (88). This assay has been used previously to identify small

molecular chaperones for the treatment of phenylketonuria (89), to identify stabilizing

additives that facilitate crystallization (90), and to measure the binding affinities of

carbonic-anhydrase inhibitors (91). Sypro orange (Sigma, Poole, UK), a fluorescent

dye, was diluted from a 5000× solution (manufacturer's concentration definition) into

a 50x solution with 10 mM HEPES, pH 8.8. This stock solution was mixed well prior

to each use. Reactions were carried out in a total volume of 20 μL and contained 5 μM

HsGalE, 100 μM DTP compound, 10 mM HEPES, pH 8.8, 1% (v/v) DMSO, 5x Sypro

orange. Melting curves were determined using a Rotor-Gene Q cycler (Qiagen,

Crawley, UK) with the following protocol: high-resolution melt run (460-nm source, 510-nm detector), 25–95 °C ramp with a 1 °C rise for each step, and no gain optimization. The melting temperatures (T_m) were calculated using the inbuilt analysis software.

Acknowledgements:

Chapter 2 is a minimally modified reprint of the material appearing in Friedman AJ, Durrant JD, Pierce LCT, McCorvie TJ, Timson DJ, McCammon JA. The Molecular Dynamics of *Trypanosoma brucei* UDP-Galactose 4'-Epimerase: A Drug Target for African Sleeping Sickness. *Chem Biol Drug Des* 2012, 80, 173-181. I was the first author and co-principal investigator of this paper.

Supporting Information

Table S2.1. *Tb*GalE Agonists. Specific methods can be found in Durrant et al. (2010) *J Med Chem 53*, 5025-5032.
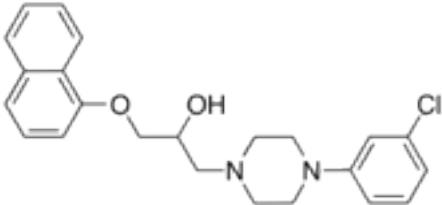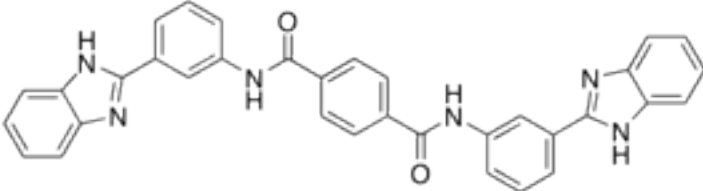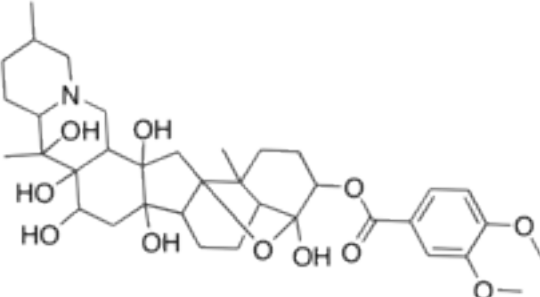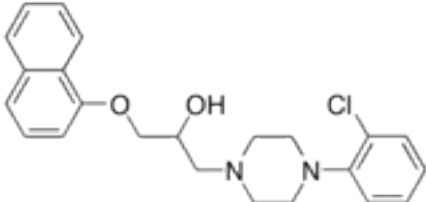
NSC ID	Structure	% inhib. @ 100 mM
91395		-167
61610		-169
7524		-191
91396		-194

Table S2.1 (continued). *Tb*GalE Agonists. Specific methods can be found in Durrant et al. (2010) *J Med Chem 53*, 5025-5032.
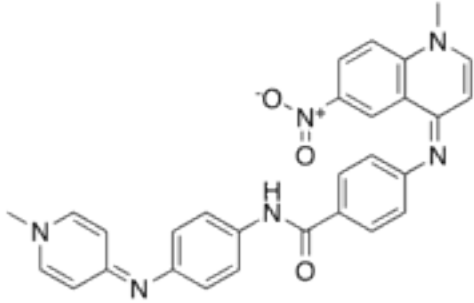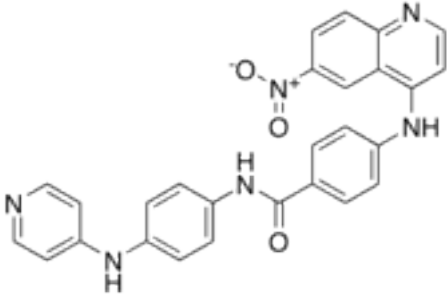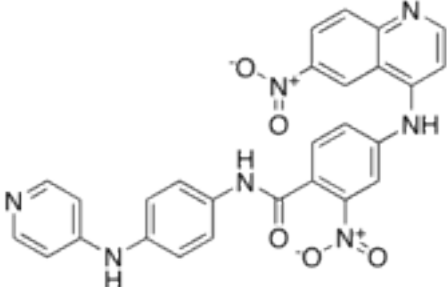
NSC ID	Structure	% inhib. @ 100 mM
260594		**-223**
146771		**-242**
202386		**-283**

Table S2.2. Percentage activity of 20 nM *Hs*GalE in the presence of different DTP compounds. The reactions contained 100 μM DTP compound, 100 μM UDP-Galactose, 10 mM NAD$^+$, 1.2 μM HsUGDH, 10 mM HEPES-NaOH, pH 8.8, 1% (v/v) DMSO. Data are reported as the mean ± SD determined from three separate experiments. No compound resulted in a statistically significant (Student's t-test) change in activity.

DTP Compound	% Activity
No compound	100 ± 17
91395	102 ± 32
61610	89 ± 27
7524	112 ± 3
91396	124 ± 30
260594[a]	30 ± 190
146771[a]	104 ± 57
202386[a]	41 ± 59

Table S2.3. Melting temperatures of *Hs*GalE in the presence of different DTP compounds. The reactions contained 5 μM *Hs*GalE, 100 μM DTP compound, 10 mM HEPES, pH 8.8, 1% (v/v) DMSO, 5× Sypro orange. The change of melting temperature, ΔT_m, due to ligand binding was calculated according to: ΔT_m = (T_m of protein without compound) - (T_m of protein with compound). Data are reported as mean ± SD determined from three experiments. If a compound bound to the enzyme, it would be expected to stabilize the protein's structure resulting in an increase in T_m. However, none of the compounds tested here resulted in a statistically significant (Student's t-test) change in T_m. [a]Compounds 260594, 146771 and 202386 formed a colored precipitate, preventing determination of the melting temperature.

DTP Compound	T_m (°C)	ΔT_m (K)
No compound	51.5 ± 0.3	N/A
91395	51.3 ± 0.3	− 0.2 ± 0.6
61610	51.4 ± 0.1	− 0.1 ± 0.4
7524	51.3 ± 0.4	− 0.2 ± 0.7
91396	51.3 ± 0.3	− 0.2 ± 0.6
260594[a]	N/D	N/D
146771[a]	N/D	N/D
202386[a]	N/D	N/D
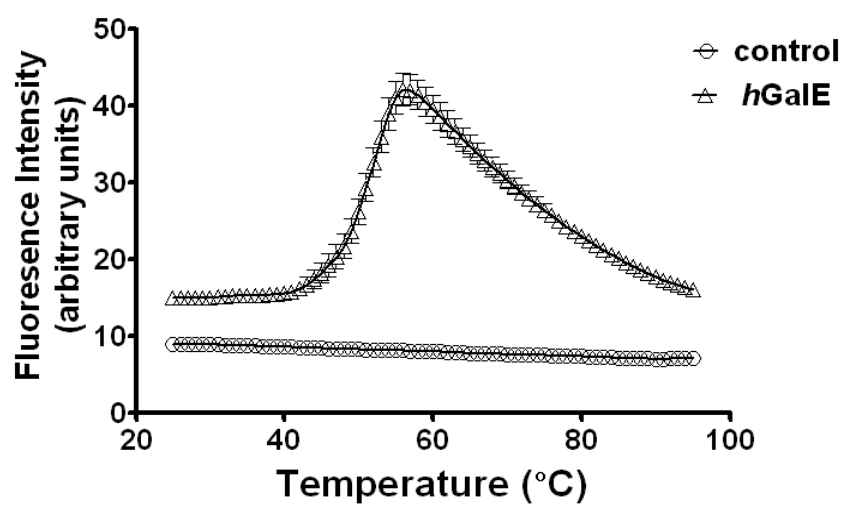
Figure S2.1. Thermal scanning fluorimetry of *Hs*GalE. 5 μM *Hs*GalE in 10 mM HEPES-NaOH, pH 8.8, 1% (v/v) DMSO, 5× Sypro orange showed a clear melting curve resulting in a T_m of 51.5±0.3 °C.

Chapter 3: CrystalDock: Leveraging Known Structural Information for Fragment-Based Drug Design

Abstract

We present a novel algorithm called CrystalDock that analyzes a molecular pocket of interest and identifies potential binding fragments. The program first identifies groups of pocket-lining receptor residues (i.e., microenvironments) and then searches for geometrically similar microenvironments present in publically available databases of ligand-bound experimental structures. Germane fragments from the crystallographic or NMR ligands are subsequently placed within the novel binding pocket. These positioned fragments can be linked together to produce ligands that are likely to be potent; alternatively, they can be joined to an inhibitor with a known or suspected binding pose to potentially improve binding affinity.

To demonstrate the utility of the algorithm, CrystalDock is used to analyze the principal binding pockets of influenza neuraminidase and *Trypanosoma brucei* RNA editing ligase 1, validated drug targets in the fight against pandemic influenza and African sleeping sickness, respectively. In both cases, CrystalDock suggests modifications to known inhibitors that may improve binding affinity.

Introduction

The computational prediction of molecular recognition is important in modern drug discovery. Computational medicinal chemists seek to answer two classes of questions. First, given an NMR, crystallographic, or homology-modeled receptor structure, can novel ligands be identified *in silico*? Second, given a ligand with a known or suspected receptor-binding pose, what chemical modifications can improve potency?

Several techniques have been developed to answer these questions. When a project requires that thousands of potential ligands be evaluated, techniques that favor speed over accuracy, such as computational docking, are often employed. These programs typically sample multiple ligand conformations and attempt to fit, or dock, each conformation into a known binding-pocket structure. Receptor–ligand interactions are subsequently evaluated using a fast scoring function to estimate binding affinity. Unfortunately, because these algorithms are optimized for speed, they are far less accurate than most experimental techniques (92). Any single docking prediction is untrustworthy; the objective of a docking study is rather to produce an enriched pool of potential binders by docking many ligands (e.g., compounds with diverse scaffolds or analogues of a known inhibitor) and considering only the top predicted binders for subsequent experimental or computational evaluation.

Other computational techniques that are similarly designed to confirm or refute molecular recognition have been optimized for accuracy rather than speed. These techniques utilize molecular dynamics (MD) simulations to probe not only the

possible conformations of the ligand, but also the conformations of the protein, water molecules, and other molecular elements that may contribute to binding (93-97). Because an accurate prediction of the binding free energy depends on adequate sampling of these many possible conformations, MD-based techniques can easily require thousands or even tens of thousands of computer hours. If sampling is adequate, however, these methods are often more accurate than the scoring functions used by computer-docking programs.

Here, we present a computer program called CrystalDock that takes a different approach to the computational identification of molecular recognition. Our technique is somewhat unique (98) in that it makes direct use of crystallographic and NMR structures from the Protein Data Bank (PDB) (65) to generate a molecular-recognition database that is used to place molecular fragments into binding pockets of interest. In this paper, we describe the CrystalDock algorithm and use the program to generate novel potential inhibitors of influenza neuraminidase and *Trypanosoma brucei* RNA editing ligase 1 (*Tb*REL1), enzymes critical to the etiological agents of pandemic influenza and African sleeping sickness, respectively.

CrystalDock is open source and Python implemented, making it easily editable, customizable, and platform independent. The program has been tested on Ubuntu Linux 10.04.1 LTS, Mac OS X 10.6.6, and Windows XP using Python versions 2.6.5, 2.6.1, and 2.6, respectively, together with NumPy/SciPy versions 1.3.0/0.7.0, 2.0.0.dev-3071eab/0.10.0.dev, and 1.6.1rc1/0.9.0, respectively. A copy can be obtained free of charge from http://www.nbcr.net/crystaldock.

Materials and Methods:

Creating a Database of Molecular Fragments and Microenvironments

A search of the PDB in February 2011 revealed 50 424 structures with bound

ligands. To generate a database of molecular fragments, models of all these receptor–

ligand complexes were downloaded. Several classes of undesirable ligands were

subsequently identified: ligands that contained multiple rotamers; ligands that were

ribonucleic, deoxyribonucleic, or amino acids; and small molecules that were not

sufficiently close to any potential receptor. After the receptor–ligand complexes

containing only these undesirable ligands were removed, 43 327 complexes containing

202 584 ligands remained for subsequent analysis. A representative example of such a

ligand positioned in a receptor binding pocket is shown in Figure 3.1A.

Each of these 202 584 ligands was subsequently fragmented into its constituent

molecular parts. All bonds between heavy atoms not belonging to the same ring were

identified; "cutting" along these bonds produced multiple molecular fragments from

each ligand model (Figure 3.1B). Any fragment with fewer than three heavy atoms

was merged with the neighboring fragment that had the fewest atoms.

While long-range electrostatic interactions certainly can influence fragment

binding, for many fragments, the predominant interactions required for molecular

recognition are with receptor atoms that immediately line the fragment-binding

pocket. Consequently, receptor microenvironments (i.e., groups of adjacent, pocket-

lining receptor residues) were identified by extending geometric rays, separated by 10°

in all directions, from each fragment atom out into space (Figure 3.1C). Whenever a

ray encountered a receptor residue, ray extension was terminated, and the residue was recorded. In addition, if a ray grew to 4 Å without encountering any receptor residue, it was similarly terminated. Testing confirmed that this process could effectively identify the receptor residues that line a fragment-binding pocket (Figure 3.1D).
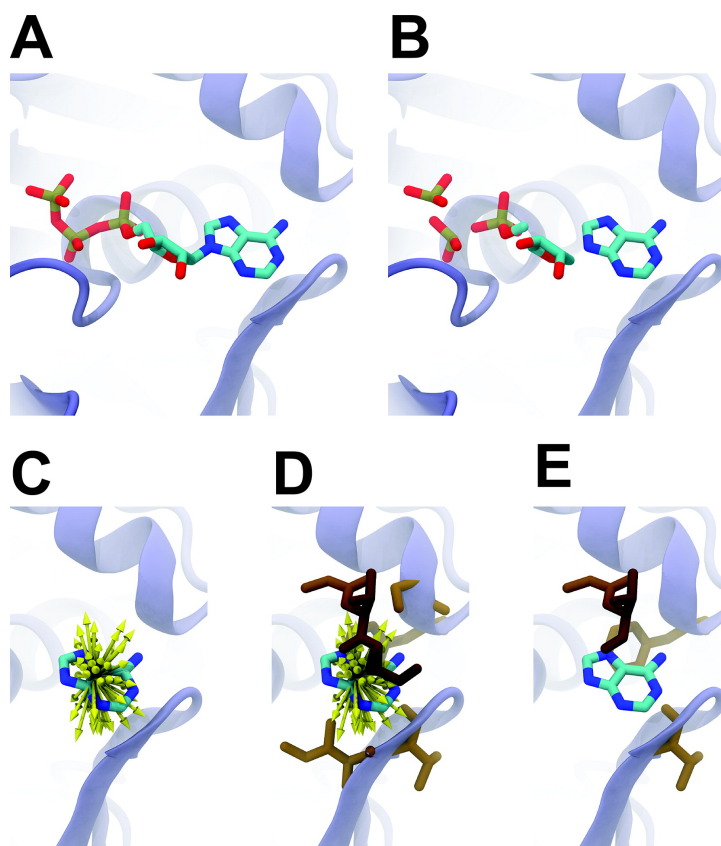
Figure 3.1. Schematic of the algorithms used to generate a database of microenvironments from available Protein Data Bank (PDB) structures: (A) 43 327 receptor–ligand complexes were identified with 202 584 total ligands; (B) each ligand was fragmented into its constituent molecular parts; (C) geometric rays, separated by 10° in all directions, were extended from each fragment atom out into space; (D) these rays were used to identify microenvironment receptor residues; and (E) a ligand–receptor distance cutoff was implemented. The cutoff was gradually scaled back from 4 Å to 0 Å, and receptor residues beyond the cutoff were discarded at every step. In this way, multiple microenvironments were identified for each molecular fragment. Subsequently, only those microenvironments with 3, 4, and 5 receptor residues (823 460 in total) were considered.

Often, the number of residues lining a binding pocket was too large. To make the number of microenvironments more manageable for future search, a ligand–

receptor distance cutoff was implemented. The cutoff was gradually scaled back from 4 Å to 0 Å, and receptor residues beyond the cutoff were discarded at every step. In this way, multiple microenvironments were identified for each molecular fragment. Subsequently, only those microenvironments with 3, 4, and 5 receptor residues (823,460 in total) were considered (Figure 3.1E).

Characterizing a New Binding Pocket

Characterization of a new binding pocket begins when the user provides three-dimensional coordinates identifying the pocket location. CrystalDock then sends out rays as described above to identify the pocket-lining receptor residues. All combinations of 3, 4, and 5 active-site residues are subsequently considered (Figure 3.2A).

CrystalDock then searches through the database of predefined microenvironments in an attempt to find geometric matches (Figure 3.2B). A root-mean-square deviation (rmsd) alignment is used to judge microenvironment similarity. Those aligned microenvironments from the database that are judged to be geometrically similar to the identified microenvironments of the binding pocket are saved for further analysis. Rather than requiring exact amino-acid matches, the user can also instruct the program to consider chemically similar amino acids to be equivalent, according to a predefined similarity matrix (see Table S3.1 in the Supporting Information), based on BLOSUM62 (2). (See the Supporting Information for more details.)
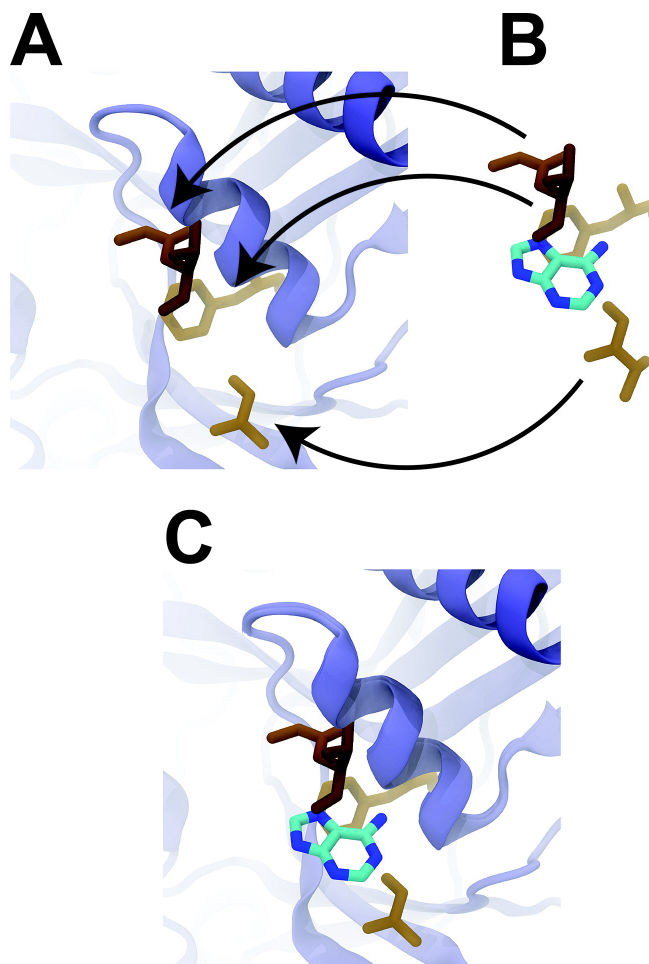
Figure 3.2. Schematic of the algorithm used to position binding fragments into a pocket of interest: (A) CrystalDock sends out rays to identify the receptor residues that line the binding pocket, and, subsequently, all combinations of 3, 4, and 5 lining residues are considered (a representative combination of 3 residues is shown); (B) CrystalDock searches through the database of predefined microenvironments in an attempt to find geometric matches; (C) Although the root-mean-square deviation (RMSD) alignment considers only receptor residues (i.e., the residues of the microenvironments), the structures include models of the original ligand fragments as well; rmsd alignment positions these molecular fragments within the binding pocket of interest.

Although the RMSD alignment considers only receptor residues (i.e., the residues of the microenvironment), the structures also include models of the original ligand fragments; RMSD alignment positions these molecular fragments within the binding pocket of interest. Once those positioned fragments that have steric clashes

with receptor residues have been discarded, a final set of molecular fragments ideally positioned within the pocket of interest remains (Figure 3.2C).

Independent-Trajectories Thermodynamic Integration Calculations

Independent-trajectories thermodynamic integration (IT-TI) (99) was used to predict the binding energies of both a known and a predicted *Trypanosoma brucei* RNA editing ligase 1 (*Tb*REL1) inhibitor. The initial model creation, minimization, and equilibration steps of the receptor-bound **V2** (100) simulation have been described previously (82). An identical protocol also was used to set up the simulation of a receptor-bound CrystalDock inhibitor.

For the solvated-ligand simulations, the same ligand parameters were used. The ligands were immersed in a TIP3 (74) water box extending 10 Å beyond the ligand atoms in all three dimensions. Na$^+$ counterions were added as needed to ensure the electrical neutrality of the system. NAMD 2.7b1 (75) was used to subject the system to 15,000 steps of conjugate-gradient energy minimization.

Following this initial preparation, six independent thermodynamic-integration (TI) runs (101)—three with bound-ligand annihilation and three with solvated-ligand annihilation—were performed for each ligand. Twenty-one (21) lambda (λ) points were used for both the receptor-bound and solvated simulations; as λ decreased from 1.00 to 0.00, in decrements of 0.05, the electrostatic and van der Waals interactions between the ligand and protein were slowly turned off (tiVdwLambdaEnd = tiElecLambdaStart = 0.5). Each simulation ran for 2 ns. Average dE/dλ values over the last 1 ns of each simulation were plotted against λ; the area under the curve, calculated

using the composite trapezoidal rule, as implemented in SciPy (102), was taken to be the free energy of the alchemical transformation. The free energies of the transformations in which the protein-bound and solvated ligands vanish are denoted $\Delta G_{\text{protein}}$ and ΔG_{water}, respectively.

When necessary, a restraint was applied to the protein and ligand to maintain the correct positional orientation. This was especially necessary at low λ values, where the interactions between the ligand and protein were almost absent. Since the protein-bound ligand was often confined to an artificially limited volume by these imposed constraints, it was necessary to correct the free energy accordingly. This was done using the formula employed by Lawrenz et al. (103). Specifically,

$$\Delta G_{corrected} = \Delta G_{protein} + RT ln \left(\frac{V_{pocket}}{1600 \text{ Å}^2} \right)$$

where $\Delta G_{\text{corrected}}$ is the corrected free energy, $\Delta G_{\text{protein}}$ the free energy prior to correction, R the gas constant, T the temperature, and V_{pocket} the volume sampled by the ligand during the simulations. The free energy of binding (ΔG_{bind}) was ultimately calculated by simply subtracting $\Delta G_{\text{corrected}}$ from ΔG_{water}.

Since each TI run was performed in triplicate, there were three $\Delta G_{\text{corrected}}$ and three ΔG_{water} values, yielding nine possible estimates of ΔG_{bind}. Histograms of these nine values were generated by simple binning; average predicted binding energies are also reported.

Results and Discussion

Here, we present a novel algorithm called CrystalDock that identifies molecular fragments likely to bind pockets of interest. First, CrystalDock identifies relevant microenvironments (i.e., groups of adjacent pocket-lining amino-acid residues) within the user-specified binding pocket (Figure 3.2A). The program then searches through a database of microenvironments derived from ligand-bound crystallographic and NMR structures deposited in the PDB (Figure 3.2B) for similar microenvironments, and positions the associated small-molecule fragments from the database within the binding pocket of interest (Figure 3.2C). The identified fragments can be joined to create novel ligands or can be fused to ligands with known binding poses to enhance potency.

CrystalDock-Suggested Oseltamivir Modifications

To demonstrate the utility of CrystalDock, we first used the program to analyze the principal binding pocket of influenza neuraminidase (Figure 3.3). Neuraminidase is a useful initial validation system because it is thoroughly represented among PDB structures; a search of the PDB for the terms "neuraminidase" and "sialidase" returned 222 structures with ~70 unique ligands that bind in the sialic acid pocket. Neuraminidase is an important drug target in the fight against influenza, including virulent strains such as those that have recently caused the H1N1 and H5N1 pandemics. Many neuraminidase inhibitors have been approved by the FDA or are otherwise progressing through clinical trials (104).
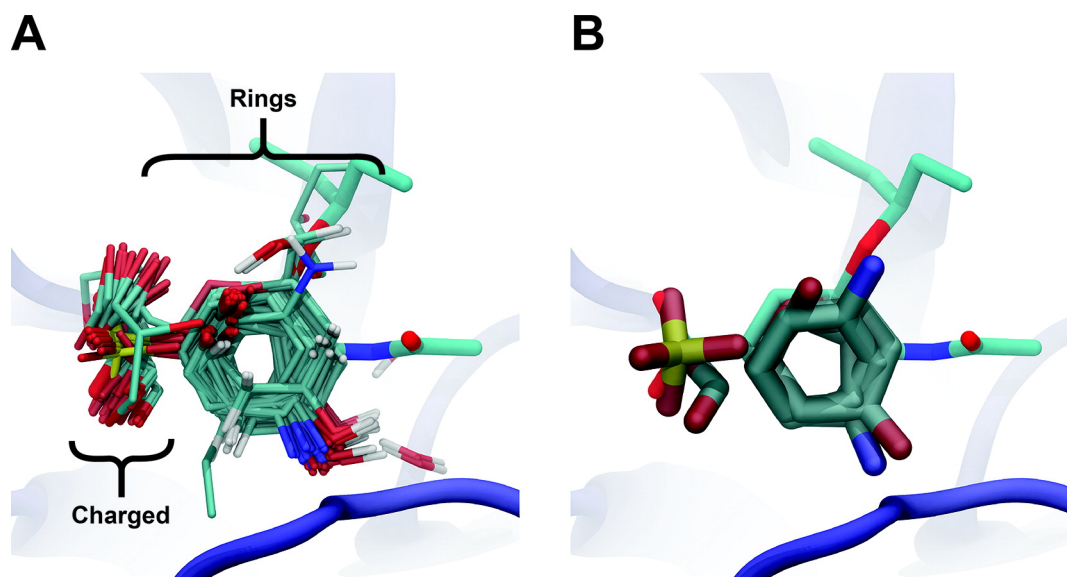
Figure 3.3. The results of an influenza neuraminidase CrystalDock run, shown together with the crystallographic pose of oseltamivir, a known inhibitor. In panel (A), CrystalDock identified many ringed fragments derived from several known neuraminidase inhibitors; interestingly, the program also placed a sulfate ion near the location of the charged oseltamivir carboxylate group. Panel (B) is the same as panel (A), but with only selected positioned fragments shown for the sake of simplicity.

CrystalDock-identified binding fragments came from 95 distinct PDB structures representing 39 unique ligands. As expected, most of the identified ringed fragments were derived from known neuraminidase ligands, including peramivir (e.g., PDB ID 3K37) (105), oseltamivir (e.g., PDB ID 3K3A) (105), zanamivir (e.g., PDB ID 3B7E) (106), sialic acid (e.g., PDB ID 2C4A), DANA (e.g., PDB ID 2F25), and other experimental inhibitors. However, positioned fragments were not derived exclusively from neuraminidase structures; a single fragment was also obtained from pentaethylene glycol bound to D-lactate dehydrogenase (PDB ID 3KB6) (107), confirming that CrystalDock is able to identify binding fragments from even distantly related proteins.

Interestingly, CrystalDock placed a sulfate ion near the location of the charged oseltamivir carboxylate group (Figure 3.3B). The Figuridea of substituting this

carboxylate group with a sulfonate is interesting because at least one known

neuraminidase inhibitor has a sulfonate group so positioned (2-[*N*-

cyclohexylamino]ethane sulfonic acid, PDB ID 2VW2) (108), and the binding pose of

another inhibitor is also thought to position a sulfonate at this same location (109). In

addition, many researchers have demonstrated that inhibitors with comparable

phosphonate groups are also potent (110-113).

CrystalDock-Suggested Modifications of a TbREL1 Naphthalene-Based Inhibitor

Because many neuraminidase structures with bound ligands have been

deposited in the PDB, the example above, while useful as a proof of concept and for

method validation, does not demonstrate the full utility of CrystalDock. Ideally, the

program should be able to extract molecular fragments from multiple structurally and

even functionally diverse receptors. As a further demonstration, we used CrystalDock

to analyze the binding pocket of the adenylation domain of *Trypanosoma brucei* RNA

editing ligase 1 (*Tb*REL1), a protein with only one PDB-deposited crystal structure

bound to a single ligand (ATP) (114).

The results of the *Tb*REL1 CrystalDock run are shown in Figure 3.4, together

with **V2**, a known low-μM naphthalene-based inhibitor (100) docked into the *Tb*REL1

crystal structure using AutoDock Vina (115). When only the lower, buried portion of

the *Tb*REL1 active site was targeted, predicted binding fragments from 55 structures

representing 45 unique ligands were identified. Interestingly, the CrystalDock-

positioned fragments can be generally clustered into three groups. The first group

contains a single sulfate that CrystalDock positioned near the predicted pose of

a **V2** sulfonate group, explaining, in part, the potency of the naphthalene-based

inhibitor. The second group is comprised principally of aromatic fragments that are

generally in the same region and plane as the **V2** naphthalene group. The third group,

which is represented by several mostly hydrophobic fragments, does not correspond to

any **V2** substructure, suggesting a possible route for improving potency.
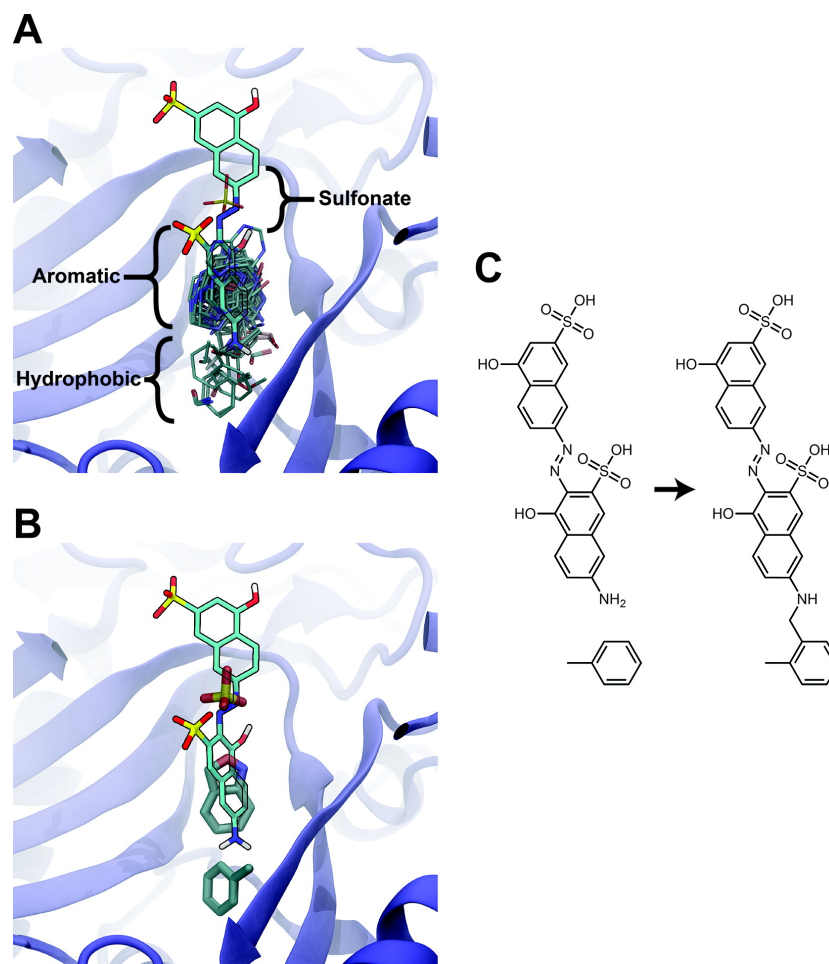
Figure 3.4. Results of a *Tb*REL1 CrystalDock run, shown together with **V2** (faintly outlined), a low-μM inhibitor docked into the crystallographic *Tb*REL1 active site. In panel (A), the CrystalDock-positioned fragments can be generally clustered into three groups: a single sulfate positioned near the predicted pose of a **V2** sulfonate group, aromatic fragments that are generally in the same region and plane as the **V2** naphthalene group, and hydrophobic fragments not corresponding to any **V2** substructure. Panel (B) is the same as panel (A), but with only selected positioned fragments shown for the sake of simplicity. Panel (C) shows that, serendipitously, the position of a toluene fragment was ideal for chemical linking to the Vina-docked **V2**.

Serendipitously, the CrystalDock poses of two of these group-three hydrophobic fragments, toluene fragments derived from two unique inhibitors of P38 mitogen-activated protein kinases (PDB IDs 2ZB1 (116) and 3LHJ (117)), were ideal for fragment addition to the Vina-docked**V2** via an intermediary methyl linker (Figure 3.4C). These toluene fragments occupied a small pocket at the buried end of the *Tb*REL1 active site that, to our knowledge, has not been previously exploited for drug design.

To test if a CrystalDock-inspired **V2** + toluene composite compound would have improved binding affinity over **V2**, we employed a computational technique known as independent-trajectories thermodynamic integration (IT-TI) (99) to predict ligand binding energies (101). IT-TI is far more computationally demanding than high-throughput methods for estimating binding affinity (e.g., computer docking programs); using the protocol described in the Materials and Methods section, calculating the binding energy of a single *Tb*REL1 ligand required ~25 000 CPU hours. In contrast, a simple docking run using AutoDock Vina (115) takes only a few minutes on a single processor, a speed up of ~5 orders of magnitude. However, if conformational sampling is adequate, IT-TI is often more accurate than docking scoring functions.

Six TI runs were executed for each system: three in which the protein-bound ligand was annihilated, and three in which the solvated ligand was annihilated. From these six TI runs, nine binding-energy estimates were calculated. Histograms of these nine values were generated by simple binning and are shown in Figure 3.5. The

predicted binding affinity of **V2** was -8.4 ± 0.5 kcal/mol, which correlates well with

the experimentally measured IC_{50} value of 1.53 μM (100). The predicted binding

energy of the new composite compound was -10.7 ± 0.9 kcal/mol, representing a 2.3

kcal/mol improvement.
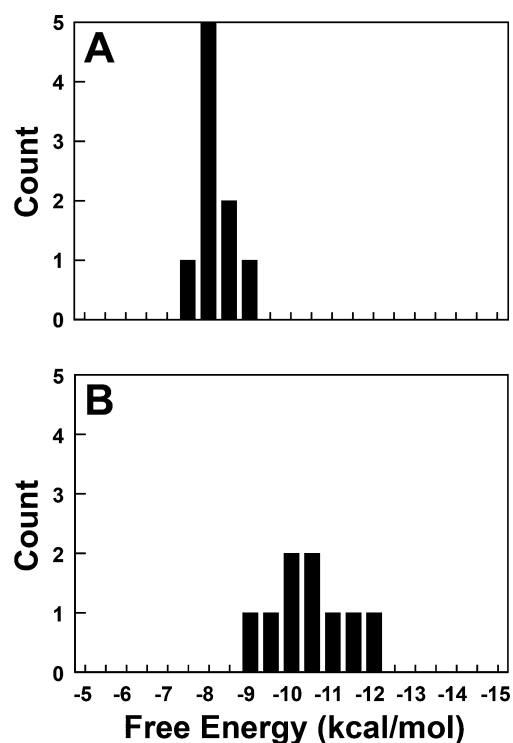
Figure 3.5. Histograms of the predicted binding energies generated using IT-TI: (A) the predicted binding energies derived from the **V2** simulation and (B) the predicted binding energies derived from the **V2** + toluene composite compound. Bin sizes of 0.5 kcal/mol were used, and the *x*-axis in each is ordered by increasing potency (i.e., decreasing predicted binding energy).

A one-tailed, homoscedastic *t*-test was subsequently used to determine if the

difference in the average predicted binding energies of **V2** and the composite

compound was statistically significant. The associated *p* value was 2.7×10^{-6},

suggesting that the CrystalDock-inspired ligand represents a genuine improvement,

possibly potent in the nanomolar (nM) range.

If these predictions are confirmed experimentally, this optimized compound may represent a significant contribution in the fight against African sleeping sickness, a deadly parasitic infection that currently afflicts as many as 70 000 sub-Saharan Africans and threatens an additional 50 million (37). In a previous study, we demonstrated that **V2** fails to kill whole-cell parasites, despite its confirmed *Tb*REL1 inhibition (100), perhaps because the compound is too negatively charged to easily cross biological membranes. However, the hydrophobic toluene group may make membrane crossing more feasible. The calculated logarithmic *P* (Log*P*) values of the electrically neutral forms of **V2** and the new composite compound are 0.305 and 2.940, respectively, supporting this theory.

The new composite compound is also a promising lead for other reasons. It is only one hydrogen-bond acceptor away from satisfying Lipinski's Rule of Five (118), a common measurement of druglikeness; the hydroxyl group connected to the aminonaphthalene forms an important hydrogen bond with the receptor, but the other hydroxyl group may be a good candidate for elimination. The CrystalDock-predicted toluene pose is also ideally positioned to react with the **V2** amine, assuming that the Vina-docked pose of **V2** is correct; we expect that the composite compound can be easily synthesized by reacting **V2** with 1-(bromomethyl)-2-methylbenzene via an amino-dehalogenation reaction (119).

Acknowledgements

Chapter 3 is a minimally modified reprint of the material appearing in Durrant JD, Friedman AJ, and McCammon JA. CrystalDock: A Novel Approach to Fragment-Based Drug Design. *J Chem Inf Model* 2011, 51(10), 2573-2580. I was the second author of this paper. Additional work in this chapter will be included in an upcoming paper in which I will be first author.

Supporting Information

Aligning Microenvironments from the Database to an Active-Site of Interest

In order to identify microenvironments from the database that are similar to those of the active site, a computationally efficient, multi-tiered filtering process is employed. First, all database microenvironments containing the same protein residues as the active-site microenvironment are identified. If desired, similar amino acids, as judged by the similarity matrix shown in Table S1, can be considered equivalent.

Next, the potential matches from the database must be evaluated for geometric similarity. For both the active-site microenvironment and the potential matches, the maximum distance between any two atoms (i.e., the span of the microenvironment) is calculated. Geometrically similar microenvironments should have similar spans; any potential match whose span differs by more than a user-specified number of angstroms (2.0 Å by default) is discarded.

For the remaining potential matches, a more sophisticated geometric comparison is subsequently employed. A unique "fingerprint" of each microenvironment is generated by creating a sorted list of all the pairwise distances between all alpha carbons. Two fingerprints can be compared by comparing each of the nth entries in the corresponding lists. If the nth entry of any of the database-microenvironment fingerprints differs substantially from the corresponding entry in the active-site fingerprint, the database microenvironment is discarded. By default, differences of up to 2.0 Å are permitted.

Next, a more computationally demanding geometric comparison is employed. An RMSD alignment is performed on the alpha carbons of the active-site microenvironment and each of the potential matches from the database. Alpha carbons are considered equivalent for alignment purposes if they belong to similar protein residues, as judged by the similarity matrix shown in Table S1. Where multiple equivalent alpha carbons exist, the ones in closest proximity at each step of the minimization are selected. Any database microenvironment with an RMSD greater than 2.5 Å (by default) following alignment is subsequently discarded.

It is possible for microenvironments to have similar alpha-carbon configurations but markedly different side-chain configurations. The next step in the filter seeks to verify that the side chains of equivalent microenvironment protein residues point in the same general direction. For each pair of aligned, equivalent microenvironment protein residues, an alpha-carbon coordinate is calculated by averaging the coordinates of the alpha carbons of the two residues. Additionally, the coordinates of selected side-chain atoms (see Table S2) are used to represent the location of each side chain. If the (sidechain)-(alpha-carbon)-(side-chain) angle is greater than a user-specified value (100° by default), the side-chain orientation of the database and active-site residues is judged to be different, and the database microenvironment is discarded.

As a final check of geometric similarity, a second RMSD alignment is performed, using the coordinates of the first alignment as a starting point. This time, however, all heavy atoms are considered rather than just the alpha carbons. As equivalent but nonidentical protein residues do not necessarily have the same number of side-chain atoms, atoms with equivalent names are aligned, and all other atoms are ignored. Any database microenvironment that differs by more than 1.5 Å RMSD (by default) from the activesite microenvironment following alignment is subsequently discarded.

The remaining database microenvironments are judged to be geometrically similar to that of the active site. As a final check, the atoms of the molecular fragment present in the aligned database microenvironment, which up to this point have been

ignored, are now compared to the atoms of the active site. If any of these atoms come within 2.0 Å (by default) of each other, the database microenvironment is discarded due to steric clashes.

Ranking the Aligned Microenvironments

A crude ranking algorithm seeks to prioritize which matches are most reliable. While we recommend rescoring with more advanced scoring functions for better results, the default ranking may still prove helpful. First, database microenvironments with the same protein residues as the active-site microenvironment take precedence over those that contain only similar residues, as judged by Table S1. If two matching database residues are equivalent under this criterion, the database microenvironments containing a greater number of protein residues are given precedence. Finally, all other things being equal, database microenvironments with lower RMSD to the active-site microenvironment are given priority.

Table S3.1. Rather than requiring exact amino-acid matches, the user can also instruct CrystalDock to consider chemically similar amino acids to be equivalent. Amino acids (in bold) that are chemically similar are marked with X's.

	C	S	T	A	V	Q	N	D	E	H	R	K	M	I	L	F	Y	W
C	X	X	X															
S	X	X	X															
T	X	X	X															
A				X	X													
V				X	X								X	X	X			
Q						X	X											
N						X	X											
D								X	X									
E								X	X									
H										X	X	X				X	X	X
R										X	X	X						
K										X	X	X						
M					X								X	X	X			
I					X								X	X	X			
L					X								X	X	X			
F										X						X	X	X
Y										X						X	X	X
W										X						X	X	X

Table S3.2. To determine the orientation of protein side chains, representative coordinates are chosen corresponding to the side-chain atoms indicated. The atom name is given across the top, and the residue name is given on the left.

	CB	CZ	CG	SG	CD	NZ	CE	OG
ALA	X							
ARG		X						
ASN			X					
ASP			X					
CYS				X				
GLU					X			
GLN					X			
HIS			X					
ILE	X							
LEU			X					
LYS						X		
MET							X	
PHE			X					
SER								X
THR	X							
TRP			X					
TYR			X					
VAL	X							

Chapter 4: The Marine Cyanobacterial Metabolite Gallinamide A is a Potent and Selective Inhibitor of Human Cathepsin L

Abstract

A number of marine natural products are potent inhibitors of proteases, an important drug target class in human diseases. Hence, marine cyanobacterial extracts were assessed for inhibitory activity to human cathepsin L, a cysteine protease involved in several disease conditions. Herein, we have shown that gallinamide A potently and selectively inhibits the human cysteine protease, cathepsin L. With 30 minutes of preincubation, gallinamide A displayed an IC_{50} of 5.0 nM, and kinetic analysis demonstrated an inhibition constant of $k_i = 9009 \pm 262$ M^{-1} s^{-1}. Preincubation-dilution and activity-probe experiments revealed an irreversible mode of inhibition, and comparative IC_{50} values display a 28- to 331- fold greater selectivity toward cathepsin L than closely related human cysteine cathepsins V or B. Molecular docking and molecular dynamics simulations were used to determine the pose of gallinamide in the active site of cathepsin L. These data resulted in the identification of a pose characterized by high stability, a consistent hydrogen bond network, and the reactive Michael acceptor enone functional group of gallinamide A positioned near the active site cysteine of the protease, leading to a proposed mechanism of covalent inhibition. These data reveal and characterize the novel activity of gallinamide A as a potent

inhibitor of human cathepsin L.

Introduction

Marine cyanobacteria are exceptionally rich in biologically active natural products (120). From a biochemical perspective, their metabolites are highly diverse, often deriving from mixtures of nonribosomal peptide synthetase (NRPS), polyketide synthase (PKS), terpene, and sugar biosynthetic pathways, and are commonly further decorated with halogen atoms, methyl groups, and interesting oxidations. As a result, they have been actively investigated for their therapeutic potential for a number of years, especially for anticancer activity. One such recently approved agent for anaplastic large cell lymphoma and Hodgkin's lymphoma, brentuximab vendroitin, was inspired by the marine cyanobacterial metabolite dolastatin 10 (121, 122). In recent years, the biological evaluation of marine-derived natural products has broadened to include inflammation, infectious and parasitic diseases, and neurological diseases (123).

In this latter regard, an emergent trend in the pharmacological mechanism of action of cyanobacterial natural products is that many are potent inhibitors of various classes of proteases (124, 125). Proteases have been implicated in the pathogenesis of many human diseases, including cancer (28, 126), neurological disorders such as Alzheimer's Disease (127, 128), and parasitic diseases (129); thus the therapeutic modulation of proteolytic activity offers an attractive potential treatment modality. However, with myriad proteases and many potential therapeutic applications,

discovery of agents with selectivity for specific proteases is crucial to the development of truly useful pharmaceuticals in this class.

Whereas freshwater cyanobacteria have yielded a number of protease inhibitors (130), their marine relatives represent an under-explored resource for modulators of this enzyme class. Hence, we have initiated a program to survey marine cyanobacterial extracts, fractions and newly isolated pure compounds for interesting profiles of protease inhibition, with a special focus on enzymes in the cysteine cathepsin and proteasome classes. We have recently reported the structures of the carmaphycins, low nanomolar epoxyketone proteasome inhibitors from the Curaçao cyanobacterium *Symploca* sp., and previously had identified the depsipeptide symplocamide A as a potent serine protease inhibitor (125, 131).

Our recent efforts in this regard have focused on the human cysteine cathepsin L protease, an important lysosomal endopeptidase with exceptionally high proteinase activity. Aside from its traditional role in protein degradation, cathepsin L is responsible for many specialized roles that make it an interesting target for drug discovery. It is upregulated in multiple cancer cell types, and has been strongly implicated in bone resorption, bone pit formation, and invasion of bone tissue by osteoclasts due to its high level of secretion and efficient hydrolysis of bone matrix proteins (132). Multiple studies have shown significant reduction in tumor invasiveness and metastasis with treatment of pan cysteine protease or selective cathepsin L inhibitors (28). Furthermore, related cysteine proteases have been identified and targeted in various infectious diseases, including malaria, leishmaniasis,

trypanosomiasis and others (133). Finally, recent evidence has mounted to elucidate murine cathepsin L's role in proneuropeptide processing, with knockout (KO) and siRNA studies indicating a particularly important role in the production of the dynorphins and neuropeptide Y (134, 135).

Despite the multitude of disease implications associated with cathepsin L, few selective inhibitors have been described, and even fewer have appropriate pharmaceutical properties for potential clinical application. Herein, we report that evaluation of cyanobacterial extracts led to the identification of gallinamide A (**1**) as a potent and selective inhibitor of human cathepsin L, and thus provides a lead structure for developing agents with highly desired subtype selectivity within the cysteine proteases. Thus, this study describes the re-isolation and identification of gallinamide A, inhibitory potency to cathepsin L and related cysteine proteases, kinetic inhibition properties, and analyses of molecular docking to cathepsin L that indicates a Michael addition-based inhibition as supported by biochemical data. The molecular features of gallinamide A will assist future structure-based optimization efforts for effective inhibitors of human cathepsin L and members of the cysteine cathepsin protease family.
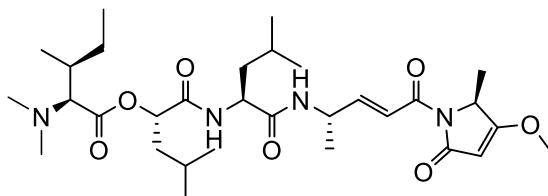
Gallinamide A (**1**)
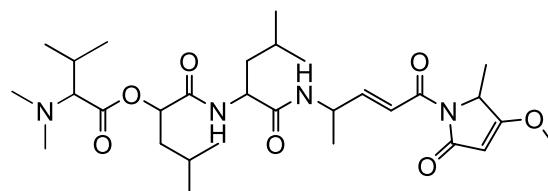
Gallinamide B (**2**)

Scheme 4.1. Gallinamide Structures

Results and Discussion

Gallinamide A isolation and structure determination.

Screening efforts of fractionated extracts from marine cyanobacteria for modulation of human cathepsin L activity identified fractions active for inhibition of the enzyme. A fraction eluting with 4:6 hexanes/EtOAc from the collection of a red-tipped *Schizothrix* sp. described by Linington et al. (136) showed 97% inhibition of cathepsin L at 3 μg/mL, and was subsequently fractionated by solid phase extraction (SPE) to produce eight subfractions. The subfractions eluting with 4:6 hexanes/EtOAc and 2:8 hexanes/EtOAc showed 99% and 99% inhibition, respectively, of cathepsin L at 3 μg/mL and were further fractionated by RP-HPLC to give five compounds, E45A-E45E. Compound E45A was inactive against cathepsin L, while compounds E45B, E45C, E45D and E45E displayed 97-99% inhibition of cathepsin L at 3 μg/mL. The major component, E45C, yielded a mass of 0.7 mg (0.14% of the crude extract), while

the minor constituents E45B, E45D, and E45E yielded 0.33 mg, 0.4 mg, and 0.38 mg, respectively (0.059%, 0.072%, and 0.068% of crude extract).

The HR-ESITOFMS spectrum of the major compound E45C gave an $[M+H]^+$ ion at m/z 593.3908, which was consistent with a molecular formula of $C_{31}H_{53}N_4O_7$ (calculated for $C_{31}H_{53}N_4O_7$, 593.3914). This psuedomolecular formula matched that of gallinamide A (**1**), a compound previously isolated from an adjacent fraction of the same collection with a reported $[M+H]^+$ ion at m/z 593.3907 (136). ^{1}H NMR analysis was used to elucidate partial structural features, which matched the previously reported subunits of gallinamide A. Furthermore, tandem LC-MS/MS analysis provided structural data consistent with the planar structure of gallinamide A (Figure 4.1A). While the mass and NMR data were consistent with the identification of compound **1** as gallinamide A, these data were insufficient for complete stereochemical assignment. However, that this material derived from a fraction adjacent to that yielding the original source of gallinamide A (**1**), it is virtually certain to be of the same absolute configuration, thus completing the structure identification of this active cathepsin L inhibitor.
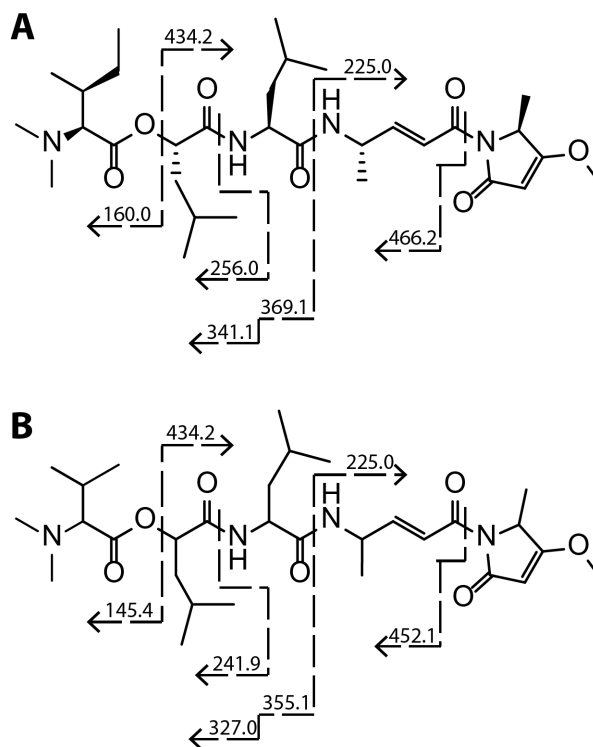
Figure 4.1. Fragmentation patterns for (A) gallinamide A (**1**) and (B) gallinamide B (**2**) by ESI-MS/MS.

Two minor compounds showed similar mass fragmentation patterns and NMR spectra to compound **1**, although they varied in their retention times from HPLC analysis. As they were not obtained in sufficient yield to allow detailed spectroscopic analysis, we predict that they represent stereo- or geometric isomers of gallinamide A. A third minor constituent, however, displayed a HR-ESITOFMS spectra with an [M+H]$^+$ ion at *m/z* 579.3756, representing a mass difference of 14 amu less than **1**. MS/MS fragmentation analysis showed that this mass loss was isolated to the *N*-terminal subunit. In gallinamide A this is an *N,N*-dimethyl isoleucine residue (Figure 1B). Previous reports of similar cyanobacterial depsipeptides, for example dolastatin 10 and symplostatin 1 (121), have been shown to occur as natural analogs containing either a terminal *N,N*-dimethyl isoleucine or *N,N*-dimethyl valine subunit. While NMR

analysis was not sufficiently robust to definitively assign this structure, the literature precedence and likely biosynthetic inclusion of a valine residue at this position strongly suggests the structure of gallinamide B (**2**) as shown (Scheme 4.1B). Due to the low yields of gallinamide B, and the remaining minor compounds, biochemical characterization of the inhibitory properties towards cathepsin L was limited to compound **1**.

Gallinamide A potently and selectively inhibits cathepsin L

Next, we demonstrated that gallinamide A (**1**) blocks the binding of the activity based probe (ABP) DCG-04, a biotin labeled derivative of the potent cysteine cathepsin inhibitor E64c (137). Unlike experiments measuring enzyme inhibition, ABP studies investigate the ability of a compound to compete with and block binding of a potent active site-directed probe. Human recombinant cathepsin L was preincubated with compound **1** for 30 minutes, followed by visualization of residual active enzyme by the labeled probe. Samples were subsequently run on gel electrophoresis and transferred to a Hybond Nitrocell membrane for detection with avidin-HRP and chemiluminescent substrate. Cathepsin L is a known target of DCG-04, and thus a reduction in band optical density revealed blockade of the active site. Compound **1** displayed a concentration-dependent inhibition of ABP labeling (Figure 4.2), with decreased labeling at 111 nM and complete loss of probe labeling at 333 nM. Interestingly, the concentration of enzyme in this assay was 100 nM; thus gallinamide A shows a high degree of efficiency in its inhibition, reducing binding at a 1:1 molar ratio of inhibitor and enzyme.
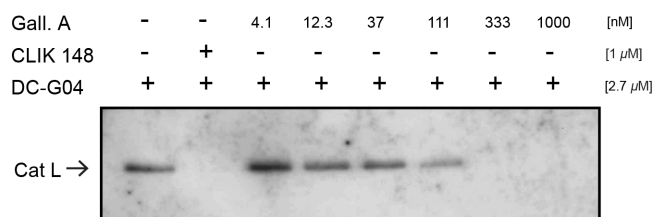
Figure 4.2. Representative blot of competitive activity-based probe labeling of cathepsin L. Gallinamide A showed reduced labeling at 111 nM and complete inhibition at 333 nM.

The potency of enzyme activity inhibition, commonly reported as an IC_{50} (inhibitory concentration for 50% activity), was found to be time-dependent in preincubation dose-response experiments. With immediate mixing of enzyme, substrate, and inhibitor, compound **1** inhibited human cathepsin L with an IC_{50} of 46.5 nM (95% CI = 40.4 nM to 53.5 nM). Following a preincubation of enzyme and inhibitor for 30 minutes prior to addition of substrate, compound **1** displayed increased potency, with an IC_{50} of 5.01 nM (95% CI = 4.18 nM to 6.02 nM) (Figure 3). Time-dependent inhibition is a hallmark of slow-binding inhibitors (138), a finding that directed the subsequent characterization of the mode of inhibition and binding affinity of gallinamide A.

Time-Dependent IC$_{50}$ Shift
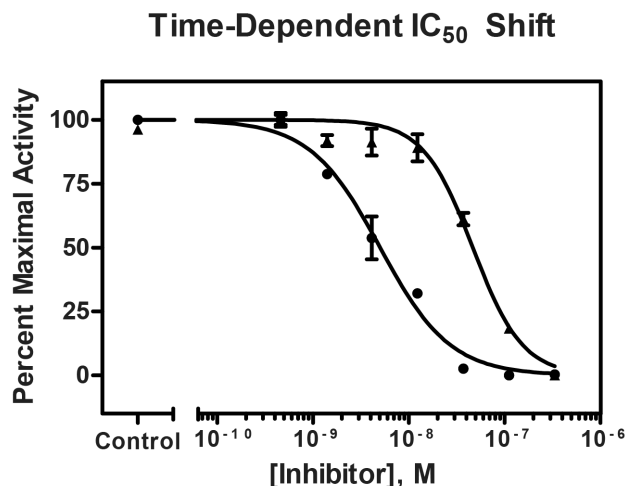
Figure 4.3. Dose response curves following gallinamide A preincubation with cathepsin L for 0 min (▲) and 30 min (●). The measured IC$_{50}$ following immediate mixing is 46.5 nM (95%CI = 40.4 nM to 53.5 nM), while 30 min preincubation results in an IC$_{50}$ of 5.01 nM (95%CI = 4.18 nM to 6.02 nM). IC$_{50}$ data are significantly different (p<0.0001).

To assess enzyme selectivity, compound **1** was tested for inhibitory activity against the highly homologous cysteine proteases cathepsin V (human) and cathepsin B (human). IC$_{50}$ values were obtained with and without preincubation of inhibitor and enzyme, as described for cathepsin L. A selectivity index was calculated as a ratio of the IC$_{50}$ of the assayed protease to that of cathepsin L for each incubation condition, summarized in Table 4.1. Interestingly, gallinamide A displays a 10-fold increase in potency for cathepsin L relative to cathepsin V without preincubation, and this metric increases to 28-fold after allowing the inhibitor to associate with the protease for 30 minutes. Cathepsin V is the most closely related human cysteine protease, sharing 77% sequence identity by BLAST analysis, and inhibitors capable of distinguishing between these two enzymes are currently limited. Selectivity is exaggerated for the more distantly related cathepsin B, sharing 30% sequence identity by BLAST, which shows a 331-fold higher IC$_{50}$ following 30 minutes of preincubation. While previous

studies have demonstrated moderate inhibition of murine cathepsin L as well as potent

inhibition of the *P. falciparum* falcipain proteases (139), a full panel of related

cysteine proteases was not able to be evaluated due to insufficient material in the

current study. To fully appreciate the selectivity of this inhibitor, additional enzymes

in this family should be evaluated.

Table 4.1. IC_{50} values and selectivity indices of gallinamide A (**1**) for cathepsins L, V and B. [A]95% confidence intervals: L(0)=0.0042-0.0060; L(30)=0.040-0.053; V(0)=0.416-0.500; V(30)=0.119-0.170; B(0)=3.69-4.71; B(30)=1.34-2.04

Enzyme	IC_{50}[A]		Selectivity Index	
	0 Min	30 Min	0 Min	30 Min
	μM			
Human Cathepsin L	0.046	0.005	1	1
Human Cathepsin V	0.456	0.142	10	28
Human Cathepsin B	4.167	1.653	91	331

Taken together, these experiments reveal human cathepsin L as a major target

of gallinamide A based on potency and selectivity. With increased focus on the

discovery of these inhibitors, gallinamide A represents an interesting lead compound

containing easily modifiable structural features. With two total syntheses now

published and a limited initial SAR study completed (139, 140), the gallinamide

structural class has great potential to yield analogs with improved potency, selectivity

and pharmaceutical properties.

Gallinamide A is an irreversible inhibitor of cathepsin L

As many slow-binding inhibitors work through covalent or tight-binding

interactions (138), the reversibility of the enzyme-inhibitor complex was subsequently

assessed. Irreversibility of gallinamide A (**1**) inhibition was determined using a preincubation-dilution experiment adapted from Copeland et al. (138), in which a concentrated sample of cathepsin L was incubated with concentrated compound **1** before a rapid dilution of the combined solution (Figure 4.4A). A preincubation of 30 minutes was used, and approximate levels of expected enzyme activity were obtained from the 30-minute preincubation dose-response experiment (Figure 4.4B). Following rapid dilution, a shift in inhibitor concentration from 10-fold IC_{50} to 0.1-fold IC_{50} value should cause a fully reversible inhibitor to immediately disassociate from the enzyme, resulting in activity levels equivalent to 90% of control. Alternatively, a slowly-reversible inhibitor will display a reduced reaction rate initially but with a gradual increase over time, whereas an irreversible inhibitor will show an initial rate of approximately 10% of the control condition, which persists over time. Immediately following the dilution, cathepsin L activity was measured at 12% of the vehicle control. The reaction rate remained linear and was unchanged for 2 hours following the dilution (Figure 4.4C). These data indicate that gallinamide is an irreversible inhibitor of the cathepsin L protease. This is congruent with previous studies showing that gallinamide A was a covalent, irreversible inhibitor of the related *Plasmodium falciparum* falcipain proteases based on kinetic model fitting (139).
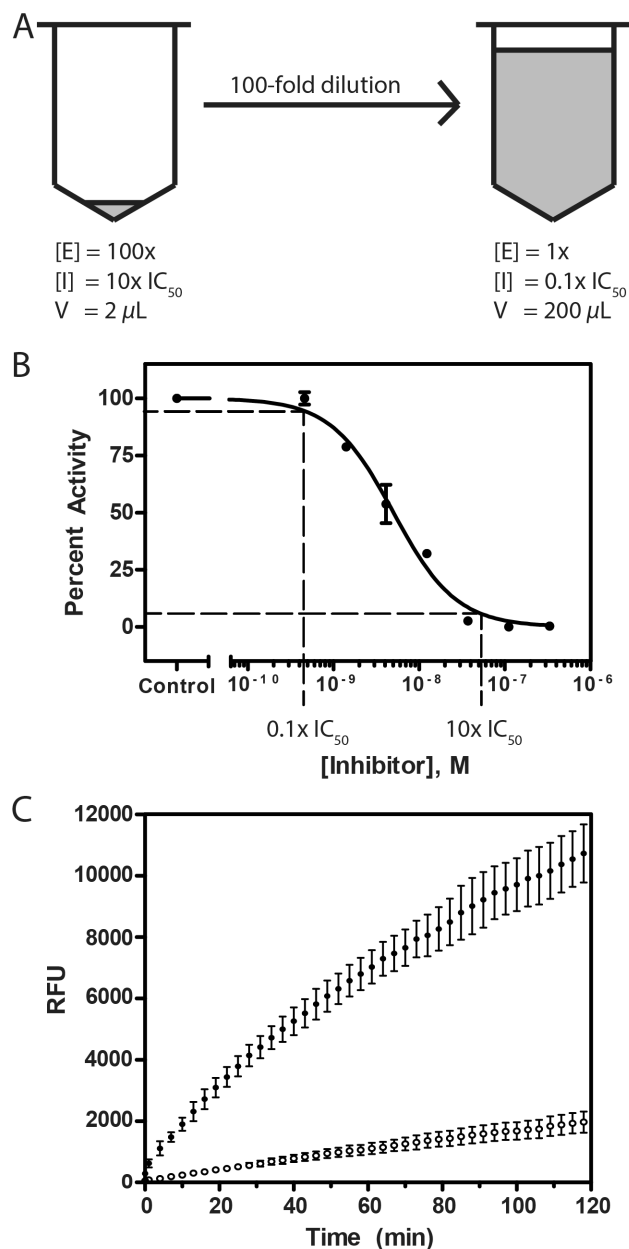
Figure 4.4. (A) A concentrated solution of enzyme and gallinamide A was incubated for 30 min and then then diluted. (B) The resulting shift in the enzymatic activity is based on the dose response curve. (C) The subsequent rate of the reaction was monitored for 2 hr, and comparison of initial reaction rates showed 12% of the activity with preincubation of gallinamide A (○) as compared to the control (●). The reaction rate was constant over the course of the two hour monitoring period, demonstrating an irreversible mode of inhibition.

To more accurately characterize the inhibitory potency of gallinamide A, the

substrate turnover kinetics of recombinant human cathepsin L were determined in the

presence of various concentrations of inhibitor (Figure 4.5). Progress curves were

fitted to a model of irreversible enzyme inactivation, and the calculated first order rate

constants k_{obs} were plotted against [I] (141, 142). The resulting relationship was linear,

and the second order rate constant represented by k_{obs}/[I] was determined to be 9009 $\pm$

262 M^{-1} s^{-1}. A k_{obs} value was not determined for the [I] = 1000 nM condition, as the

level of inhibition did not produce a progress curve capable of regression by the model.

These data are consistent with two key aspects of this study: 1) gallinamide A is a

potent and efficient inhibitor of human cathepsin L, and 2) the inactivation is

irreversible.
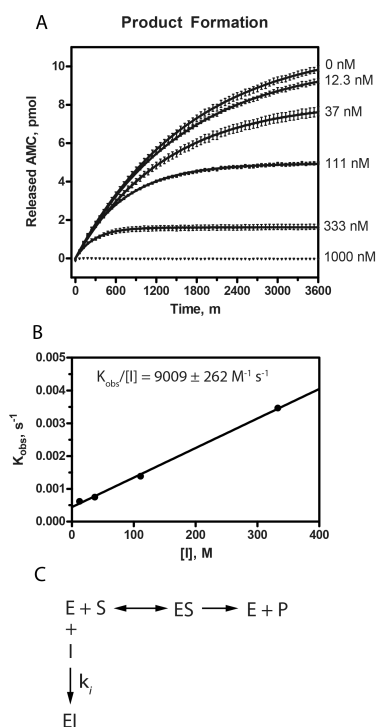
Figure 4.5. (A) Product formation from the turnover of substrate by cathepsin L in the presence of various concentrations of gallinamide A was monitored over time. The resulting plots were fitted to a model of irreversible inhibition, and the obtained k_{obs} values were plotted against [I]. (B) This produced a linear relationship, the slope of which represents the second order inhibition constant, k_i = 9009 $\pm$ 135.6 s^{-1} M^{-1}. (C) These data fit a simple model of irreversible enzyme inhibition.

Gallinamide A inhibits cathepsin L via Michael addition.

Leveraging the knowledge that gallinamide A binds irreversibly to cysteine cathepsins, a hypothesis that the reaction occurs via a Michael addition to the reactive gallinamide A enone was generated. In previous studies, it has been shown that the top scoring pose *in silico* is not necessarily the correct pose *in vitro* (143). With this in mind, docked poses in which the enone was positioned acceptably close (<4 Å) to the catalytic cysteine were selected. The induced fit docking protocol produced two different poses with similar predicted binding affinities, referred henceforth as 'top' (XPGscore -8.539) and 'bottom' (XPGScore -7.359) orientations (Figure 4.6A-B, respectively). In no circumstance did Glide position other potential reactive groups, specifically the two gallinamide A esters, near the catalytic cysteine.

Both of these poses possessed structurally desirable qualities for a reaction to occur. The 'top' orientation, in which the cyclic head group interacted with the S1' pocket, was positioned such that the enone carbonyl, which is negatively charged in the predicted reaction intermediate, is stabilized in an oxyanion hole by the Gln19 and positively charged His163 sidechains, as well as the backbone NH of Cys25. Additionally, intermolecular hydrogen bonds are evident throughout the ligand-enzyme interface, which may help stabilize the reactive complex. In the 'bottom' orientation, the enone carbonyl is stabilized by a hydrogen bond to Gly164. Additionally, the cyclic head group appears to be sterically accommodated very nicely into the S2 pocket of cathepsin L. Although the predicted hydrogen bond network was

found to be less extensive in this orientation, the pose remained a viable option and was not discarded.

Due to the importance of identifying the correct pose for structure-guided lead optimization, molecular dynamic simulations were employed to elucidate the most probable ligand orientation. We hypothesized that a noncovalent complex must persist for a chemical reaction to occur between the enzyme and inhibitor. Specifically, a complex must exist where the ligand position does not significantly fluctuate and the catalytic thiol on Cys25 is continually well positioned to react with the gallinamide A enone. The use of multiple simulations enabled a more robust assessment of stability than using a single trajectory, which might become trapped in a local minimum. As binding between gallinamide A and cysteine cathepsins does not occur 100% of the time, it is reasonable to expect that not every simulation would result in a stable binding pose. Consequently, it became our goal to assess whether one pose consistently provided a more stable, reaction-ready configuration. The first 10 ns of each trajectory were discarded for equilibration, which left 120 ns of usable simulation time for pose analysis. All trajectories were aligned by Cα RMSD to their average structure, and these aligned structures were used in the analysis. Stability of complexes were assessed using three metrics: RMSD of gallinamide A in the binding pocket (Figure 4.6C), distance between the Cys25-SH and the reactive Michael addition carbon, hence forward referred to as C* (Figure 4.6D), and hydrogen bond persistence (Figure 4.6E). In all three metrics, the orientation where the head group

was oriented near the S1' pocket appeared to be better suited for covalent reaction to occur between the enzyme and inhibitor.
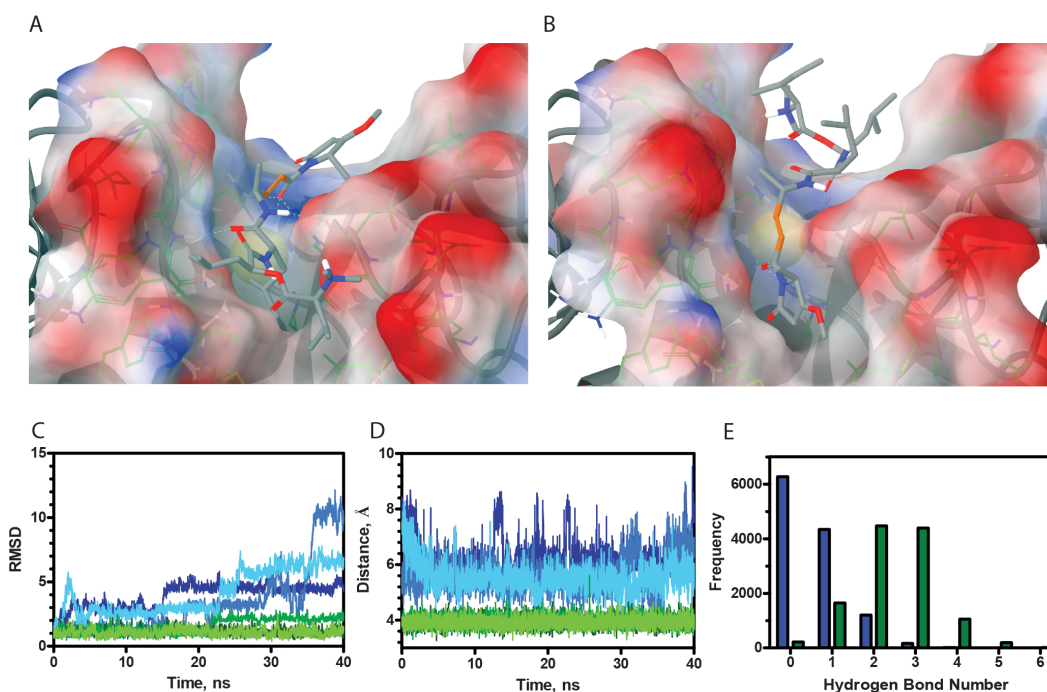
Figure 4.6. (A-B) Induced fit docked poses of representative structures for two gallinamide A (**1**) conformations docked into cathepsin L resulted in 'top' and 'bottom' poses, respectively. In C-E, the scores for the 'top' pose are represented by green and the 'bottom' pose blue. (C) RMSD values were obtained for each pose, corresponding to differences between the structure at a given time and the original pose, and thus are inversely correlated with stability. (D) The distance between Cys25-SH and C* for gallinamide A bound to cathepsin L. (E) A histogram of hydrogen bond count for both orientations of gallinamide A docked into the cathepsin L active site.

To first assess stability of the complexes, the movement of gallinamide A within the binding pocket was analyzed. Theoretically, ligands that have more stable interactions with their corresponding receptors move less within a pocket. The RMSD, a metric computing similarity of two structures, of gallinamide A complexed to cathepsin L was used to compare stability of each pose (Figure 4.6C). All three simulations of gallinamide A in the 'top' orientation proved to be more stable than in the 'bottom' orientation based on lower RMSD values.

The most important reaction criterion is that the catalytic Cys25 must be positioned adjacent to a reactive center on gallinamide A. Gallinamide A has three potentially reactive centers: two enones and an ester. In all of the computationally docked poses, only the reactive carbon on the enone was positioned adjacent to Cys25-SH, likely due to the geometric requirements of the cathepsin binding pocket. To assess stability of the bound complexes, we first looked at the percentage of time Cys25-SH was within 4 Å of the enone carbon (Figure 4.6D). While this distance cutoff is not enough to ensure reactivity, it is close enough that slight changes in the conformation of the bound complex could lead to a productive reaction. Distances were measured for each frame in the trajectories (12 000 per complex). For the 'bottom' pose, the enone was positioned within 4 Å of Cys25-SH only 0.28% of the time, as opposed to 65.80% for the 'top' pose. It should be noted that these results were not the result of the initial bias toward the first frame of the analysis, as the latter half of all trajectories were not statistically different from the first half.

Additionally, hydrogen bond interactions contribute significantly to the binding free energy of a complex, and can heavily dictate selectivity. In this regard, the two distinct poses of a protein-ligand complex could result in similar contributions from other contributing forces (e.g. van der Waals), yet a protein should have a more persistent hydrogen bond network with the correctly oriented ligand than one incorrectly oriented. For all trajectories, hydrogen bonds between cathepsin L and gallinamide A were counted and a histogram was generated of hydrogen bond count (Figure 4.6E). Hydrogen bonds were more persistent in the 'top' orientation compared

to the 'bottom' orientation (p <<< 0.001, T-test), further supporting the 'top' orientation.

Favorable attributes of a correct docking pose are also consistent with the 'top' orientation. Most importantly, the utilization of the oxyanion hole to stabilize the reaction is consistent with what is observed with the natural proneuropeptide substrates. In our simulations, the enone carbonyl of gallinamide A was observed to hydrogen bond with the oxyanion hole 55.4% of the time in the 'top' orientation, as compared to 17.06% in the 'bottom' orientation. The positively charged His163 is an especially good stabilizing residue for the transition state compared to Gly24 in the 'bottom' orientation, as individual electrostatic interactions are stronger than individual hydrogen bonds. Additionally, the positively charged tertiary amine in gallinamide A is positioned relatively close to the S2 pocket, which accommodates a basic residue in many of its natural proneuropeptide substrates. This positive charge may also help confer selectivity toward cathepsin L over cathepsin V, as the S2 pocket in cathepsin V is not as negatively charged.

Other atomistic interactions between gallinamide A and cathepsin L lend support to the head group 'top' pose being the correct orientation. Stolze et al.[21] conducted a limited SAR of gallinamide A inhibition of a *Plasmodium falciparum* cysteine protease and found that the cyclic head group was essential for gallinamide A-based inhibition. In the 'top' orientation, the carbonyl in the cyclic head group was found to hydrogen bond to either Gln19 or His189 for 83.55% of the simulation. The high persistence of this hydrogen bond suggests that removal would significantly

reduce the inhibitory potential of gallinamide A with cathepsin L. In contrast, hydrogen bonds with the cyclic head group were absent in the 'bottom' orientation, which would not support the aforementioned SAR. In other structural studies of cathepsin L inhibitors, Gly68 has been shown to hydrogen bond with inhibitors (144). In our simulations, a hydrogen bond between Gly68 and the carbonyl of the leucine residue appears 82.75% of the time in the 'top' pose, consistent with these findings.

The identification of the correct pose and proposed reaction mechanism between gallinamide A (**1**) and cathepsin L will serve as a platform on which to base the development of therapeutically relevant agents. The potential applications of gallinamide A as an antimalarial agent have been investigated since its discovery, though the potency and selectivity of this compound make it a valuable structural class in other fields. Inhibitors with selectivity between various human cysteine cathepsin proteases are lacking, and compounds with such a property may prove valuable in the elucidation of the relative roles of these enzymes in proneuropeptide processing (145). Furthermore, inhibitors of cathepsin L are under investigation for the reduction of tumor metastasis.[14] In this capacity, compounds do not need to enter cells to work as effective agents, as cathepsin L promotes metastasis through its degradation of matrix proteins in the extracellular space (132). Due to the easily modifiable carbon skeleton of **1**, this structure could serve as a starting place for the development of compounds with such favorable pharmaceutical properties. Lastly, this study identifies potential off-target binding sites for this marine natural product, an important consideration in

its development as a potential antimalarial agent, and thus will help guide the development of more selective and potent therapeutic agents.

Materials and Methods

General Experimental Procedures

Ultraviolet/visual-light spectra were measured on a Beckman Coulter DU 800 Spectrophotometer. NMR spectra were collected on a Varian Unity 500 MHz (500 MHz and 125 MHz for ^{1}H and ^{13}C NMR respectively) using CDCl$_3$ from Cambridge Isotope Laboratories, Inc. 99.8% D containing 0.03% v/v trimethylsilane (δ_H 0.0 and δ_C 77.16 as internal standards using trimethylsilane and CDCl$_3$, respectively). LCMS data were obtained with a Phenomenex Kinetex 5 μm C18(2) 100Å column (4.6 x 250 mm) with a Thermo Finnigan Surveyor Autosampler-Plus/LC-Pump-Plus/PDA-Plus system and a Thermo Finnigan LCQ Advantage Max mass spectrometer. HPLC purification was carried out with a Waters 515 HPLC Pump with a Waters 996 Photoiode Array Detector using Empower Pro software. All solvents were HPLC grade except for 99.8% acetone from Fisher which was distilled before use, and H$_2$O which was purified by a Millipore Milli-Q system before use.

Collection, Extraction and Isolation

The collection, extraction and VLC fractionation information can be found in the original isolation report (136). The fraction eluting with 4:6 hexanes/EtOAc (12.3 mg) showed the strongest inhibition (3% activity remaining) of cathepsin L at 3 μg/ml. LC-MS and ^{1}H- NMR spectra were obtained, and then the sample was further

fractionated using 200 mg Si NP solid phase extraction (SPE), producing 8 subfractions. These eluted with 3 mL washes of: 9:1 hexanes/EtOAc; 8:2 hexanes/EtOAc; 6:4 hexanes/EtOAc; 4:6 hexanes/EtOAc; 2:8 hexanes/EtOAc; 100% EtOAc; 1:1 EtOAc/MeOH; 100% MeOH. LC-MS and ^{1}H NMR traces were again obtained for each sub-fraction. Sub-fractions 6 through 8 were combined based on similarities in their LC-MS and NMR data, and all sub-fractions were subjected to the cathepsin L inhibition assay. The fourth and fifth fractions, eluting with 4:6 hexanes/EtOAc (1.64 mg) and 2:8 hexanes/EtOAc (0.52 mg), respectively, demonstrated the greatest inhibition of cathepsin L (99% and 99%, respectively), and were thus selected for additional purification. These two fractions were separately subjected to C18 RP-HPLC (Phenomenex Luna C18 4.6 x 250 mm RP-HPLC column, 5 μm ACN/H_2O 40:60, 1 mL/min) under both neutral and acidic (0.01% TFA) conditions. Separation was more robust under acidic conditions, and collections were tested in the cathepsin L inhibition assay to ensure protonation did not affect bioactivity. Collections from both neutral and acidic conditions demonstrated inhibition of cathepsin L (data not shown). Analytical RP-HPLC was used to profile each subfraction, and preparatory HPLC was performed to isolate and collect the major components (Phenomenex Luna C18 4.6 x 250 mm, 5 μm, 35% ACN/65% H2O/0.01%TFA, 1 mL/min).

Gallinamide A (1): colorless amorphous solid; UV, ^{1}H NMR, and high resolution ESITOFMS matched reported values within experimental error and are available in the Supporting Information.

Gallinamide B (2): colorless amorphous solid; $[\alpha]^{25}_D = 54.5$ (*c* 0.15, MeOH); UV (MeOH) λ_{max} (log ε) 201 (3.97), 248 (3.36) nm; HR-ESITOFMS *m/z* [M+H]$^+$ 579.3756 (calcd for $C_{30}H_{50}N_4O_7$, 579.3758).

Cathepsin L assay

Z-Phe-Arg-AMC substrate and E-64c were purchased from Bachem Americas (Torrance, CA). Human recombinant cathepsin L was purchased from R&D Systems (Minneapolis, MN). Assays were carried out using 20 µM Z-Phe-Arg-AMC and 3.0 ng/mL human recombinant cathepsin L. Assay buffer consisted of 50 mM sodium acetate, 100 mM NaCl, 1.0 mM EDTA and 4 mM dithiotreitol, pH 5.5. The enzymatic reaction (25˚C) was monitored on a SpecraMax Gemini or SpectraMax microplate reader (PerkinElmer Life Sciences, Waltham, MA) and the fluorescent signal was measured at the excitation and emission wavelengths of 365 and 450 nm, respectively.

Inhibitor potency determination

An 8-point, 3-fold serial dilution dose-response assay was performed in triplicate. Each well of a 96-well assay plate contained 50 μL of 2X concentrated assay buffer, 20 μL of water, and 10 μL of inhibitor in 10% DMSO. Positive and negative controls of 10 μL of 10 μM E64c or 10% DMSO, respectively, replaced inhibitor as internal controls. A 10 μL aliquot of a 30 ng/mL cathepsin L solution and 10 μL of 200 μM Z-Phe-Arg-AMC in assay buffer were added sequentially to initiate the protease reaction. Assay buffer (10 μL) was added in place of cathepsin L as a substrate blank for baseline correction. The resultant dose response concentration range was 333 nM to 0.46 nM inhibitor in a 100 μL final reaction volume. Data were

scaled to internal controls and a four-parameter logistic model (GraphPad vs. 5.0, Prism) was used to fit the measured data and determine IC_{50} values.

Selectivity

Gallinamide was tested for inhibition of human cathepsin V (100 ng/ml; R&D Systems) and human cathepsin B (100 ng/mL; R&D Systems). Reactions were performed in 50 mM sodium acetate buffer containing 100 mM NaCl, 4 mM DTT, and 1 mM EDTA, pH 5.5, using 10 μM Z-Phe-Arg-AMC substrate. IC_{50} scores were determined for both 0 min and 30 min preincubations, and selectivity index values were calculated relative to cathepsin L. All values were obtained in technical triplicate, as compound supply was insufficient to perform replicate plates.

Reversibility

A preincubation-dilution experiment was adapted from Copeland *et al.* (138) Cathepsin L at 100-fold its final assay concentration was incubated with gallinamide A at 10-fold its IC_{50} value for 30 min in a volume of 2 μL in a 96-well plate. This mixture was diluted 100-fold with assay buffer containing 10 μM Z-FR-AMC substrate to a final volume of 200 μL, resulting in a standard concentration of enzyme and 0.1-times the IC_{50} value of gallinamide A. A rapidly reversible inhibitor will dissociate from the enzyme to restore approximately 90% of enzymatic activity following the dilution event, while an irreversible inhibitor will maintain approximately 10% of enzymatic activity. Fluorescence intensities of the 200-μL wells were monitored continuously for AMC hydrolysis on a Spectramax plate reader in kinetic mode for 2 hr.

Active-site directed probe competition binding assay

Gallinamide A was prepared in a 6-point, 3-fold serial dilution in 10% DMSO. Assay buffer consisted of 100 mM sodium acetate, 1.0 mM EDTA, and 4 mM DTT, pH 5.5. One μL of 40 ng/μL cathepsin L solution was added to 8.5 μL of assay buffer and preincubated with 1.5 μL of inhibitor for 30 min at room temperature. Positive and negative controls of 1.5 μL of 100 μM CLIK-148 or 10% DMSO were included. After preincubation, 4 μL of 10 μM Biotin-Lys-C5 alkyl linker-Tyr-Leu-epoxide (DCG-04) was added and incubated for 30 min at room temperature. The incubated sample was mixed with 4X NuPAGE sample buffer and 50 mM DTT and denatured at 70 ˚C for 5 min. SDS-polyacrylamide gel separation was performed with Xcell SureLock system (Invitrogen) on 12% Bis-Tris gel (Invitrogen) at 200 V for 50 min. After transfer to a Hybond Nitrocell membrane (Amersham) at 30 V (1 hr), the membrane was blocked in 5% milk in TBS + 0.05% Tween for 1 h and then incubated with avidin and biotinylated horse-radish peroxidase, followed by washing (4X) and detection with ECL+ chemiluminescent substrate (Amersham).

Kinetic Analysis

Kinetic characterization of the interaction between gallinamide A and cathepsin L was performed to obtain an accurate second order inhibition constant corresponding to k_{obs}/[I]. Continuous monitoring of substrate hydrolysis in the presence of the inhibitor was used to measure the decrease in enzyme activity over time. Gallinamide A was tested in concentrations ranging from 12.3-1000 nM with

simultaneous mixing of enzyme, substrate and inhibitor. Progression curves of product

formation were fitted to a simple model of irreversible inhibition:

$$[P] = \frac{v_i}{k_{obs}}(1 - e^{-k_{obs}*t}) \qquad\qquad \text{Equation 1}$$

in which [P] is concentration of product, v_i is the initial reaction rate, k_{obs} is the

observed first order rate constant, and t is time in seconds. The observed first order

inhibition constant K_{obs} was plotted against [I] to obtain a linear relationship, the slope

of which represents the second order rate constant k_{inact}/K_I, a measure of affinity for

the slow binding inhibitor. Kinetic analysis was performed in technical quadruplicate.

Receptor and ligand preparation

A cathepsin L structure (PDB ID: 2XU3) (144) was downloaded from the

Protein Data Bank (65). All ligands and water molecules were removed and hydrogens

were subsequently added to the protein structure using PROPKA (67, 68, 146, 147)

and PDB2PQR (66) at pH 5.5. The proteins were then processed using Schrödinger

Maestro's Protein Preparation Wizard (www.schrodinger.com). Grids were generated

using the catalytic cysteine as the center. The inner box, which imposes restrictions on

the location of the center of the ligand, was set to a 10 Å cube and the outer box,

which restricts the possible locations of all ligand atoms, was set to a 30 Å cube.

Gallinamide A was built using Schrödinger Maestro's 2D builder and processed using

LigPrep at pH 5.5 +/- 2.0 to mimic assay conditions.

Docking protocol

Gallinamide A was positioned into cathepsin L active site using the Induced Fit Docking module of Schrödinger's glide (148). The center of the catalytic residues Cys25 and His163 was used as the box center. No hydrogen bond constraints were applied. Side chains within 5.0 Å of docked ligands were refined with Prime. XP Precision was used to score poses in the final re-docking step.

Molecular dynamics simulations

Selected complexes were extracted from the docking runs. In all poses, Na^+ counter-ions were added to neutralize the solution, and the complex was solvated in a TIP4PEW water box extending 10 Å in all x,y,z directions beyond the complex (149). Minimization and equilibration were done similarly to previous MD studies (150). In brief, the system was first minimized to remove any steric issues. Harmonic restraints of 2.0 kcal/mol/$Å^2$ were applied to the protein as the system was heated to 300K. A 1 ns simulation to equilibrate the system was then run with restraints removed. The SHAKE algorithm was used to constrain bonds involving hydrogens (151), and Particle Mesh Ewald was used to treat long-range electrostatics (80). Simulations of 50 ns duration were run in triplicate from the minimized structure. For each simulation, the first 10 ns were discarded for equilibration of the docked pose, resulting in 120 ns of simulation per system that was analyzed.

Simulation snapshots were taken every 10 ps. All analysis was performed using either the *ptraj* module of AMBER12 (152) or analysis tools in VMD (69). Trajectories were aligned by Ca RMSD to their average structure. Hydrogen bonds

were determined for polar hydrogen acceptor atoms (O, N, S) with 3.0 Å and 120°
cutoffs.

Acknowledgements

Chapter 4 is a minimally modified reprint of the material appearing in Miller
B, Friedman AJ, Choi H, Hogan J, McCammon JA, Hook V, and Gerwick WH. The
Marine Cyanobacterial Metabolite Gallinamide A is a Potent and Selective Inhibitor of
Human Cathepsin L. Submitted. I was the second author, and principal computational
author, on this paper.

Supporting Information
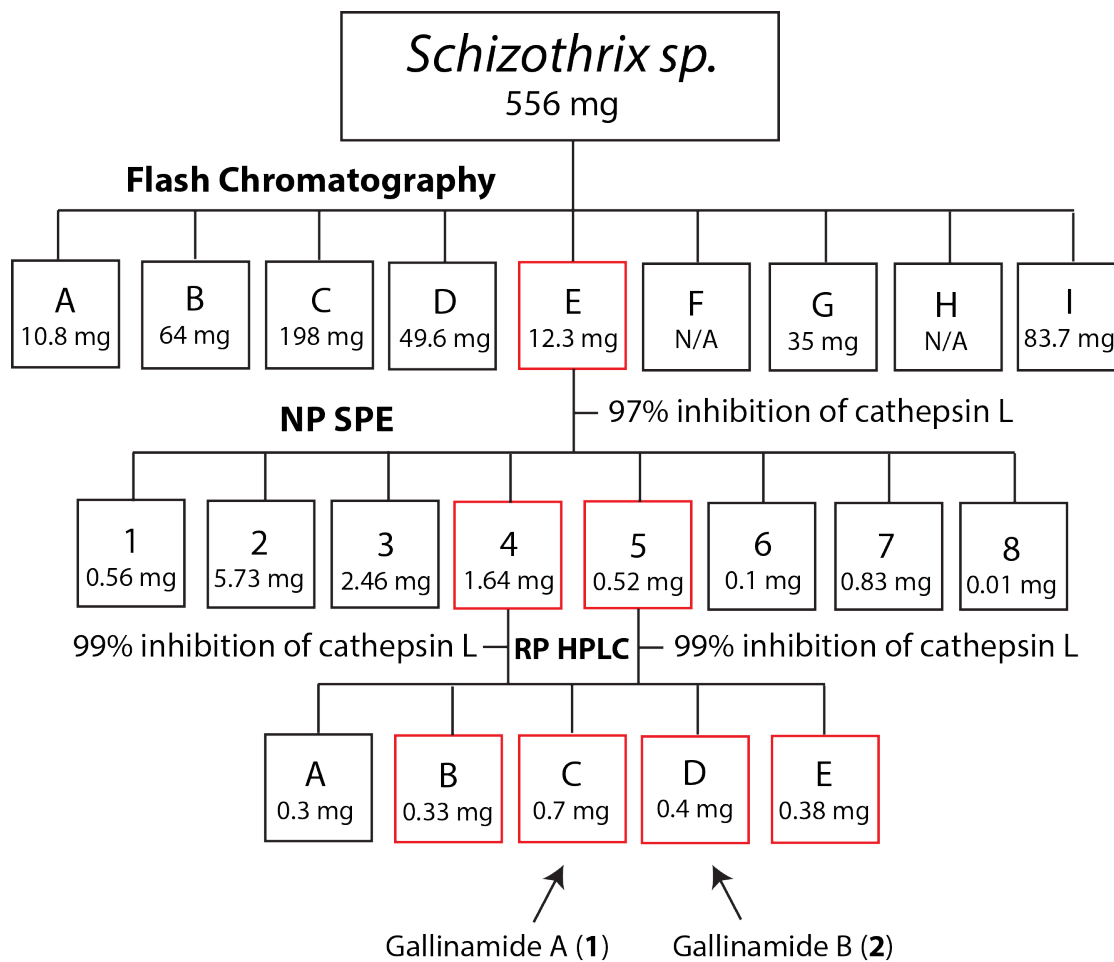
Scheme S4.1. Fractionation and isolation scheme
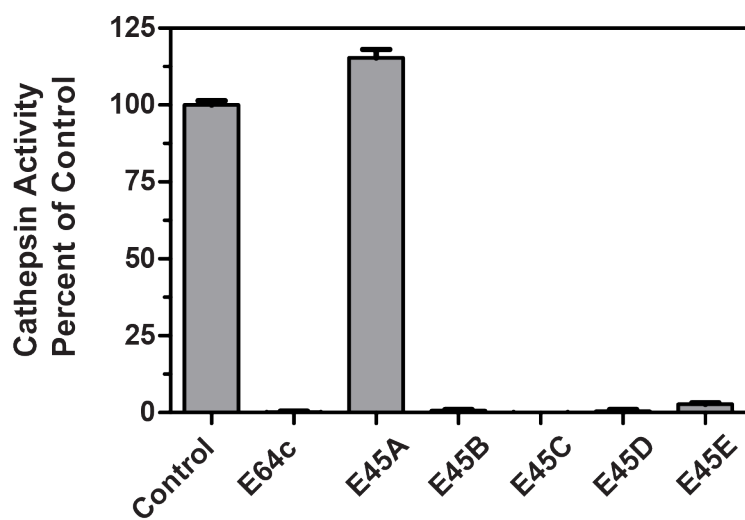
Figure S4.1. Cathepsin L activity assay for HPLC collections
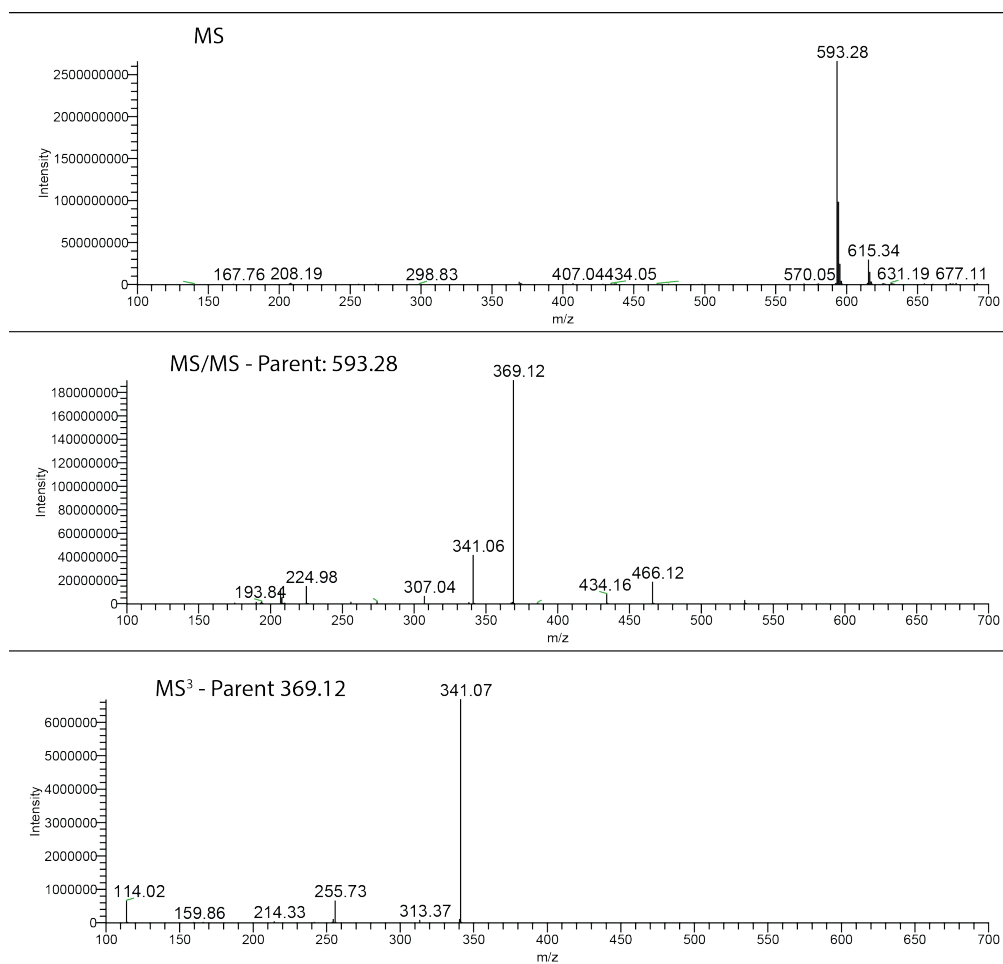
Figure S4.2. ESI-MS/MS and MS^3 spectra of Gallinamide A **(1)**
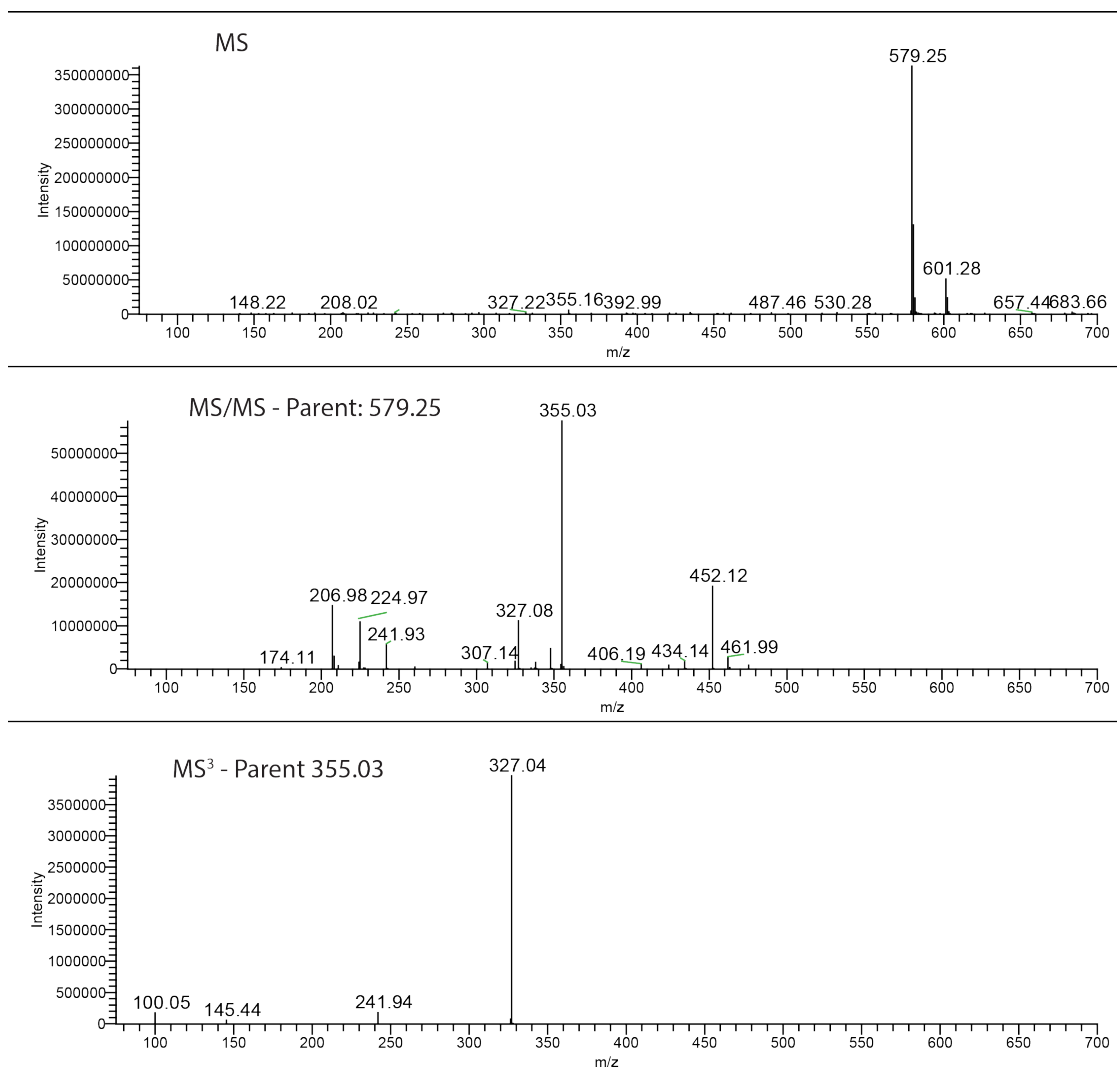
Figure S4.3. ESI-MS/MS and MS^3 spectra of Gallinamide B (**2**)
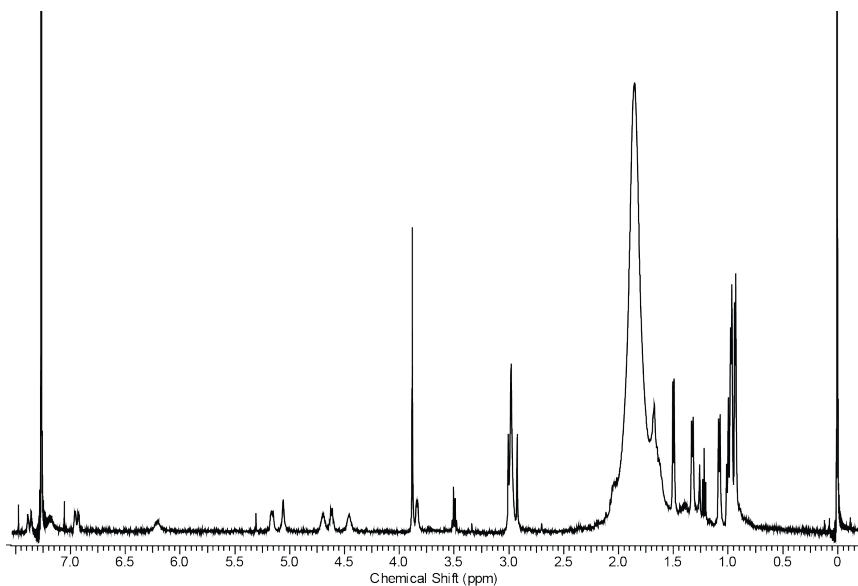
Figure S4.4. ^{1}H NMR spectrum of gallinamide A **(1)** in CDCl$_3$ (500 MHz)
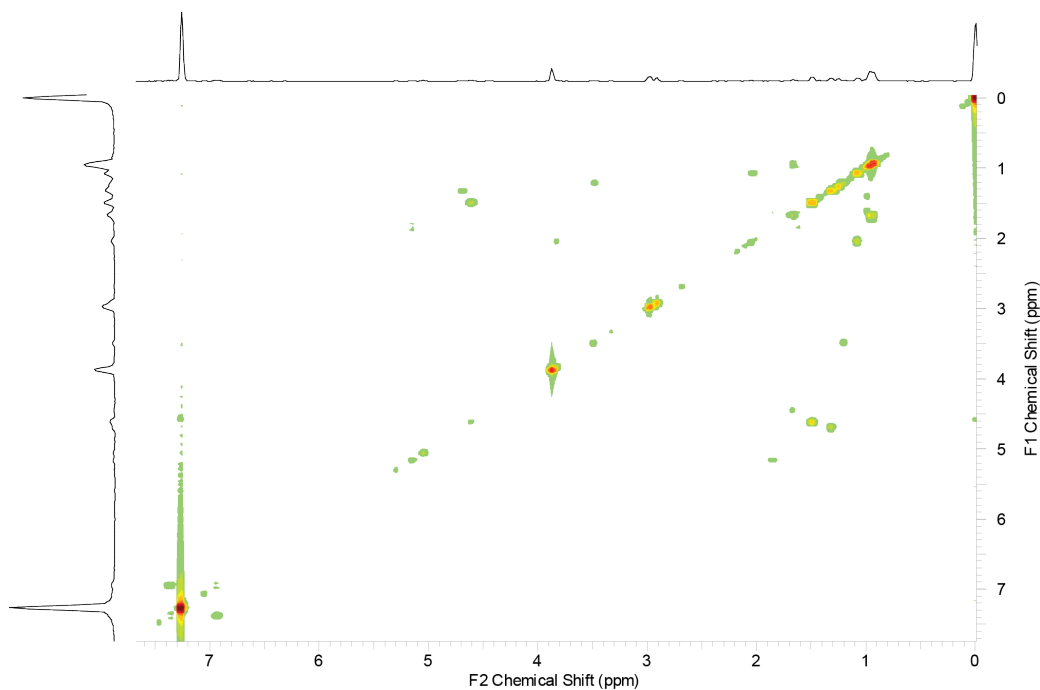
Figure S4.5. ^{1}H-^{1}H COSY spectrum of gallinamide A (**1**) in CDCl$_3$ (500 MHz).^{1}H NMR (CDCl$_3$, 500 MHz) δ 7.42 (1H, d, *J* = 15.8 Hz, H-8), 7.14 (1H, brs, 13-NH), 6.94 (1H, dd, *J* = 15.8, 4.3 Hz, H-9), 6.21 (1H, brd, 10-NH), 5.16 (1H, dd, *J* = 9.3, 4.0 Hz, H-19), 5.05 (1H, s, H-2), 4.69 (1H, m, H-10), 4.61 (1H, q, *J* = 6.5 Hz, H-4), 4.45 (1H, brdd, H-13), 3.88 (3H, s, O-Me), 3.83 (1H, d, *J* = 5.9 Hz, H-25), 2.98 (6H, brs, H-30), 2.04 (1H, m, H-26), 1.84 (2H, m, H-20a), 1.67 (2H, m, H-14), 1.66 (1H, m, H-15), 1.63 (2H, m, H-20b), 1.62 (1H, m, H-21), 1.5 (3H, d, *J* = 6.5 Hz, H-5), 1.4 (2H, brm, H-27), 1.32 (3H, d, *J* = 6.5 Hz, H-11), 1.08 (3H, d, *J* = 6.5 Hz, H-29), 0.99 (3H, t, *J* = 7.3 Hz, H-28), 0.96 (3H, d, H-22), 0.96 (3H, d, *J* = 6.2 Hz, H-16), 0.93 (3H, d, *J* = 6.2 Hz, H-17), 0.92 (3H, d, *J* = 6.2 Hz, H-23)

Chapter 5: Conclusions

Two fundamental challenges in modern day drug discovery are reducing off-target effects and overcoming resistance. Both of these problems are rooted in evolution and pose significant challenges to the design of effective drugs. Many methods have been employed to overcome this problem, to limited success. Recent advances in computational power have increased the potential for such analyses. Combinations of system-specific analysis and algorithmic development have the potential to improve such efforts. Two fields that will benefit the most from merging evolution and drug discovery are infectious disease and oncology. In both, the development of resistance poses a significant problem. Consequently, either new targets need to be identified, or compounds that leverage and/or work around these genetic perturbations need to be designed.

In this work, we described several efforts to understand drug discovery under the context of evolution. Specifically, disease targets for both infectious disease and cancer were analyzed using both developed and existing methods. In studying the dynamics of *Tb*GalE, we found that targeting evolutionarily conserved residues, as well as those that make important ligand interactions with their backbone will increase the likelihood of evading resistance mechanisms. Selecting for compounds that preferentially bind a recently evolved negatively charged pocket, such as in cathepsin L, can help increase specificity over highly related proteins, such as cathepsin V. To analyze the protein-ligand coevolution as it pertains to drug discovery, we developed a novel algorithm, CrystalDock, which eschews traditional binding free-

energy calculations, instead leveraging existing structural information to "dock" molecular fragments. We hypothesized that much of a ligand's overall binding affinity can be attributed to the localized enthalpic interactions between functional moieties on a ligand and the receptor microenvironments they interact with. By mining a structural database of known ligand-fragment/receptor-microenvironment pairs for matches, we demonstrated that ligand modifications suggested by our approach can improve the potency of known inhibitors. While not yet demonstrated, it is our anticipation that linking together these fragments, which individually are weak binders, can also form high affinity compounds.

Further integration of evolutionary information, particularly resistance information, with drug discovery will increase the efficacy of prescribed treatments. Currently, the development of resistance to a treatment is often discovered after the fact, a suboptimal process. Efficiently translating from genomic variation to structural perturbation to drug response represents a significant unmet need in the treatment of many diseases.

As the cost of genomic sequencing continues to drop, the long-awaited goal of personalized medicine is rapidly becoming a reality. Ideally, one would sequence diseased cells or organisms, identify variations driving both the disease and treatment, and subsequently administer drugs best suited to combat that genomic profile. The development of automated tools to account for such perturbations is still in its infancy, and has the potential to transform the way in which treatments are administered. In certain diseases such as cancer, leveraging genomic profiles to identify driving and resistant mutations to inform drug discovery will likely increase the efficacy of

treatment. Specifically, the creation of a database that pairs genetic profiles and drug-response data will allow for real-time analysis of patients' tumors leading to the recommendation of better therapies.

Often, resistant mutations disrupt binding by creating steric clashes in ligands, eliminating a key interaction, or creating a disruptive interaction. Currently, it is infeasible to computationally or experimentally screen compounds against all models of resistant targets. Alternatively, one may screen high-affinity compounds against wild type structures and subsequently map resistant mutations onto the structure. Newly generated protein-ligand pairs would be evaluated by several scoring functions (153-155) to assess the effect of the mutations on binding. It is important to note that such approaches would work best with chemical series, as initial docked structures could be validated for convergence. After the generation of the initial database, new compounds could be docked using ligand-based approaches to accurately position themselves in binding pockets already containing ligands of relative similarity.

Eventually, the development of automated pipelines will allow for the rapid identification of genomic variants that cause disease. In order to maximize the therapeutic potential of such pipelines, concurrent approaches need to be developed to combat resistance and minimize drug off target effects. Understanding how evolution impacts these challenges will become increasingly important future drug discovery efforts.

References

1.	Chen S, Krinsky BH, Long M. New genes as drivers of phenotypic evolution. Nature reviews Genetics. 2013;14(9):645-60. doi: 10.1038/nrg3521. PubMed PMID: 23949544.

2.	Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A. 1992;89(22):10915-9. PubMed PMID: 1438297; PubMed Central PMCID: PMC50453.

3.	Styczynski MP, Jensen KL, Rigoutsos I, Stephanopoulos G. BLOSUM62 miscalculations improve search performance. Nat Biotechnol. 2008;26(3):274-5. doi: 10.1038/nbt0308-274. PubMed PMID: 18327232.

4.	Massova I, Kollman PA. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. Journal of the American Chemical Society. 1999;121(36):8133-43. doi: 10.1021/ja990935j.

5.	Pazos F, Ranea JA, Juan D, Sternberg MJ. Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome. J Mol Biol. 2005;352(4):1002-15. doi: 10.1016/j.jmb.2005.07.005. PubMed PMID: 16139301.

6.	Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE. Co-evolution of proteins with their interaction partners. J Mol Biol. 2000;299(2):283-93. doi: 10.1006/jmbi.2000.3732. PubMed PMID: 10860738.

7.	Ramani AK, Marcotte EM. Exploiting the co-evolution of interacting proteins to discover interaction specificity. J Mol Biol. 2003;327(1):273-84. PubMed PMID: 12614624.

8.	Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. J Comput Aided Mol Des. 2008;22(9):693-705. doi: 10.1007/s10822-007-9159-2. PubMed PMID: 18196463; PubMed Central PMCID: PMC2516539.

9.   Jorgensen WL, Tirado-Rives J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. J Comput Chem. 2005;26(16):1689-700. doi: 10.1002/jcc.20297. PubMed PMID: 16200637.

10.   Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE, 3rd. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Accounts of chemical research. 2000;33(12):889-97. PubMed PMID: 11123888.

11.   Shoichet BK. Virtual screening of chemical libraries. Nature. 2004;432(7019):862-5. doi: 10.1038/nature03197. PubMed PMID: 15602552; PubMed Central PMCID: PMC1360234.

12.   Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. 2004;3(11):935-49. doi: 10.1038/nrd1549. PubMed PMID: 15520816.

13.   Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol. 2008;18(2):178-84. doi: 10.1016/j.sbi.2008.01.004. PubMed PMID: 18302984; PubMed Central PMCID: PMC2396190.

14.   C BR, Subramanian J, Sharma SD. Managing protein flexibility in docking and its applications. Drug Discov Today. 2009;14(7-8):394-400. doi: 10.1016/j.drudis.2009.01.003. PubMed PMID: 19185058.

15.   Sotriffer CA, Sanschagrin P, Matter H, Klebe G. SFCscore: scoring functions for affinity prediction of protein-ligand complexes. Proteins. 2008;73(2):395-419. doi: 10.1002/prot.22058. PubMed PMID: 18442132.

16.   Harris R, Olson AJ, Goodsell DS. Automated prediction of ligand-binding sites in proteins. Proteins. 2008;70(4):1506-17. doi: 10.1002/prot.21645. PubMed PMID: 17910060.

17.   Xie L, Bourne PE. A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. BMC Bioinformatics. 2007;8 Suppl 4:S9. doi: 10.1186/1471-2105-8-S4-S9. PubMed PMID: 17570152; PubMed Central PMCID: PMC1892088.

18.   Xie L, Bourne PE. Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. Proc Natl Acad Sci U S A. 2008;105(14):5441-6. doi: 10.1073/pnas.0704422105. PubMed PMID: 18385384; PubMed Central PMCID: PMC2291117.

19.   Xie L, Xie L, Bourne PE. A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery. Bioinformatics. 2009;25(12):i305-12. doi: 10.1093/bioinformatics/btp220. PubMed PMID: 19478004; PubMed Central PMCID: PMC2687974.

20.   Liu T, Altman RB. Using multiple microenvironments to find similar ligand-binding sites: application to kinase inhibitor binding. PLoS Comput Biol. 2011;7(12):e1002326. doi: 10.1371/journal.pcbi.1002326. PubMed PMID: 22219723; PubMed Central PMCID: PMC3248393.

21.   Brenke R, Kozakov D, Chuang GY, Beglov D, Hall D, Landon MR, Mattos C, Vajda S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. Bioinformatics. 2009;25(5):621-7. Epub 2009/01/30. doi: btp036 [pii] 10.1093/bioinformatics/btp036. PubMed PMID: 19176554; PubMed Central PMCID: PMC2647826.

22.   Durrant JD, Amaro RE, Xie L, Urbaniak MD, Ferguson MA, Haapalainen A, Chen Z, Di Guilmi AM, Wunder F, Bourne PE, McCammon JA. A multidimensional strategy to detect polypharmacological targets in the absence of structural and sequence homology. PLoS Comput Biol. 2010;6(1):e1000648. doi: 10.1371/journal.pcbi.1000648. PubMed PMID: 20098496; PubMed Central PMCID: PMC2799658.

23.   Kinnings SL, Xie L, Fung KH, Jackson RM, Xie L, Bourne PE. The Mycobacterium tuberculosis drugome and its polypharmacological implications. PLoS Comput Biol. 2010;6(11):e1000976. doi: 10.1371/journal.pcbi.1000976. PubMed PMID: 21079673; PubMed Central PMCID: PMC2973814.

24.   Xie L, Xie L, Kinnings SL, Bourne PE. Novel computational approaches to polypharmacology as a means to define responses to individual drugs. Annu Rev Pharmacol Toxicol. 2012;52:361-79. doi: 10.1146/annurev-pharmtox-010611-134630. PubMed PMID: 22017683.

25.   Cutler DM. The demise of the blockbuster? N Engl J Med. 2007;356(13):1292-3. doi: 10.1056/NEJMp078020. PubMed PMID: 17392299.

26.   Xie L, Li J, Xie L, Bourne PE. Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. PLoS Comput Biol. 2009;5(5):e1000387. doi: 10.1371/journal.pcbi.1000387. PubMed PMID: 19436720; PubMed Central PMCID: PMC2676506.

27.   Kindy MS, Yu J, Zhu H, El-Amouri SS, Hook V, Hook GR. Deletion of the Cathepsin B Gene Improves Memory Deficits in a Transgenic Alzheimer's Disease Mouse Model Expressing AbetaPP Containing the Wild-Type beta-Secretase Site Sequence. J Alzheimers Dis. 2012. Epub 2012/02/18. doi: 1J1654L311GT5NL2 [pii] 10.3233/JAD-2012-111604. PubMed PMID: 22337825.

28.   Lankelma JM, Voorend DM, Barwari T, Koetsveld J, Van der Spek AH, De Porto AP, Van Rooijen G, Van Noorden CJ. Cathepsin L, target in cancer treatment? Life Sci. 2010;86(7-8):225-33. Epub 2009/12/05. doi: 10.1016/j.lfs.2009.11.016. PubMed PMID: 19958782.

29.   Viken MK, Sollid HD, Joner G, Dahl-Jorgensen K, Ronningen KS, Undlien DE, Flato B, Selvaag AM, Forre O, Kvien TK, Thorsby E, Melms A, Tolosa E, Lie BA. Polymorphisms in the cathepsin L2 (CTSL2) gene show association with type 1 diabetes and early-onset myasthenia gravis. Hum Immunol. 2007;68(9):748-55. doi: 10.1016/j.humimm.2007.05.009. PubMed PMID: 17869649.

30.   Reiser J, Adair B, Reinheckel T. Specialized roles for cysteine cathepsins in health and disease. J Clin Invest. 2010;120(10):3421-31. doi: 10.1172/JCI42918. PubMed PMID: 20921628; PubMed Central PMCID: PMC2947230.

31.   Tsuge H, Nishimura T, Tada Y, Asao T, Turk D, Turk V, Katunuma N. Inhibition mechanism of cathepsin L-specific inhibitors based on the crystal structure of papain-CLIK148 complex. Biochem Biophys Res Commun. 1999;266(2):411-6. doi: 10.1006/bbrc.1999.1830. PubMed PMID: 10600517.

32.   Uinuk-Ool TS, Takezaki N, Kuroda N, Figueroa F, Sato A, Samonte IE, Mayer WE, Klein J. Phylogeny of antigen-processing enzymes: cathepsins of a cephalochordate, an agnathan and a bony fish. Scandinavian journal of immunology. 2003;58(4):436-48. PubMed PMID: 14507309.

33.   Delespaux V, de Koning HP. Drugs and drug resistance in African trypanosomiasis. Drug Resist Updat. 2007;10(1-2):30-50. Epub 2007/04/06. doi: S1368-7646(07)00021-0 [pii]
10.1016/j.drup.2007.02.004. PubMed PMID: 17409013.

34.   Legros D, Ollivier G, Gastellu-Etchegorry M, Paquet C, Burri C, Jannin J, Buscher P. Treatment of human African trypanosomiasis--present situation and needs for research and development. Lancet Infect Dis. 2002;2(7):437-40. Epub 2002/07/20. doi: S1473309902003213 [pii]. PubMed PMID: 12127356.

35.   Maser P, Luscher A, Kaminsky R. Drug transport and drug resistance in African trypanosomes. Drug Resist Updat. 2003;6(5):281-90. Epub 2003/12/04. doi: S1368764603000864 [pii]. PubMed PMID: 14643298.

36.   Vansterkenburg EL, Coppens I, Wilting J, Bos OJ, Fischer MJ, Janssen LH, Opperdoes FR. The uptake of the trypanocidal drug suramin in combination with low-density lipoproteins by Trypanosoma brucei and its possible mode of action. Acta Trop. 1993;54(3-4):237-50. Epub 1993/09/01. PubMed PMID: 7902661.

37.   Gehrig S, Efferth T. Development of drug resistance in Trypanosoma brucei rhodesiense and Trypanosoma brucei gambiense. Treatment of human African trypanosomiasis with natural products (Review). Int J Mol Med. 2008;22(4):411-9. Epub 2008/09/25. PubMed PMID: 18813846.

38.   Control and surveillance of African trypanosomiasis. Report of a WHO Expert Committee. World Health Organ Tech Rep Ser. 1998;881:I-VI, 1-114. Epub 1999/03/10. PubMed PMID: 10070249.

39.   Chappuis F, Udayraj N, Stietenroth K, Meussen A, Bovier PA. Eflornithine is safer than melarsoprol for the treatment of second-stage Trypanosoma brucei gambiense human African trypanosomiasis. Clin Infect Dis. 2005;41(5):748-51. Epub 2005/08/05. doi: CID36204 [pii]
10.1086/432576. PubMed PMID: 16080099.

40.   Iten M, Matovu E, Brun R, Kaminsky R. Innate lack of susceptibility of Ugandan Trypanosoma brucei rhodesiense to DL-alpha-difluoromethylornithine (DFMO). Trop Med Parasitol. 1995;46(3):190-4. Epub 1995/09/01. PubMed PMID: 8533023.

41.   Mehlert A, Zitzmann N, Richardson JM, Treumann A, Ferguson MA. The glycosylation of the variant surface glycoproteins and procyclic acidic repetitive proteins of Trypanosoma brucei. Mol Biochem Parasitol. 1998;91(1):145-52. Epub 1998/05/09. doi: S0166-6851(97)00187-4 [pii]. PubMed PMID: 9574932.

42.   Tetaud E, Barrett MP, Bringaud F, Baltz T. Kinetoplastid glucose transporters. Biochem J. 1997;325 (Pt 3):569-80. Epub 1997/08/01. PubMed PMID: 9271074; PubMed Central PMCID: PMC1218597.

43.   Shaw MP, Bond CS, Roper JR, Gourley DG, Ferguson MA, Hunter WN. High-resolution crystal structure of Trypanosoma brucei UDP-galactose 4'-epimerase: a potential target for structure-based development of novel trypanocides. Mol Biochem Parasitol. 2003;126(2):173-80. Epub 2003/03/05. doi: S0166685102002438 [pii]. PubMed PMID: 12615316.

44.   Durrant JD, Urbaniak MD, Ferguson MA, McCammon JA. Computer-aided identification of Trypanosoma brucei uridine diphosphate galactose 4'-epimerase inhibitors: toward the development of novel therapies for African sleeping sickness. J Med Chem. 2010;53(13):5025-32. Epub 2010/06/10. doi: 10.1021/jm100456a. PubMed PMID: 20527952; PubMed Central PMCID: PMC2895357.

45.   Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. Ber Dtsch Chem Ges. 1894;27:2895-993.

46.   Durrant JD, McCammon JA. Computer-aided drug-discovery techniques that account for receptor flexibility. Curr Opin Pharmacol. 2010;10(6):770-4. Epub 2010/10/05. doi: 10.1016/j.coph.2010.09.001. PubMed PMID: 20888294; PubMed Central PMCID: PMC3026641.

47.   Koshland DE, Jr., Ray WJ, Jr., Erwin MJ. Protein structure and enzyme action. Fed Proc. 1958;17(4):1145-50. Epub 1958/12/01. PubMed PMID: 13619786.

48.   Kumar S, Ma B, Tsai CJ, Wolfson H, Nussinov R. Folding funnels and conformational transitions via hinge-bending motions. Cell Biochem Biophys. 1999;31(2):141-64. Epub 1999/12/11. doi: 10.1007/BF02738169. PubMed PMID: 10593256.

49.   Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. Protein Eng. 1999;12(9):713-20. Epub 1999/10/03. PubMed PMID: 10506280.

50.   Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci. 2002;11(2):184-97. Epub 2002/01/16. doi: 10.1110/ps.21302. PubMed PMID: 11790828; PubMed Central PMCID: PMC2373439.

51.   Tsai CJ, Ma B, Nussinov R. Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci U S A. 1999;96(18):9970-2. Epub 1999/09/01. PubMed PMID: 10468538; PubMed Central PMCID: PMC33715.

52.   Mongan J. Interactive essential dynamics. J Comput Aided Mol Des. 2004;18(6):433-6. Epub 2005/01/25. PubMed PMID: 15663003.

53.   Alphey MS, Burton A, Urbaniak MD, Boons GJ, Ferguson MA, Hunter WN. Trypanosoma brucei UDP-galactose-4'-epimerase in ternary complex with NAD+ and the substrate analogue UDP-4-deoxy-4-fluoro-alpha-D-galactose. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2006;62(Pt 9):829-34. Epub 2006/09/02. doi: S1744309106028740 [pii]
10.1107/S1744309106028740. PubMed PMID: 16946458; PubMed Central PMCID: PMC2242870.

54.   Allard ST, Giraud MF, Naismith JH. Epimerases: structure, function and mechanism. Cell Mol Life Sci. 2001;58(11):1650-65. Epub 2001/11/15. PubMed PMID: 11706991.

55.   Liu Y, Thoden JB, Kim J, Berger E, Gulick AM, Ruzicka FJ, Holden HM, Frey PA. Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from Escherichia coli. Biochemistry. 1997;36(35):10675-84. Epub 1997/09/02. doi: 10.1021/bi970430a
bi970430a [pii]. PubMed PMID: 9271498.

56.   Barrett MP, Fairlamb AH. The biochemical basis of arsenical-diamidine crossresistance in African trypanosomes. Parasitol Today. 1999;15(4):136-40. Epub 1999/05/14. doi: S0169-4758(99)01414-3 [pii]. PubMed PMID: 10322334.

57.   Carter NS, Fairlamb AH. Arsenical-resistant trypanosomes lack an unusual adenosine transporter. Nature. 1993;361(6408):173-6. Epub 1993/01/14. doi: 10.1038/361173a0. PubMed PMID: 8421523.

58.   Matovu E, Geiser F, Schneider V, Maser P, Enyaru JC, Kaminsky R, Gallati S, Seebeck T. Genetic variants of the TbAT1 adenosine transporter from African trypanosomes in relapse infections following melarsoprol therapy. Mol Biochem Parasitol. 2001;117(1):73-81. Epub 2001/09/12. doi: S0166-6851(01)00332-2 [pii]. PubMed PMID: 11551633.

59.   Stewart ML, Krishna S, Burchmore RJ, Brun R, de Koning HP, Boykin DW, Tidwell RR, Hall JE, Barrett MP. Detection of arsenical drug resistance in Trypanosoma brucei with a simple fluorescence test. Lancet. 2005;366(9484):486-7. Epub 2005/08/09. doi: S0140-6736(05)66793-1 [pii] 10.1016/S0140-6736(05)66793-1. PubMed PMID: 16084257.

60.   Vincent IM, Creek D, Watson DG, Kamleh MA, Woods DJ, Wong PE, Burchmore RJ, Barrett MP. A molecular mechanism for eflornithine resistance in African trypanosomes. PLoS Pathog. 2010;6(11):e1001204. Epub 2010/12/03. doi: 10.1371/journal.ppat.1001204. PubMed PMID: 21124824; PubMed Central PMCID: PMC2991269.

61.   Langmead CJ. Screening for positive allosteric modulators: assessment of modulator concentration-response curves as a screening paradigm. J Biomol Screen. 2007;12(5):668-76. Epub 2007/05/05. doi: 1087057107301854 [pii] 10.1177/1087057107301854. PubMed PMID: 17478477.

62.   Surig U, Gaal K, Kostenis E, Trankle C, Mohr K, Holzgrabe U. Muscarinic allosteric modulators: atypical structure-activity-relationships in bispyridinium-type compounds. Arch Pharm (Weinheim). 2006;339(4):207-12. Epub 2006/03/31. doi: 10.1002/ardp.200600005. PubMed PMID: 16572483.

63.   Brahma A, Banerjee N, Bhattacharyya D. UDP-galactose 4-epimerase from Kluyveromyces fragilis--catalytic sites of the homodimeric enzyme are functional and regulated. FEBS J. 2009;276(22):6725-40. Epub 2009/10/22. doi: EJB7386 [pii] 10.1111/j.1742-4658.2009.07386.x. PubMed PMID: 19843183.

64.	Hay M, Bhaduri A. UDPGlucose 4-epimerase from Saccharomyces fragilis. Allosteric kinetics with UDP-glucose as substrate. J Biol Chem. 1975;250(11):4373-5. Epub 1975/06/10. PubMed PMID: 1126955.

65.	Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42. Epub 1999/12/11. PubMed PMID: 10592235; PubMed Central PMCID: PMC102472.

66.	Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 2004;32(Web Server issue):W665-7. Epub 2004/06/25. doi: 10.1093/nar/gkh381. PubMed PMID: 15215472; PubMed Central PMCID: PMC441519.

67.	Bas DC, Rogers DM, Jensen JH. Very fast prediction and rationalization of pKa values for protein-ligand complexes. Proteins. 2008;73(3):765-83. Epub 2008/05/24. doi: 10.1002/prot.22102. PubMed PMID: 18498103.

68.	Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. Proteins. 2005;61(4):704-21. Epub 2005/10/19. doi: 10.1002/prot.20660. PubMed PMID: 16231289.

69.	Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14(1):33-8, 27-8. Epub 1996/02/01. doi: 0263785596000185 [pii]. PubMed PMID: 8744570.

70.	Pavelites JJ, Gao J, Bash PA, Jr. ADM. A molecular mechanics force field for NAD+, NADH, and the pyrophosphate groups of nucleotides. Journal of Computational Chemistry. 1997;18(2):221-39.

71.	Walker RC, de Souza MM, Mercer IP, Gould IR, Klug DR. Large and Fast Relaxations inside a Protein:‚Äâ Calculation and Measurement of Reorganization Energies in Alcohol Dehydrogenase. The Journal of Physical Chemistry B. 2002;106(44):11658-65. doi: 10.1021/jp0261814.

72.	Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem. 2004;25(9):1157-74. Epub 2004/04/30. doi: 10.1002/jcc.20035. PubMed PMID: 15116359.

73.   Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 2006;65(3):712-25. Epub 2006/09/19. doi: 10.1002/prot.21123. PubMed PMID: 16981200.

74.   Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics. 1983;79(2):926-35.

75.   Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem. 2005;26(16):1781-802. Epub 2005/10/14. doi: 10.1002/jcc.20289. PubMed PMID: 16222654; PubMed Central PMCID: PMC2486339.

76.   KalÈ L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. Journal of Computational Physics. 1999;151(1):283-312.

77.   Andersen HC. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. Journal of Computational Physics. 1983;52(1):24-34.

78.   Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. Journal of Computational Chemistry. 1992;13(8):952-62. doi: 10.1002/jcc.540130805.

79.   Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. The Journal of Chemical Physics. 1995;103(11):4613-21.

80.   Darden T, York D, Pedersen L. Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. The Journal of Chemical Physics. 1993;98(12):10089-92.

81.   Christen M, Hunenberger PH, Bakowies D, Baron R, Burgi R, Geerke DP, Heinz TN, Kastenholz MA, Krautler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF. The GROMOS software for biomolecular simulation: GROMOS05. J Comput Chem. 2005;26(16):1719-51. Epub 2005/10/08. doi: 10.1002/jcc.20303. PubMed PMID: 16211540.

82.   Durrant JD, McCammon JA. HBonanza: a computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis. J Mol Graph Model. 2011;31:5-9. Epub 2011/09/02. doi: 10.1016/j.jmgm.2011.07.008. PubMed PMID: 21880522; PubMed Central PMCID: PMC3218086.

83.   Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS. The Universal Protein Resource (UniProt). Nucleic Acids Res. 2005;33(Database issue):D154-9. Epub 2004/12/21. doi: 33/suppl_1/D154 [pii] 10.1093/nar/gki070. PubMed PMID: 15608167; PubMed Central PMCID: PMC540024.

84.   Roberts E, Eargle J, Wright D, Luthey-Schulten Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC Bioinformatics. 2006;7:382. Epub 2006/08/18. doi: 1471-2105-7-382 [pii] 10.1186/1471-2105-7-382. PubMed PMID: 16914055; PubMed Central PMCID: PMC1586216.

85.   Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-8. Epub 2007/09/12. doi: 10.1093/bioinformatics/btm404. PubMed PMID: 17846036.

86.   McCorvie TJ, Wasilenko J, Liu Y, Fridovich-Keil JL, Timson DJ. In vivo and in vitro function of human UDP-galactose 4'-epimerase variants. Biochimie. 2011;93(10):1747-54. Epub 2011/06/28. doi: 10.1016/j.biochi.2011.06.009. PubMed PMID: 21703329; PubMed Central PMCID: PMC3168732.

87.   Timson DJ. Functional analysis of disease-causing mutations in human UDP-galactose 4-epimerase. FEBS J. 2005;272(23):6170-7. Epub 2005/11/24. doi: 10.1111/j.1742-4658.2005.05017.x. PubMed PMID: 16302980.

88.   Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P, Salemme FR. High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen. 2001;6(6):429-40. Epub 2002/01/15. doi: 10.1089/108705701753364922. PubMed PMID: 11788061.

89.   Pey AL, Ying M, Cremades N, Velazquez-Campoy A, Scherer T, Thony B, Sancho J, Martinez A. Identification of pharmacological chaperones as potential therapeutic agents to treat phenylketonuria. J Clin Invest. 2008;118(8):2858-67. Epub 2008/07/04. doi: 10.1172/JCI34355. PubMed PMID: 18596920; PubMed Central PMCID: PMC2441854.

90.   Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal Biochem. 2006;357(2):289-98. Epub 2006/09/12. doi: 10.1016/j.ab.2006.07.027. PubMed PMID: 16962548.

91.   Matulis D, Kranz JK, Salemme FR, Todd MJ. Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. Biochemistry. 2005;44(13):5258-66. Epub 2005/03/30. doi: 10.1021/bi048135v. PubMed PMID: 15794662.

92.   Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol. 2008;153 Suppl 1:S7-26. Epub 2007/11/27. doi: 10.1038/sj.bjp.0707515. PubMed PMID: 18037925; PubMed Central PMCID: PMC2268060.

93.   Deng Y, Roux B. Computations of standard binding free energies with molecular dynamics simulations. J Phys Chem B. 2009;113(8):2234-46. Epub 2009/01/17. doi: 10.1021/jp807701h. PubMed PMID: 19146384.

94.   Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical free energy methods for drug discovery: progress and challenges. Curr Opin Struct Biol. 2011;21(2):150-60. Epub 2011/02/26. doi: 10.1016/j.sbi.2011.01.011. PubMed PMID: 21349700; PubMed Central PMCID: PMC3085996.

95.   Aleksandrov A, Thompson D, Simonson T. Alchemical free energy simulations for biological complexes: powerful but temperamental. J Mol Recognit. 2010;23(2):117-27. Epub 2009/08/21. doi: 10.1002/jmr.980. PubMed PMID: 19693787.

96.   Steinbrecher T, Labahn A. Towards accurate free energy calculations in ligand protein-binding studies. Curr Med Chem. 2010;17(8):767-85. Epub 2010/01/22. PubMed PMID: 20088755.

97.   Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. BMC biology. 2011;9:71. Epub 2011/11/01. doi: 10.1186/1741-7007-9-71. PubMed PMID: 22035460; PubMed Central PMCID: PMC3203851.

98.   Moriaud F, Doppelt-Azeroual O, Martin L, Oguievetskaia K, Koch K, Vorotyntsev A, Adcock SA, Delfaud F. Computational fragment-based approach at PDB scale by protein local similarity. J Chem Inf Model. 2009;49(2):280-94. Epub 2009/05/13. doi: 10.1021/ci8003094. PubMed PMID: 19434830.

99.   Lawrenz M, Baron R, McCammon JA. Independent-Trajectories Thermodynamic-Integration Free-Energy Changes for Biomolecular Systems: Determinants of H5N1 Avian Influenza Virus Neuraminidase Inhibition by Peramivir. J Chem Theory Comput. 2009;5(4):1106-16. Epub 2009/05/23. doi: 10.1021/ct800559d. PubMed PMID: 19461872; PubMed Central PMCID: PMC2669418.

100.   Durrant JD, Hall L, Swift RV, Landon M, Schnaufer A, Amaro RE. Novel naphthalene-based inhibitors of Trypanosoma brucei RNA editing ligase 1. PLoS Negl Trop Dis. 2010;4(8):e803. Epub 2010/09/03. doi: 10.1371/journal.pntd.0000803. PubMed PMID: 20808768; PubMed Central PMCID: PMC2927429.

101.   Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev. 2006;106(5):1589-615. Epub 2006/05/11. doi: 10.1021/cr040426m. PubMed PMID: 16683746; PubMed Central PMCID: PMC2547409.

102.   Jones E, Oliphant T, Peterson P, others. SciPy: Open Source Scientific Tools for Python2001.

103.   Lawrenz M, Baron R, Wang Y, McCammon JA. Effects of Biomolecular Flexibility on Alchemical Calculations of Absolute Binding Free Energies. J Chem Theory Comput. 2011;7(7):2224-32. Epub 2011/08/04. doi: 10.1021/ct200230v. PubMed PMID: 21811708; PubMed Central PMCID: PMC3146769.

104.   Bearman GM, Shankaran S, Elam K. Treatment of severe cases of pandemic (H1N1) 2009 influenza: review of antivirals and adjuvant therapy. Recent patents on anti-infective drug discovery. 2010;5(2):152-6. Epub 2010/03/26. PubMed PMID: 20334616.

105.   Oakley AJ, Barrett S, Peat TS, Newman J, Streltsov VA, Waddington L, Saito T, Tashiro M, McKimm-Breschkin JL. Structural and functional basis of resistance to neuraminidase inhibitors of influenza B viruses. J Med Chem. 2010;53(17):6421-31. Epub 2010/08/11. doi: 10.1021/jm100621s. PubMed PMID: 20695427; PubMed Central PMCID: PMC2932999.

106.   Xu X, Zhu X, Dwek RA, Stevens J, Wilson IA. Structural characterization of the 1918 influenza virus H1N1 neuraminidase. J Virol. 2008;82(21):10493-501. Epub 2008/08/22. doi: 10.1128/JVI.00959-08. PubMed PMID: 18715929; PubMed Central PMCID: PMC2573172.

107.   Antonyuk SV, Strange RW, Ellis MJ, Bessho Y, Kuramitsu S, Inoue Y, Yokoyama S, Hasnain SS. Structure of D-lactate dehydrogenase from Aquifex aeolicus complexed with NAD(+) and lactic acid (or pyruvate). Acta Crystallogr Sect F Struct Biol Cryst Commun. 2009;65(Pt 12):1209-13. Epub 2010/01/08. doi: 10.1107/S1744309109044935. PubMed PMID: 20054113; PubMed Central PMCID: PMC2802865.

108.   Xu G, Potter JA, Russell RJ, Oggioni MR, Andrew PW, Taylor GL. Crystal structure of the NanB sialidase from Streptococcus pneumoniae. J Mol Biol. 2008;384(2):436-49. Epub 2008/10/07. doi: 10.1016/j.jmb.2008.09.032. PubMed PMID: 18835278.

109.   Platis D, Smith BJ, Huyton T, Labrou NE. Structure-guided design of a novel class of benzyl-sulfonate inhibitors for influenza virus neuraminidase. Biochem J. 2006;399(2):215-23. Epub 2006/06/17. doi: 10.1042/BJ20060447. PubMed PMID: 16776653; PubMed Central PMCID: PMC1609911.

110.   Crennell SJ, Garman EF, Philippon C, Vasella A, Laver WG, Vimr ER, Taylor GL. The structures of Salmonella typhimurium LT2 neuraminidase and its complexes with three inhibitors at high resolution. J Mol Biol. 1996;259(2):264-80. Epub 1996/06/07. PubMed PMID: 8656428.

111.   White CL, Janakiraman MN, Laver WG, Philippon C, Vasella A, Air GM, Luo M. A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies. J Mol Biol. 1995;245(5):623-34. Epub 1995/02/03. PubMed PMID: 7844831.

112.   Shie JJ, Fang JM, Wang SY, Tsai KC, Cheng YS, Yang AS, Hsiao SC, Su CY, Wong CH. Synthesis of tamiflu and its phosphonate congeners possessing potent anti-influenza activity. J Am Chem Soc. 2007;129(39):11892-3. Epub 2007/09/14. doi: 10.1021/ja073992i. PubMed PMID: 17850083.

113.   Carbain B, Collins PJ, Callum L, Martin SR, Hay AJ, McCauley J, Streicher H. Efficient synthesis of highly active phospha-isosteres of the influenza neuraminidase inhibitor oseltamivir. ChemMedChem. 2009;4(3):335-7. Epub 2009/01/22. doi: 10.1002/cmdc.200800379. PubMed PMID: 19156651.

114.   Deng J, Schnaufer A, Salavati R, Stuart KD, Hol WG. High resolution crystal structure of a key editosome enzyme from Trypanosoma brucei: RNA editing ligase 1. J Mol Biol. 2004;343(3):601-13. Epub 2004/10/07. doi: 10.1016/j.jmb.2004.08.041. PubMed PMID: 15465048.

115.   Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-61. Epub 2009/06/06. doi: 10.1002/jcc.21334. PubMed PMID: 19499576; PubMed Central PMCID: PMC3041641.

116.   Angell RM, Bamborough P, Cleasby A, Cockerill SG, Jones KL, Mooney CJ, Somers DO, Walker AL. Biphenyl amide p38 kinase inhibitors 1: Discovery and binding mode. Bioorg Med Chem Lett. 2008;18(1):318-23. Epub 2007/11/17. doi: 10.1016/j.bmcl.2007.10.076. PubMed PMID: 18006306.

117.   Pettus LH, Wurz RP, Xu S, Herberich B, Henkle B, Liu Q, McBride HJ, Mu S, Plant MH, Saris CJ, Sherman L, Wong LM, Chmait S, Lee MR, Mohr C, Hsieh F, Tasker AS. Discovery and evaluation of 7-alkyl-1,5-bis-aryl-pyrazolopyridinones as highly potent, selective, and orally efficacious inhibitors of p38alpha mitogen-activated protein kinase. J Med Chem. 2010;53(7):2973-85. Epub 2010/03/12. doi: 10.1021/jm100095x. PubMed PMID: 20218619.

118.   Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery

and development settings. Advanced drug delivery reviews. 2001;46(1-3):3-26. Epub 2001/03/22. PubMed PMID: 11259830.

119.   Smith M, Smith MB, March J. Aliphatic Substitution: Nucleophilic and Organometallic. In March's Advanced Organic Chemistry: Reactions, Mechanisms, and Structure. Hoboken, NJ: John Wiley & Sons, Inc.; 2007.

120.   Nunnery JK, Mevers E, Gerwick WH. Biologically active secondary metabolites from marine cyanobacteria. Curr Opin Biotechnol. 2010;21(6):787-93. Epub 2010/10/30. doi: 10.1016/j.copbio.2010.09.019. PubMed PMID: 21030245; PubMed Central PMCID: PMC3034308.

121.   Luesch H, Moore RE, Paul VJ, Mooberry SL, Corbett TH. Isolation of dolastatin 10 from the marine cyanobacterium Symploca species VP642 and total stereochemistry and biological evaluation of its analogue symplostatin 1. J Nat Prod. 2001;64(7):907-10. Epub 2001/07/28. PubMed PMID: 11473421.

122.   Pettit GR, Kamano Y, Herald CL, Tuinman AA, Boettner FE, Kizu H, Schmidt JM, Baczynskyj L, Tomer KB, Bontems RJ. The isolation and structure of a remarkable marine animal antineoplastic constituent: dolastatin 10. Journal of the American Chemical Society. 1987;109(22):6883-5. doi: 10.1021/ja00256a070.

123.   Villa FA, Gerwick L. Marine natural product drug discovery: Leads for treatment of inflammation, cancer, infections, and neurological disorders. Immunopharmacology and immunotoxicology. 2010;32(2):228-37. Epub 2010/05/06. doi: 10.3109/08923970903296136. PubMed PMID: 20441539.

124.   Gunasekera SP, Miller MW, Kwan JC, Luesch H, Paul VJ. Molassamide, a depsipeptide serine protease inhibitor from the marine cyanobacterium Dichothrix utahensis. J Nat Prod. 2010;73(3):459-62. Epub 2009/12/22. doi: 10.1021/np900603f. PubMed PMID: 20020755.

125.   Linington RG, Edwards DJ, Shuman CF, McPhail KL, Matainaho T, Gerwick WH. Symplocamide A, a potent cytotoxin and chymotrypsin inhibitor from the marine Cyanobacterium Symploca sp. J Nat Prod. 2008;71(1):22-7. Epub 2008/01/01. doi: 10.1021/np070280x. PubMed PMID: 18163584; PubMed Central PMCID: PMC2832912.

126.   Gondi CS, Rao JS. Cathepsin B as a cancer target. Expert opinion on therapeutic targets. 2013;17(3):281-91. Epub 2013/01/09. doi: 10.1517/14728222.2013.740461. PubMed PMID: 23293836; PubMed Central PMCID: PMC3587140.

127.   Hook VY, Kindy M, Hook G. Inhibitors of cathepsin B improve memory and reduce beta-amyloid in transgenic Alzheimer disease mice expressing the wild-type, but not the Swedish mutant, beta-secretase site of the amyloid precursor protein. J Biol Chem. 2008;283(12):7745-53. Epub 2008/01/11. doi: 10.1074/jbc.M708362200. PubMed PMID: 18184658.

128.   Sathya M, Premkumar P, Karthick C, Moorthi P, Jayachandran KS, Anusuyadevi M. BACE1 in Alzheimer's disease. Clin Chim Acta. 2012;414:171-8. Epub 2012/08/29. doi: 10.1016/j.cca.2012.08.013. PubMed PMID: 22926063.

129.   Rosenthal PJ. Proteases of malaria parasites: new targets for chemotherapy. Emerg Infect Dis. 1998;4(1):49-57. Epub 1998/03/21. doi: 10.3201/eid0401.980107. PubMed PMID: 9452398; PubMed Central PMCID: PMC2627653.

130.   Kwan JC, Eksioglu EA, Liu C, Paul VJ, Luesch H. Grassystatins A-C from marine cyanobacteria, potent cathepsin E inhibitors that reduce antigen presentation. J Med Chem. 2009;52(18):5732-47. Epub 2009/09/01. doi: 10.1021/jm9009394. PubMed PMID: 19715320; PubMed Central PMCID: PMC2756064.

131.   Pereira AR, Kale AJ, Fenley AT, Byrum T, Debonsi HM, Gilson MK, Valeriote FA, Moore BS, Gerwick WH. The carmaphycins: new proteasome inhibitors exhibiting an alpha,beta-epoxyketone warhead from a marine cyanobacterium. Chembiochem. 2012;13(6):810-7. Epub 2012/03/03. doi: 10.1002/cbic.201200007. PubMed PMID: 22383253; PubMed Central PMCID: PMC3393770.

132.   Leto G, Sepporta MV, Crescimanno M, Flandina C, Tumminello FM. Cathepsin L in metastatic bone disease: therapeutic implications. Biological chemistry. 2010;391(6):655-64. Epub 2010/04/08. doi: 10.1515/BC.2010.069. PubMed PMID: 20370324.

133.   Rosenthal PJ. Cysteine Proteases of Pathogenic Organisms. Robinson MW, Dalton JP, editors: Springer US; 2011.

134.	Funkelstein L, Toneff T, Hwang SR, Reinheckel T, Peters C, Hook V. Cathepsin L participates in the production of neuropeptide Y in secretory vesicles, demonstrated by protease gene knockout and expression. J Neurochem. 2008;106(1):384-91. Epub 2008/04/16. doi: 10.1111/j.1471-4159.2008.05408.x. PubMed PMID: 18410501; PubMed Central PMCID: PMC2661760.

135.	Minokadeh A, Funkelstein L, Toneff T, Hwang SR, Beinfeld M, Reinheckel T, Peters C, Zadina J, Hook V. Cathepsin L participates in dynorphin production in brain cortex, illustrated by protease gene knockout and expression. Mol Cell Neurosci. 2010;43(1):98-107. Epub 2009/10/20. doi: 10.1016/j.mcn.2009.10.001. PubMed PMID: 19837164.

136.	Linington RG, Clark BR, Trimble EE, Almanza A, Urena LD, Kyle DE, Gerwick WH. Antimalarial peptides from marine cyanobacteria: isolation and structural elucidation of gallinamide A. J Nat Prod. 2009;72(1):14-7. Epub 2009/01/24. doi: 10.1021/np8003529. PubMed PMID: 19161344; PubMed Central PMCID: PMC2760338.

137.	Yasothornsrikul S, Greenbaum D, Medzihradszky KF, Toneff T, Bundey R, Miller R, Schilling B, Petermann I, Dehnert J, Logvinova A, Goldsmith P, Neveu JM, Lane WS, Gibson B, Reinheckel T, Peters C, Bogyo M, Hook V. Cathepsin L in secretory vesicles functions as a prohormone-processing enzyme for production of the enkephalin peptide neurotransmitter. Proc Natl Acad Sci U S A. 2003;100(16):9590-5. Epub 2003/07/19. doi: 10.1073/pnas.1531542100. PubMed PMID: 12869695; PubMed Central PMCID: PMC170962.

138.	Copeland RA. Evaluation of enzyme inhibitors in drug discovery. A guide for medicinal chemists and pharmacologists. Methods Biochem Anal. 2005;46:1-265. Epub 2005/12/15. PubMed PMID: 16350889.

139.	Stolze SC, Deu E, Kaschani F, Li N, Florea BI, Richau KH, Colby T, van der Hoorn RA, Overkleeft HS, Bogyo M, Kaiser M. The antimalarial natural product symplostatin 4 is a nanomolar inhibitor of the food vacuole falcipains. Chem Biol. 2012;19(12):1546-55. Epub 2012/12/25. doi: 10.1016/j.chembiol.2012.09.020. PubMed PMID: 23261598; PubMed Central PMCID: PMC3601557.

140.	Conroy T, Guo JT, Linington RG, Hunt NH, Payne RJ. Total synthesis, stereochemical assignment, and antimalarial activity of gallinamide A. Chemistry. 2011;17(48):13544-52. Epub 2011/10/19. doi: 10.1002/chem.201102538. PubMed PMID: 22006835.

141.   Leytus SP, Toledo DL, Mangel WF. Theory and experimental method for determining individual kinetic constants of fast-acting, irreversible proteinase inhibitors. Biochim Biophys Acta. 1984;788(1):74-86. Epub 1984/07/17. PubMed PMID: 6204689.

142.   Tian WX, Tsou CL. Determination of the rate constant of enzyme modification by measuring the substrate reaction in the presence of the modifier. Biochemistry. 1982;21(5):1028-32. Epub 1982/03/02. PubMed PMID: 7074045.

143.   Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. A critical assessment of docking programs and scoring functions. J Med Chem. 2006;49(20):5912-31. Epub 2006/09/29. doi: 10.1021/jm050362n. PubMed PMID: 17004707.

144.   Hardegger LA, Kuhn B, Spinnler B, Anselm L, Ecabert R, Stihle M, Gsell B, Thoma R, Diez J, Benz J, Plancher JM, Hartmann G, Banner DW, Haap W, Diederich F. Systematic investigation of halogen bonding in protein-ligand interactions. Angew Chem Int Ed Engl. 2011;50(1):314-8. Epub 2010/12/25. doi: 10.1002/anie.201006781. PubMed PMID: 21184410.

145.   Hook V, Funkelstein L, Lu D, Bark S, Wegrzyn J, Hwang SR. Proteases for processing proneuropeptides into peptide neurotransmitters and hormones. Annu Rev Pharmacol Toxicol. 2008;48:393-423. Epub 2008/01/11. doi: 10.1146/annurev.pharmtox.48.113006.094812. PubMed PMID: 18184105; PubMed Central PMCID: PMC2731677.

146.   Olsson MHM, Sondergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. Journal of Chemical Theory and Computation. 2011;7(2):525-37. doi: Doi 10.1021/Ct100578z. PubMed PMID: ISI:000287049200025.

147.   Sondergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values. Journal of Chemical Theory and Computation. 2011;7(7):2284-95. doi: Doi 10.1021/Ct200133y. PubMed PMID: ISI:000292617900029.

148. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem. 2006;49(2):534-53. Epub 2006/01/20. doi: 10.1021/jm050540c. PubMed PMID: 16420040.

149. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, Hura GL, Head-Gordon T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys. 2004;120(20):9665-78. Epub 2004/07/23. doi: 10.1063/1.1683075. PubMed PMID: 15267980.

150. Sinko W, de Oliveira C, Williams S, Van Wynsberghe A, Durrant JD, Cao R, Oldfield E, McCammon JA. Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase, an antibacterial target. Chem Biol Drug Des. 2011;77(6):412-20. Epub 2011/02/08. doi: 10.1111/j.1747-0285.2011.01101.x. PubMed PMID: 21294851; PubMed Central PMCID: PMC3095679.

151. Lippert RA, Bowers KJ, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Shaw DE. A common, avoidable source of error in molecular dynamics integrators. J Chem Phys. 2007;126(4):046101. Epub 2007/02/09. doi: 10.1063/1.2431176. PubMed PMID: 17286520.

152. D.A. Case TAD, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman. AMBER 12. University of California, San Francisco. 2012.

153. Durrant JD, McCammon JA. NNScore 2.0: a neural-network receptor-ligand scoring function. J Chem Inf Model. 2011;51(11):2897-903. doi: 10.1021/ci2003889. PubMed PMID: 22017367; PubMed Central PMCID: PMC3225089.

154. Durrant JD, McCammon JA. NNScore: a neural-network-based scoring function for the characterization of protein-ligand complexes. J Chem Inf Model. 2010;50(10):1865-71. doi: 10.1021/ci100244v. PubMed PMID: 20845954; PubMed Central PMCID: PMC2964041.

155.   Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009;30(16):2785-91. doi: 10.1002/jcc.21256. PubMed PMID: 19399780; PubMed Central PMCID: PMC2760638.