



# Design of e-pharmacophore models using compound fragments for the *trans*-sialidase of *Trypanosoma cruzi*: Screening for novel inhibitor scaffolds

Bill R. Miller III, Adrian E. Roitberg\*

Department of Chemistry, Quantum Theory Project, University of Florida, Gainesville, FL 32611-7200, USA



## ARTICLE INFO

### Article history:

Accepted 6 August 2013

Available online 16 August 2013

### Keywords:

*trans*-Sialidase  
Pharmacophore  
Docking  
Molecular dynamics  
MM-GBSA

## ABSTRACT

Chagas' is a fatal disease that affects millions of people worldwide. The lack of safe and effective treatments for Chagas' highlights the need for the discovery of new drugs to fight the disease. *Trypanosoma cruzi*, the parasitic cause of Chagas' disease, synthesizes a *trans*-sialidase (TcTS) enzyme responsible for the transfer of sialic acids from the host cell surface to glycoconjugates on the parasitic cell surface. TcTS has no human analogs and is vital to the life cycle of *T. cruzi*, making TcTS an important enzyme for drug design against Chagas' disease. We use fragment docking to generate various e-pharmacophore hypotheses depicting protein residues important for ligand binding. Virtual screening of the ZINC Clean Leads database with more than 4 million compounds using the e-pharmacophore models found 82 potential inhibitors of TcTS. Molecular dynamics and free energy of binding calculations were used to rank the compounds based on their affinity for TcTS. Two compounds—ZINC13359679 and ZINC02576132—were found to be the most promising lead candidates for TcTS inhibition, and their binding modes are analyzed in detail.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

Chagas' disease [1] (American trypanosomiasis) is a lethal condition that affects nearly 10 million people, mostly in rural regions of South America, while more than 25 million people are at risk of infection worldwide [2]. Due to its low prevalence in developed countries, Chagas' disease has not received the same attention as more notorious diseases such as AIDS, causing the World Health Organization to consider it one of seventeen neglected tropical diseases [3]. Currently, successful treatment of Chagas' disease can only be accomplished if discovered in the acute phase, and diagnosis is difficult given the lack of resources near those infected. Furthermore, the only available drugs—benznidazole and nifurtimox—can have severe adverse side effects [4]. Thus, there exists a need for new drugs to be discovered that can help treat individuals infected with Chagas' disease.

The causative agent of Chagas' disease is the protozoan parasite *Trypanosoma cruzi*, which is mostly transferred between mammalian hosts by triatomine bugs, but can also be transmitted through blood [5] or organ transfusions [6,7], from mother to infant during pregnancies [8], or by ingesting contaminated food or drink [9,10]. *T. cruzi* is unable to synthesize sialic acid, a monosaccharide

involved in many cellular interactions including regulation of the immune system. To combat this deficiency, *T. cruzi* expresses the *trans*-sialidase (TcTS) enzyme that catalyzes the transfer of sialic acid from host sialoglycoconjugates to molecules on the parasitic cell surface, allowing the parasite to successfully evade attacks by the host immune system [11]. Due to its importance in the viability of the infectious parasite and because there is no analogous human enzyme, TcTS is an ideal drug target for the treatment of Chagas' disease. However, despite its high structure and sequence similarity to related sialidases [12], no compounds have been discovered that effectively inhibit TcTS [13]. Initially, potential drug compounds against TcTS were based on the structure of 2,3-dehydro-3-deoxy-*N*-acetylneuraminic acid (DANA), which is a strong transition-state inhibitor of influenza neuraminidase, but a very poor inhibitor of TcTS ( $K_i = 12.3$  mM) [14,15]. Given the lack of success at designing inhibitors that mimic the substrate of TcTS, both experimental and theoretical studies have been reported attempting to discover novel compound scaffolds with inhibitory activity with little overall success [16–18]. To date, the strongest reported inhibitor of TcTS is an anthraquinone derivative with a measured  $IC_{50}$  value of 0.58  $\mu$ M, with a >170-fold specificity for TcTS over human neuraminidase [19]. Despite the discovery of several novel inhibitor scaffolds, we still have not found a compound with inhibitory activity strong enough to be considered a good drug candidate for TcTS.

Molecular docking has become a highly useful tool for drug discovery [20]. In particular, docking molecular fragments has aided in the discovery of new lead compounds for drug design [21]. Due

\* Corresponding author at: 442 Leigh Hall, University of Florida, Gainesville, FL 32611-7200, USA. Tel.: +1 352 392 6972.

E-mail addresses: [roitberg@ufl.edu](mailto:roitberg@ufl.edu), [roitberg@qtp.ufl.edu](mailto:roitberg@qtp.ufl.edu) (A.E. Roitberg).

to their relatively small size, fragments help identify new regions of an enzyme's active site that were previously untargeted [22]. Although most docking scoring functions are not robust, a recent study showed that Schrödinger's Glide [23] can be a useful tool for docking small molecular fragments [24].

Structure-based pharmacophore models have also become increasingly important in computational drug design [25]. Pharmacophore models are used to identify the key features vital to inhibition activity for drug molecules. Traditionally, pharmacophore models are designed using a training set of active molecules. Thus, the question naturally arises what to do prior to any actives being identified for a given enzyme. Recently, Loving et al. developed a method that uses docked fragments to create structure-based energy-optimized pharmacophore (e-pharmacophore) hypotheses [26]. This eliminates the need for known active molecules when generating pharmacophore models and helps identify new regions of the active site to target for drug design. Once developed, these e-pharmacophore models may also be used to screen for new lead drug molecules using the key features of the pharmacophore, which is a fast technique typically used for screening large databases of compounds [27].

Unlike docking that usually provides a single conformation, molecular dynamics (MD) provides insights into the dynamical behavior of protein–ligand complexes. MD is also used to refine biological structures using classical laws of motion to describe the movement of atoms in time. It naturally arises that molecular dynamics can be used to further refine a protein–ligand structure from molecular docking calculations. Additionally, the MD simulations can be used to calculate the relative free energies of binding for a ligand bound non-covalently to a receptor, using such methods as Molecular Mechanics–Poisson Boltzmann Surface Area (MM-PBSA) or Molecular Mechanics–Generalized Born Surface Area (MM-GBSA) [28]. Finally, it has been shown that using these free energy calculations from MD simulations on docked protein–ligand poses can aid the virtual screening process of drug discovery [29].

In this study, we introduce the first e-pharmacophore models of *T. cruzi* trans-sialidase using fragments docked to the active site. Pharmacophore models were generated for the holo crystal structure (PDB code 1SOI), as well as holo and apo conformations of TcTS from previous MD simulations [30]. Using the e-pharmacophore models, we determined the key chemical features important for strong inhibition of the enzyme for each conformation and used those features to screen the Clean Leads database from ZINC [31], followed by MD and relative binding free energies calculations on all hit compounds. The best compounds from binding free energy calculations are presented in detail and proposed as novel lead drug candidates of TcTS. The paper is organized as follows: the computational details of docking, pharmacophore development, screening, molecular dynamics and free energy calculations will be described in Section 2, followed by a thorough analysis of the pharmacophore models, MD simulations, and binding free energy calculations conducted on the hit compounds in Section 3.

## 2. Methods

### 2.1. Receptor structures and grid generation

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

structure of TcTS with the substrate bound, which is often considered the best conformation for inhibitor design. The structures from MD simulations represent relaxed conformations of the protein with the potential to reveal novel inhibitors that a pharmacophore model from the crystal structure is unable to discover. The X-ray crystal structure of TcTS with the substrate (sialyl-lactose) bound was obtained from the Protein Data Bank (PDB code 1SOI [32]) for e-pharmacophore model generation (called the '1SOI' structure from now on), but the unliganded (apo) crystal structure (PDB code 1MS3) was not used because of its inability to capture the flexibility of the catalytic cleft [30]. Instead, a structure of TcTS in the apo conformation was obtained from a MD simulation previously published by Demir and Roitberg [30]. Because the simulation produced thousands of conformations of the protein, clustering was performed with GROMACS [33] to obtain a more manageable number of statistically different conformations. The snapshots from the apo simulation were clustered with a root-mean-square deviation (RMSD) cutoff of 0.125 Å for all heavy atoms in the active site (defined as any residue within 15 Å of the nucleophilic Tyr342 residue). The representative conformation of the most populated cluster (62% of the total structures) was chosen as the apo structure of TcTS for pharmacophore model creation, and will be referred to as the 'apo' structure henceforth. For comparison purposes, clustering was also performed on a MD trajectory of TcTS bound to the substrate with a RMSD cutoff of 0.085 Å to maintain approximately the same number of structures in the most populated cluster between the two simulations. The most populated cluster corresponded to 56% of the total structures from the simulation, and the representative structure from this cluster was chosen for pharmacophore hypothesis generation, and will be referred to as the 'holo' structure.

All three TcTS structures were prepared using the Protein Preparation Wizard (PrepWizard) in Schrödinger, which is shown to improve virtual screening enrichment by fixing common problems with initial protein structures [34]. The PrepWizard added any necessary hydrogen atoms and removed all water molecules in preparation for the receptor grid generation. The substrate location in the 1SOI structure was used to choose the center and size of the receptor grid, which was generated using Schrödinger's Glide with default settings for all parameters. The grid size was chosen sufficiently large to include all active site residues involved in substrate binding, including residues at the mouth of the active site used to position the acceptor/donor lactose molecule during catalysis. The grids for the apo and holo structures were chosen to have the same size and corresponding centroid point of the active site for consistency.

### 2.2. Docking fragments to TcTS structures

The Clean Fragments library of 504,074 compounds was downloaded in structure-data file (SDF) format from the ZINC [31] website. LigPrep [35], which prepared ligands for docking with Glide, was used to generate reasonable conformers of each fragment using the default settings. All ligand conformers were docked to each of the three receptor grid files (1SOI, apo, and holo) using Schrödinger's Glide [23] in High-Throughput Virtual Screening (HTVS) mode with default settings to filter out completely unreasonable ligands from being considered. The highest-ranking 120,000 compounds from virtual screening were re-docked using Glide SP with default settings, followed by re-docking of the top 6513 fragments from Glide SP using Glide XP [36]. The Glide XP settings were chosen based on the protocol of Loving et al. [26] determined to be ideal for generating e-pharmacophores.

### 2.3. E-pharmacophore generation

The results from fragment docking with Glide XP were used to generate e-pharmacophore hypotheses for each TcTS conformation using the E-pharmacophore script [37] available from the Schrödinger Script Center. The E-pharmacophore script works by mapping the Glide XP energies for the top 2000 docked fragment poses onto the corresponding ligand atoms, aligned to the receptor conformations, followed by volume clustering of the fragments into 15 distinct clusters. Phase [38] was used to generate the pharmacophore hypotheses from the docked poses of the clusters using the default set of six chemical features in Phase: hydrogen bond acceptor, hydrogen bond donor, negative ionizable, positive ionizable, aromatic ring, and hydrophobic. The Glide XP energies for each pharmacophore site in the hypothesis are summed and ranked based on their total energetic contributions. A van der Waals scaling of 0.5 was used for receptor-based excluded volumes in each hypothesis to account for the shape of the active site. For each hypothesis, the 15 most favorable pharmacophore sites were kept, which is more than typical hypotheses. For TcTS, we were attempting to discover new regions of the active site that could lead to novel inhibitors, so analyzing more sites increases our likelihood of doing so.

### 2.4. Pharmacophore screening

We searched the Clean Leads library of 4,230,832 compounds in order to find novel inhibitor scaffolds of TcTS. This library represents much of chemical space, which is ideal for this type of screening. Molecules were required to match a minimum of 3 or 4 sites on each hypothesis to be considered a “hit compound” (see Section 3 for a detailed explanation of the required sites chosen for each hypothesis), but preference was given to any molecules that matched more than the required sites. The chemical features of the required pharmacophore sites were used to filter the Leads library with Schrödinger’s Maestro v9.3 interface for only compounds that contained those features. Filtering the initial database saves time, but should have no effect on the screening results since the excluded compounds would never be chosen by Phase as hit compounds anyway. During the screening process, conformers were generated for each ligand using a rapid conformational sampling procedure. Ten conformers were created for each rotatable bond of each ligand, with a maximum of 100 conformers generated per structure (both defaults), which was a reasonable value given that the lead compounds are all relatively small and typically contained less than 10 rotatable bonds. Excluded volumes, required matches, and an intersite distance matching tolerance of 2.0 Å for all sites were applied during the screening process. Phase attempts to maximize overlap between the ligand conformations and the locations of the pharmacophore sites, which means the protein structure is not explicitly included during the virtual screening process. Only one structure was returned for each compound that matched the hypothesis. Finally, default settings were used for all coefficients when calculating the Phase fitness score and ranking the hit compounds.

### 2.5. Molecular dynamics and free energy calculations

To refine the ligand structures from Phase, the GPU accelerated *pmemd* code [39] of Amber 12 [40] was used to perform MD on each hit compound in complex with TcTS. Because the screening process does not include the protein structure, several ligands had atoms that overlapped with protein atoms, so the atom positions were manually moved using VMD [41] to remove any overlap prior to MD. The *antechamber* module of Amber was used to calculate the AM1-BCC charges for all ligands. The *gaff* [42] and *ff99SB* [43]

force fields were used to define all remaining parameters for the hit compounds and TcTS, respectively. All complexes were neutralized with counter ions and solvated with TIP3P water [44] in a truncated octahedron with a minimum of 12.0 Å between the solute and the edge of the unit cell. All initial structures underwent a seven-step minimization procedure using positional restraints on all solute heavy atoms, and slowly reducing the restraint weight from 10.0 to 0.0 kcal/mol/Å<sup>2</sup>. After minimization, the protein was restrained with a positional restraint of 10.0 kcal/mol/Å<sup>2</sup> while the system was linearly heated from 10 K to 300 K over 2.0 ns. A seven-step equilibration period over 3.5 ns was used to slowly remove the positional restraints on the protein from 10.0 to 0.0 kcal/mol/Å<sup>2</sup>. Finally, 10.0 ns of unrestrained MD was performed on all solvated protein–ligand complexes with coordinates saved every 1.0 ps, which was used for all analyses. The SHAKE algorithm [45] was applied to fix all covalent bond distances involving hydrogen, allowing a time step of 2 fs for dynamics. All simulations were performed using periodic boundary conditions under constant pressure and temperature using an Andersen thermostat [46].

Hydrogen bonding analysis was performed using the *cpptraj* module of AmberTools using default hydrogen bond definitions. Single trajectory free energy of binding calculations were performed on 200 evenly spaced snapshots from the unrestrained MD simulation with the *MMPBSA.py* program [47] released alongside AmberTools using the Generalized-Born (GB) model [48] to calculate solvation free energies. GB is a good approximation of solvation energies compared to more rigorous free energy perturbation calculations when computing relative free energies of binding [49]. We assume the solute entropy for all systems remains the same, so all entropic contributions to binding were neglected during free energy calculations [28]. It is important to note that the calculated free energies are not expected to be true binding free energies, and are used only to compare relative to one another.

## 3. Results and discussion

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

### 3.1. Pharmacophore hypotheses

#### 3.1.1. 1S0I pharmacophore model

This pharmacophore model was generated using the TcTS conformation from the 1S0I PDB X-ray crystal structure. Table 1 provides a breakdown of all 15 pharmacophore sites generated by Phase, including the corresponding interaction with TcTS each site represents. Several of the pharmacophore sites found by Phase were expected. For example, it is well known that the arginine triad (Arg35, Arg245, and Arg314) is important for stabilizing the substrate carboxylic acid during ligand binding for both sialidases and TcTS [50], so it was not surprising that feature N10 was the most favorable site. Furthermore, the acceptor binding site between residues Tyr119 and Trp312 at the mouth of the binding site was previously shown to be important for substrate positioning in the active site [30], explaining the presence of feature R20 in Table 1. The occurrence of these two features (N10 and R20) is a testament to the validity of the method at selecting reasonable pharmacophore sites.

Clearly, no single drug molecule would be able to form all 15 interactions described by the pharmacophore model. Thus, when screening for new inhibitors, we decided to mandate that a compound match four of these features to be considered a “hit compound” because we found that using more than four features resulted in no hit compounds. Phase provides the ability to not

**Table 1**

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

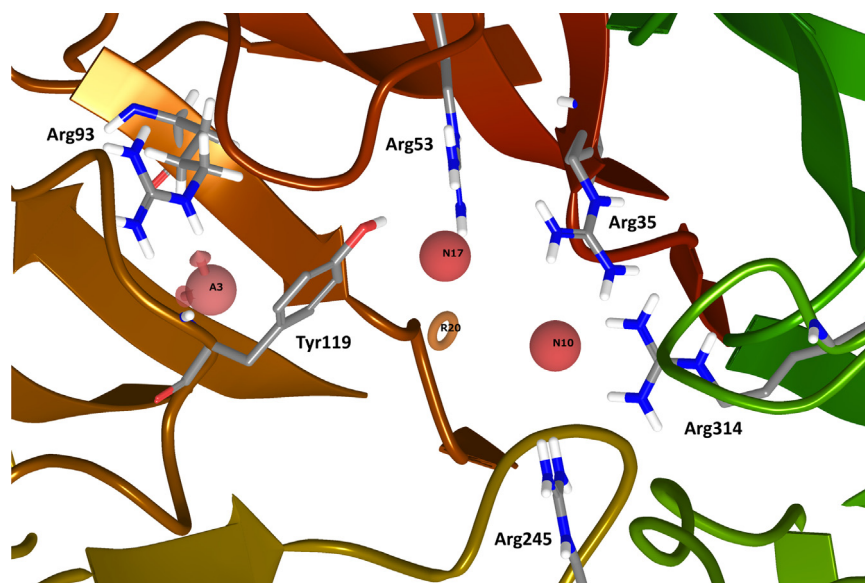
Rank	Feature label	Score	Interaction with TcTS
1	N10*	−13.39	Arginine triad (Arg35, Arg245, Arg314)
2	N109	−4.19	Arg35 and Arg53
3	A18	−3.14	Backbone —NH of Asp59 and side chain —OH of Tyr119
4	H84	−2.52	Hydrophobic pocket near Val95 and Leu176
5	D69	−2.20	Side chain carbonyl oxygen of Asn60
6	A3*	−1.94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

specify the four mandated features, allowing any four features to be matched. However, this is very computationally demanding using a pharmacophore model with 15 features and seems unnecessary since we have knowledge of several important features *a priori*. We decided to select the four required features based on our understanding of the protein structure and catalysis. Initially, we chose both N10 and R20 because of their previously mentioned importance for substrate binding. Our third choice was N17 (negative ionizable and hydrogen acceptors were both considered matches) because Pierdominici-Sottile et al. recently found Arg53 to be the most stabilizing non-catalytic residue of the transition state complex for TcTS [51]. We wanted the final required feature to be a new site that has not been previously targeted by inhibitor design on TcTS. For this feature, we wanted to avoid parts of the active site that were solvent exposed and near the edge of the active site to avoid non-specific interactions and competition with the solvent. Feature A3 was chosen because it was in a back pocket of the active site and contained a key target residue—Arg93. This pocket is mostly hydrophobic except for the Arg93 residue, and is known to bind the *N*-acetyl group of the substrate [50]. From our understanding, Arg93 has never been targeted for drug design and thus has the ability to aid in the discovery of novel inhibitors against

TcTS. It would have been easy to choose the four features with the highest score from Phase, but we believe that the four sites chosen here (N10, R20, N17, and A3—Fig. 1) are logical choices based on the available information on TcTS.

### 3.1.2. Holo pharmacophore model

The holo pharmacophore model was generated using a TcTS structure from clustering of an MD simulation with TcTS bound to sialyl-lactose. The pharmacophore sites found by Phase for this model are described in Table 2. There are many similarities between this model and the 1S0I hypothesis described above. Once again, the top-ranking site was a negative ionizable feature (N39) interacting with the arginine triad. Other sites from the holo pharmacophore model that were very similar to the 1S0I model include N38, D41, A34, D55, and R137. Given that the 1S0I and holo pharmacophore models were generated using different conformations of TcTS, differences in the features were anticipated. The major differences included the presence of features interacting with Trp120, Arg145 and Ser57 only in the holo pharmacophore, and the presence of sites interacting with Ala59, Asn60, Ser229, Gln195, Gln362, and Asp247 only in the 1S0I hypothesis. However, several of these differences are considered irrelevant because the TcTS residues involved



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



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

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

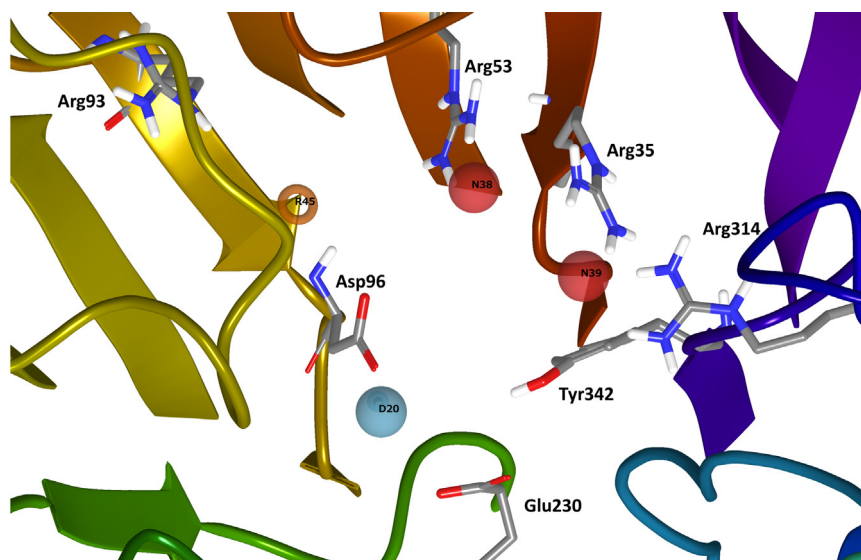
are mobile and/or solvent exposed, such as Ser57, Asn60, and Gln362.

When selecting four sites as required matches when screening possible inhibitors, we wanted to choose sites that were different from the 1S0I model, but still maintained at least two features known as important for substrate binding to TcTS. The differences were to ensure the same hit compounds were not discovered while screening compounds against the pharmacophore models. Thus, features N39 and N38 were chosen as the two sites that were consistent from the 1S0I model, while the acceptor binding site feature was excluded to increase hit compound diversity. As mentioned previously, the hydrophobic pocket near Val95 and Leu176 contains the *N*-acetyl group of the substrate, but Phase found the pocket large and favorable enough to accommodate an aromatic ring. This finding is contrary to most TcTS drug design efforts that typically try to place a similar *N*-acetyl group at this location, so we selected R45 to represent this region for screening. Finally, feature D20 was chosen as a new region of the active site to target. Investigations of the TcTS structure near feature D20 discovered that this region is an ideal location for a hydrogen donor because of the presence of nearby negatively charged residues Asp96 and Glu230. Furthermore, Glu230 is important to the catalytic mechanism as a proton acceptor for the nucleophilic Tyr342 [52] and

stabilizing the substrate transition state [51], which makes it an ideal drug design target residue. To the best of our knowledge, our screening approach is novel because no one has attempted placing an aromatic ring in the hydrophobic pocket near Val95/Leu176 or targeted the Asp96/Glu230 motif when designing drugs against TcTS, which makes these four features—N39, N38, R45, and D20 (Fig. 2)—ideal for screening new inhibitors.

### 3.1.3. Apo pharmacophore model

The apo conformation of TcTS was the most unique of the three studied because of the flexibility of the unliganded enzyme [30]. This was the only pharmacophore model that did not include a negative ionizable site interacting with the arginine triad as the top-ranking feature (Table 3). Instead, a negative ionizable charge interacting with the side chain of Arg93 was the most favorable site (N136). The negative site interacting with the arginine triad (N132) unexpectedly ranked fifth among sites for the apo model. Asp59 and Glu227 are the only residues involved in an interaction with a pharmacophore site in the apo model that was not present in either of the other two models. Given the role of Asp59 as an acid/base catalyst for TcTS, it is interesting to note that the other pharmacophore models did not produce any sites interacting with the side chain of Asp59, while the apo model did. However, in the



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

**Table 3**

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

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

apo conformation, Asp59 is unlikely to be a residue of high interest to target for drug design because it is almost completely solvent exposed. Expectedly, Phase found no pharmacophore site in the acceptor binding site between Tyr119 and Trp312 because the flexibility of the Trp312 loop precludes an aromatic ring from binding to this site in this conformation [30]. The re-ranking and scoring of several similar sites among the three pharmacophore models further enhances the idea that the pure rankings should not be the ultimate deciding factor when considering what pharmacophore sites to select as required matches during the screening process.

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

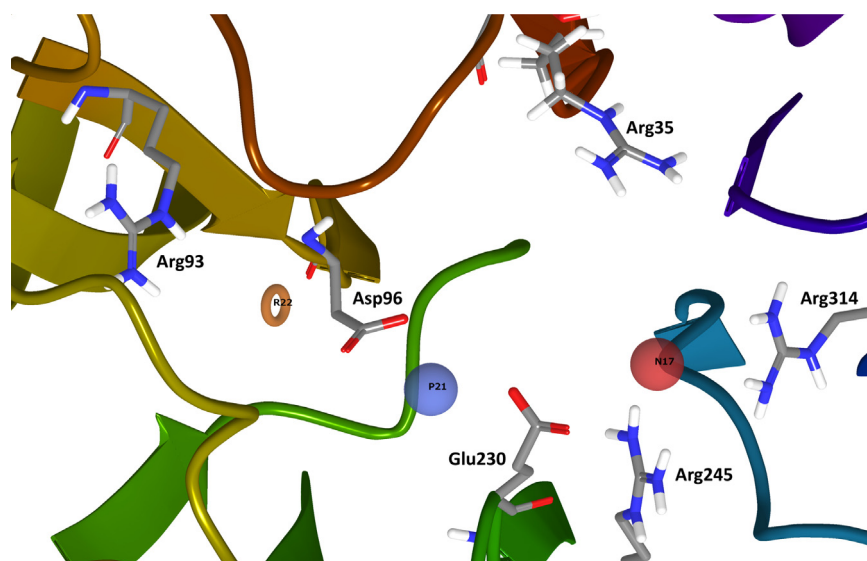
filled by the *N*-acetyl group of sialyl-lactose. Choosing both sites allowed us the ability to screen for compounds that were specifically suited for the sialic acid binding site and the *N*-acetyl binding region. Unfortunately, using these four sites—N17, P21, N136, and R22—resulted in no hit compounds during screening of the leads database. As Demir and Roitberg [30] noted, the active site of apo TcTS is much larger and more open than the holo structure, and thus any hit compounds from screening would need to be larger to match the sites in this active site compared to the other pharmacophore models. Large molecules often do not make ideal drugs and the ZINC Leads database screened here primarily consists of smaller lead scaffolds, not large molecules. Consequently, we trimmed the required matches down from four to three and attempted using both sites as the third and final required site. The only combination that resulted in any hit compounds was using N17, P21, and R22 (Fig. 3). Thus, these three sites were chosen as the required features for the screening results discussed below for the apo pharmacophore model.

### 3.2. Screening results using TcTS pharmacophore models

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

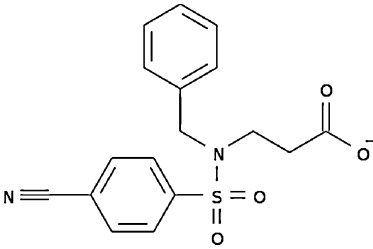
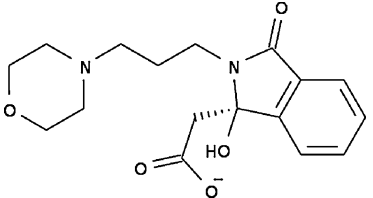
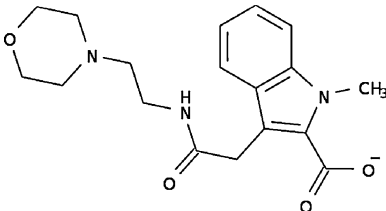
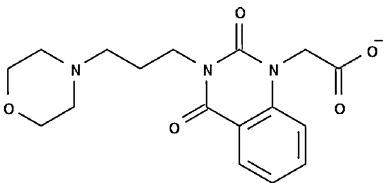
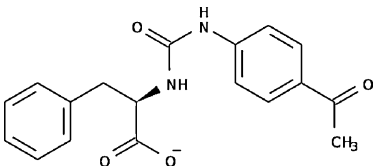
#### 3.2.1. Screening results for 1S0I pharmacophore

Screening of the ZINC Leads database using the 1S0I pharmacophore resulted in 25 hit compounds. After performing MD and calculating the MM-GBSA binding free energy for each compound, the potential inhibitors were re-ranked and the best five compounds are shown in Table 4 (see Supporting Information for all compounds). For the Phase Fitness scores, the higher value the better the compound fits the shape and topology of the pharmacophore model. Conversely, the more negative the MM-GBSA binding free energy the better the compound is expected to bind to TcTS. Aside from the compounds, Table 4 makes it clear why MD and MM-GBSA binding free energy calculations were necessary—there is no correlation between the Phase Fitness score and the MM-GBSA binding free energies. This observation is consistent for all hit compounds found by Phase from the 1S0I pharmacophore as well as the holo and apo (see Supporting Information). Because the pharmacophore screening includes no specific details of the receptor it



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

**Table 4**  
The best five hit compounds from screening results of the 1S0I pharmacophore according to MM-GBSA binding free energies. The Phase Fitness scores are unitless, while the MM-GBSA have units of kcal/mol. The uncertainties reported represent the standard error of the mean. See Supporting Information for the results of all 25 hit compounds. The binding mode of ZINC13359679 is discussed in more detail in Section 3.3.1.

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

would be unrealistic to expect the Phase Fitness to correspond with the MM-GBSA binding energies that are calculated as an average over 200 conformations of the compound noncovalently bound to the active site of TcTS. For drug design and discovery, Phase Fitness scores should be disregarded and replaced by a more sophisticated and rigorous binding affinity calculation that is known to accurately predict relative free energy differences, such as MM-GBSA [53].

As all of these compounds are derived from the 1S0I pharmacophore, they have some structural similarities because of the required features imposed during the screening process. Per the pharmacophore model, every hit compound contains a negative ionizable functional group (typically a carboxylic acid) that interacts with the arginine triad, a hydrogen acceptor or negative ionizable group that interacts with Arg53, a hydrogen acceptor interacting with Arg93, and an aromatic ring designed to fit into the acceptor binding site between Tyr119 and Trp312. Overall, the 1S0I pharmacophore model resulted in the best hit compounds of the three pharmacophore models according to the binding energies, with four compounds with binding free energies better than

-50 kcal/mol, and ten compounds better than -40 kcal/mol (once again, these are not considered real binding free energies and are only used to compare relative to one another). The better TcTS inhibitors likely arise from the 1S0I model because of the pharmacophore sites chosen as requirements during the screening process. The 1S0I pharmacophore was the only hypothesis that required hit compounds to have a presence in both the sialic acid and acceptor binding sites of TcTS, which is believed to yield more specific TcTS inhibitors than inhibitors that simply target the sialic acid binding site. The results here further support the idea that a strong TcTS inhibitor should target both of these sites.

The MM-GBSA binding free energies reported here are relative to one another and are used to compare directly to the calculated binding free energy of the TcTS substrate, sialyl-lactose. MM-GBSA calculations performed on the MD simulation of sialyl-lactose bound to TcTS by Demir and Roitberg [30] determined that the average binding free energy was  $-54.8 \pm 0.1$  kcal/mol for sialyl-lactose using the same protocol reported here for the hit compounds. Thus, the three most promising hit compounds discovered

using the 1S0I pharmacophore—ZINC13359679, ZINC23187566, and ZINC20107259—are projected to bind TcTS stronger than the natural substrate. ZINC13359679 is calculated to have a significantly stronger binding affinity (−76.3 kcal/mol) than the substrate (−54.8 kcal/mol), and thus the MD of this compound will be analyzed in more detail in Section 3.3.1. As is, ZINC23187566, ZINC20107259, and ZINC73658621 are likely to compete with sialyl lactose binding *in vivo*, but are not likely to be much better than the substrate. However, all of these compounds are good scaffolds for drug design efforts against TcTS because their structures are unique compared to any previously reported TcTS potential inhibitors. Modifications could be made to optimize each of the hit compounds in Table 4 to improve their binding affinity for TcTS.

It is important to note that unlike for the screening process, the binding energies are averages based on a dynamic process. Many of the hit compounds are initially in a single, ideal conformation within the TcTS binding pocket from screening that is not maintained during the MD simulation. Instead, the binding free energy often gets worse during the simulation as the contacts between the protein and ligand relax until the ligand reaches a state of equilibrium with the protein that results in a binding free energy that oscillates around a single value. This natural worsening of the binding free energy during simulations is a result of the screening process only producing a single ligand conformation and that ligand conformation is not even directly based on the protein conformation.

### 3.2.2. Screening results for holo pharmacophore

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

These compounds all have a set of common features that can align to the required pharmacophore sites of the holo e-pharmacophore. Each compound has a negative ionizable group positioned to interact with the arginine triad, a hydrogen acceptor or negative ionizable group positioned to hydrogen bond with Arg53, an aromatic ring to fill the hydrophobic pocket near Val95 and Leu176, and a hydrogen donor group that can form interactions with the side chains of Asp96 and/or Glu230. Overall, the hit compounds from the holo pharmacophore are not as strong of inhibitors compared to the compounds from the 1S0I pharmacophore screening. Only two compounds—ZINC02576132 and ZINC40351842—resulted in binding energies that were better than sialyl-lactose, while the other 28 hit compounds were worse. Similar to the results from the 1S0I pharmacophore screening, one compound, ZINC02576132, stands out as significantly better than sialyl-lactose and the other hit compounds with a binding free energy of −73.2 kcal/mol. For this reason, the MD and binding mode of ZINC02576132 will be analyzed in more detail in Section 3.3.2. The free energy of binding results suggests that ZINC40351842 and ZINC60181779 have a similar binding affinity as the substrate, and could compete for binding *in vivo*, while the other hit compounds are unlikely to be anything but very weak inhibitors of TcTS. However, rational drug design efforts could use these compounds as new scaffolds for modification and optimization that improve their TcTS binding affinities.

### 3.2.3. Screening results for the apo pharmacophore

Screening of the ZINC Leads database using the apo pharmacophore resulted in 27 hit compounds. After performing MD and calculating the average MM-GBSA binding free energy for each compound, the potential inhibitors were re-ranked and the best five compounds are shown in Table 6 (see Supporting Information for all compounds).

The hit compounds from the apo pharmacophore screening were expected to be more chemically diverse compared to the 1S0I and holo pharmacophore results because only three pharmacophore sites were required for hit compounds instead of four like the 1S0I and holo models. Each hit compound was required to have a negative ionizable charge that could interact with the arginine triad, an aromatic ring for the hydrophobic pocket near Val95 and Leu176, and a positive ionizable group for hydrogen bonds with the side chains of Asp96 and Glu230. Unexpectedly, many of the hit compounds had very similar chemical structures, such as ZINC72447098, ZINC72432649, ZINC72477708, and ZINC72407055 (see Supporting Information). The likely cause of this result is the shape and topology of the apo pharmacophore required sites. The TcTS active site is known to be larger and more open in the apo conformation, which reduces the number of likely hit compounds because any potential hit compound must be large enough to accommodate the bigger active site. Furthermore, one of the required pharmacophore features included a positive ionizable group, which significantly reduces the number of compounds that could match this pharmacophore model.

Unlike the screening results from the 1S0I and holo pharmacophores, no hit compound from the apo pharmacophore screening was discovered with a binding free energy significantly better than sialyl-lactose. In fact, the best hit compound, ZINC72447098, had an average binding free energy identical to sialyl-lactose (−54.8 kcal/mol). At best, ZINC72447098 would bind to TcTS with about the same affinity as sialyl-lactose, while all the other hit compounds would bind worse. Of course, these compounds should all be considered lead candidates for rational drug design against TcTS, regardless of their relatively poor calculated binding free energies compared to the 1S0I and holo screening results. Specifically, the chemical scaffolds that are observed multiple times in Table 6 should be considered for optimization given their prevalence as hit compounds discovered using the apo pharmacophore. The repeated occurrence of compounds similar to ZINC72447098 may indicate that this scaffold is viable as an inhibitor of TcTS with some optimization.

## 3.3. Analysis of the most promising TcTS inhibitors

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

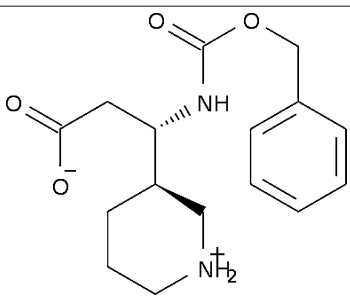
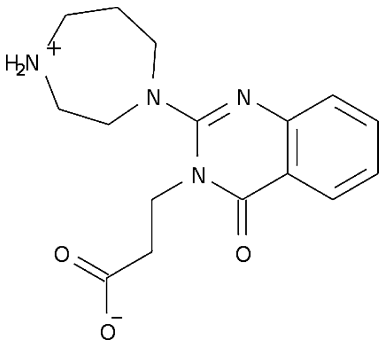
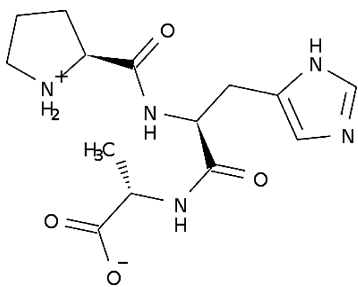
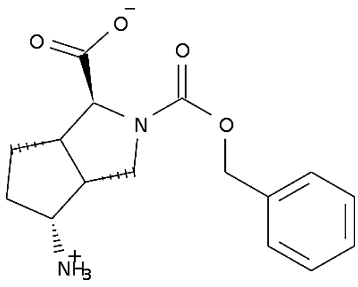
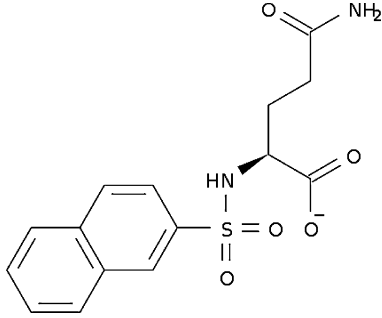
### 3.3.1. ZINC13359679

ZINC13359679 was the top hit compound discovered using the 1S0I pharmacophore, with a MM-GBSA binding free energy of −76.3 kcal/mol (Table 4). The protein and ligand were found to be relatively stable during the 10 ns production trajectory, with an RMSD of less than 2.0 Å for most of the simulation (Fig. 4). As the RMSD plot shows, neither the protein nor the ligand are undergoing any major conformational changes after the initial relaxation and equilibration steps, suggesting ZINC13359679 is stable within the active site of TcTS. The two spikes in RMSD near 1.5 ns and 4.3 ns are related to movements near the periphery of the enzyme and



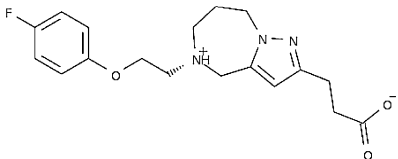
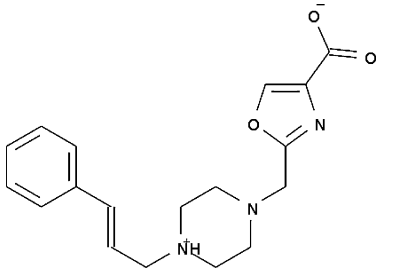
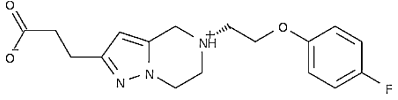
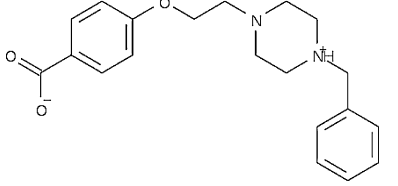
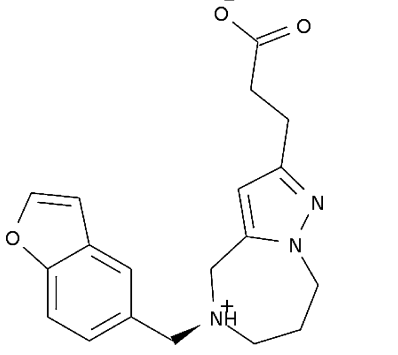
**Table 5**

The best five hit compounds from screening results of the holo pharmacophore according to MM-GBSA binding free energies. The Phase Fitness scores are unitless, while the MM-GBSA energies have units of kcal/mol. The uncertainties reported represent the standard error of the mean. See Supporting Information for the results of all 30 hit compounds. The binding mode of ZINC02576132 is discussed in more detail in Section 3.3.2.

Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
1	ZINC02576132		0.194	−73.2 ± 0.5
2	ZINC40351842		0.145	−55.1 ± 0.4
3	ZINC60181779		0.198	−52.2 ± 0.4
4	ZINC19795370		0.184	−48.5 ± 1.0
5	ZINC03406782		0.197	−45.0 ± 0.5

**Table 6**

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

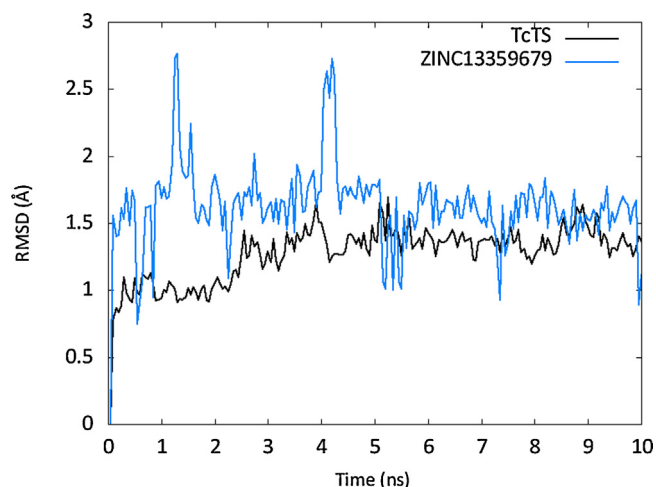
Rank	ZINC code	Structure	Phase Fitness	MM-GBSA
1	ZINC72447098		0.155	−54.8 ± 0.3
2	ZINC67946325		0.186	−50.1 ± 0.4
3	ZINC72145690		0.144	−44.0 ± 0.5
4	ZINC19850530		0.168	−42.0 ± 0.3
5	ZINC72432649		0.168	−41.8 ± 0.3

should not be associated with instability of the enzyme or active site.

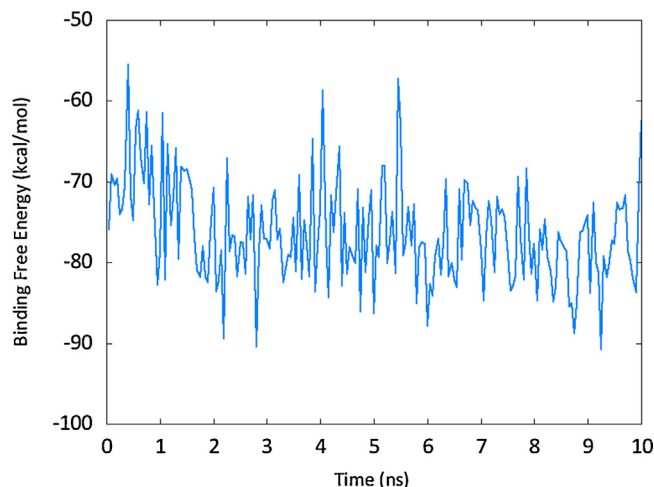
Since the binding free energy reported represents an average value of a dynamic property, it is also informative to examine the binding free energy as a function of time (Fig. 5). The binding energy fluctuates as the simulation progresses, but the value oscillates around the average value and does not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

The contacts between TcTS and ZINC13359679 are a clear mixture of polar and nonpolar interactions (Fig. 6). The strongest set of polar interactions naturally exists between the carboxylic acid

and the arginine triad, which is also important for stabilizing the carboxylic acid of DANA (a poor inhibitor of TcTS) in the active site. In fact, hydrogen bonding analysis shows that hydrogen bonds exist between ZINC13359679 and the three residues that comprise the arginine triad (Arg35, Arg245, and Arg314) more than 80% of the simulation (Fig. 6a). Interactions with the arginine triad are a key component of any potential inhibitor of TcTS. The initial protein–ligand conformation from pharmacophore screening included the sulfate functional group forming hydrogen bonds with Arg53, and the cyano group hydrogen bonded to Arg93. However, relaxation of the ligand complexed with TcTS significantly reduced both of these interactions. The sulfate group is never close enough to form a hydrogen bond with Arg53 during the simulation, but is likely close enough (~4.25 Å) to make favorable Coulombic interactions for the majority of the simulation. ZINC13359679 shifted slightly during relaxation to a conformation with the cyano group of the ligand hydrogen bonding to the side chain of Thr121 for



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

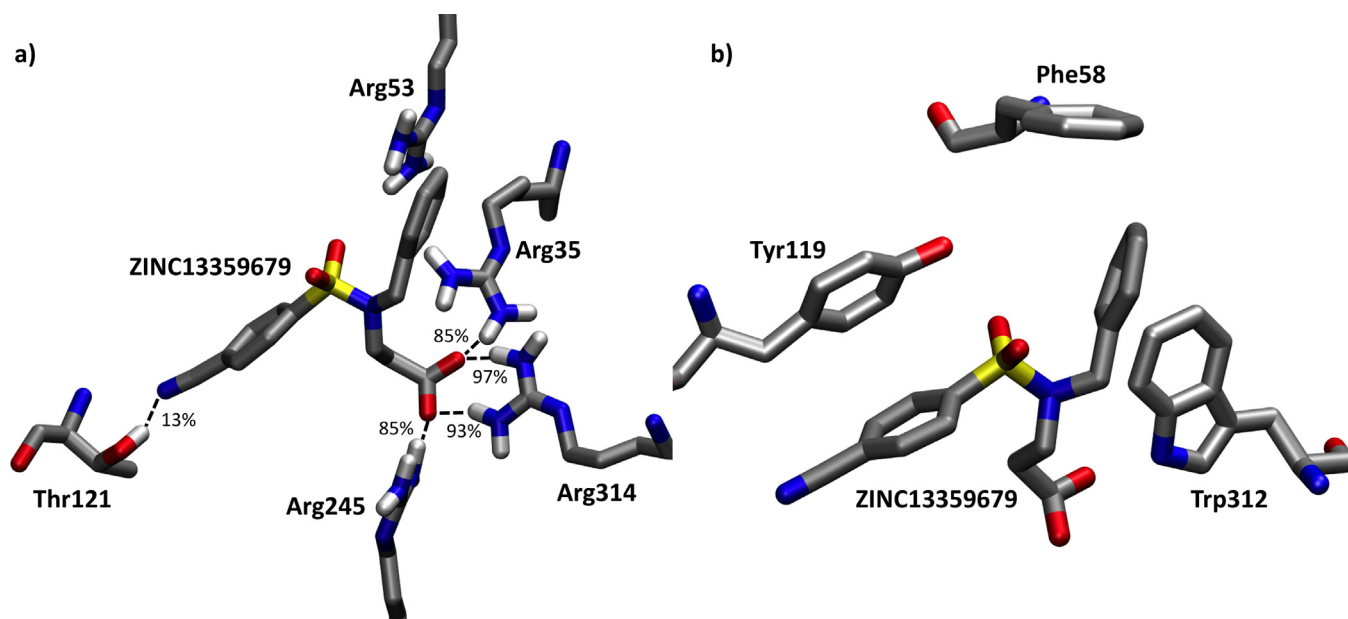


**Fig. 5.** The MM-GBSA binding free energy of ZINC13359679 bound to TcTS as a function of time for a 10-ns MD simulation.

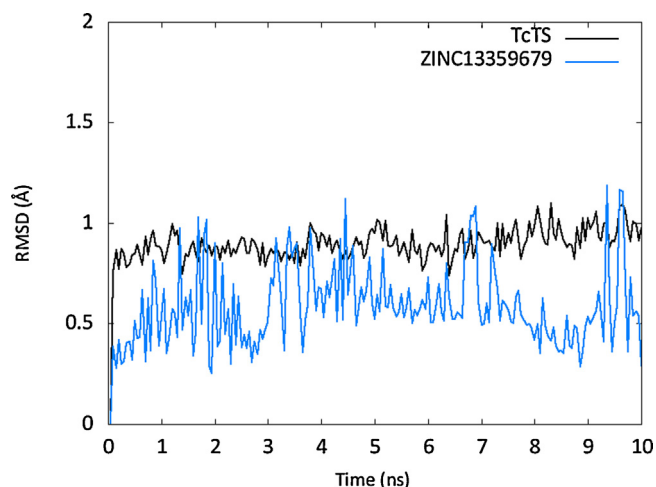
approximately 13% of the simulation instead of Arg93. Neither of these interactions are seen between DANA and TcTS, which helps suggest ZINC13359679 would make a better inhibitor than TcTS.

The dominating nonpolar protein–ligand interactions involve  $\pi$ -stacking between aromatic rings on the ligand and three TcTS residues—Phe58, Tyr119, and Trp312 (Fig. 6b). These favorable protein–ligand aromatic interactions are not present in many inhibitors of TcTS discussed in the literature, including DANA, because they lack an aromatic ring in the acceptor binding site. The 1S0I pharmacophore contained a required feature involving an aromatic ring in the acceptor binding site between Tyr119 and Trp312, but the  $\pi$ -stacking interaction with Phe58 was unexpected. In the initial conformation, Phe58 is not involved in any protein–ligand interactions, but  $\sim 6.7$  ns into the simulation, the side chain of Phe58 changes conformation and forms a T-shaped  $\pi$ -stacking interaction with a phenyl ring on ZINC13359679. In this conformation, Phe58 is positioned over the active site like a lid preventing the ligand from exiting. Trp312 is also making a T-shaped  $\pi$ -stacking

interaction with the same phenyl ring of ZINC13359679 as Phe58, while Tyr119 forms a parallel  $\pi$ -stacking interaction with the other ligand aromatic ring. Phe58 has no known role in substrate binding, and is unlikely to form a similar interaction with the substrate given that the lactose in the acceptor binding site is always attached to another sugar, preventing Phe58 from “closing” the top of the active site. ZINC13359679 forces Tyr119 and Trp312 of the acceptor binding site to deviate from their traditional conformation with sialyl-lactose bound. Typically, Tyr119 and Trp312 ‘sandwich’ a lactose acceptor molecule, which means they are forming nonpolar interactions with the same sugar ring, but with ZINC13359679 they are  $\pi$ -stacking with different aromatic rings. This is a reasonable observation given the known flexibility of the Trp312 loop on TcTS [30]. Although ZINC13359679 did not maintain all of the interactions originally described by the 1S0I pharmacophore model, it was able to form many favorable polar and nonpolar interactions with TcTS that make ZINC13359679 a strong candidate to inhibit TcTS.



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



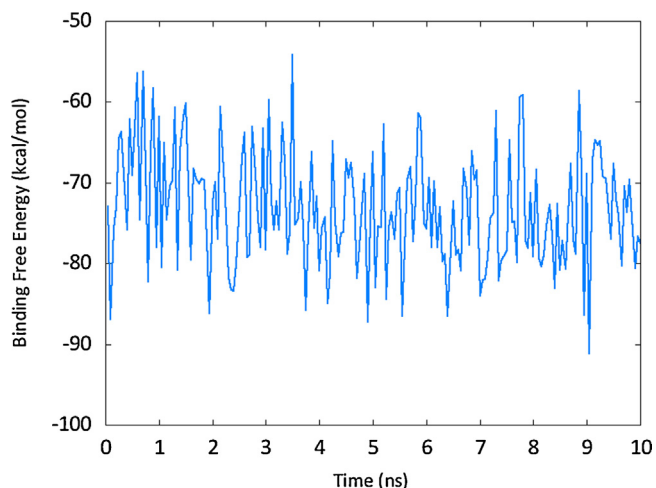
**Fig. 7.** The RMSD of TcTS and ZINC02576132 using the initial frame as reference. The RMSD was calculated using only the backbone atoms for TcTS, and all atoms for ZINC02576132.

### 3.3.2. ZINC02576132

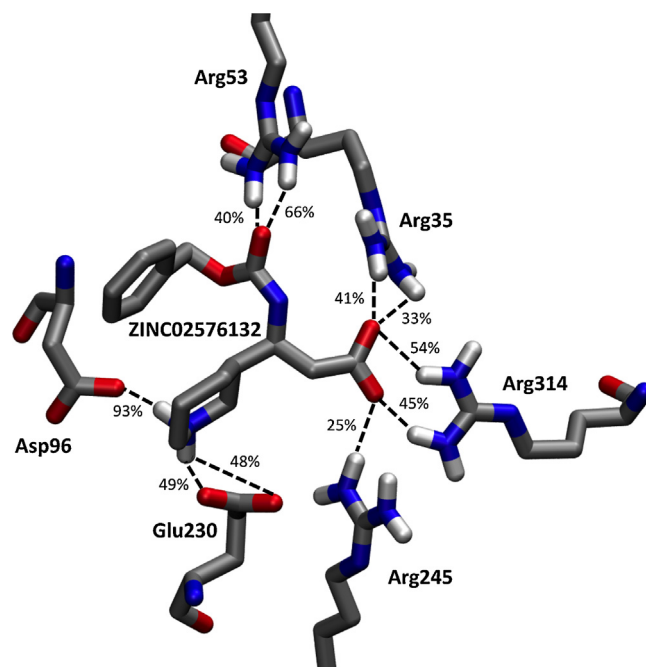
ZINC02576132 was the top hit compound discovered using the holo pharmacophore, with a MM-GBSA binding free energy of  $-73.2$  kcal/mol (Table 5). During dynamics both TcTS and ZINC02576132 were very stable, producing RMSD values less than  $1.0$  Å (Fig. 7), suggesting that neither the protein nor the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

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

Unlike ZINC13359679, the interface between ZINC02576132 and TcTS appears to be dominated by polar interactions (Fig. 9).



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



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

The nonpolar interactions present are simple van der Waals contacts with no real possibilities for protein–ligand  $\pi$ -stacking since the ZINC02576132 phenyl group is located in the hydrophobic pocket near Val95 and Leu176, not in the acceptor binding site between Tyr119 and Trp312. The hydrogen bonds between the ligand carboxylate and the arginine triad do not have as high occupancy as ZINC13359679 (all more than 80% for ZINC13359679; none more than 54% occupied for ZINC02576132), but this is compensated for by increased hydrogen bonds with other protein residues. For example, Arg53 is ideally positioned to form two hydrogen bonds (66% and 44% occupied) with the carbonyl oxygen on ZINC02576132. In contrast, the substrate and DANA are only able to form a single hydrogen bond between a ligand hydroxyl group and Arg53. Likely more important than interactions with Arg53 is the  $-\text{NH}_2^+$  group of the ligand that forms a total of three hydrogen bonds with the Asp96/Glu230 motif, including a hydrogen bond with an occupancy over 90% with Asp96 (the most sustained TcTS–ZINC02576132 hydrogen bond). The holo pharmacophore specifically required a hydrogen donor be present at this location for any hit compounds with the intent of interacting with this motif, which is consistent with even poor inhibitors (*i.e.* DANA) of TcTS. The important role that Glu230 plays in acid/base catalysis combined with the close proximity of Asp96 makes this area of the active site a key target for drug design, and ideally suited to accommodate a positive ionizable group from an inhibitor (such as the  $-\text{NH}_2^+$  group of ZINC02576132). The strong binding free energy observed for ZINC02576132 provides further evidence that the Asp96/Glu230 motif could be a key to designing low nM inhibitors of TcTS. Although ZINC02576132 is stable in the active site and has a strong binding affinity for TcTS, key optimization could still make it a better inhibitor. Specifically, a clever addition of a phenyl group near the carboxylate could allow ZINC02576132 to also form  $\pi$ -stacking interactions with Tyr119 and Trp312 in the important acceptor binding site.



## 4. Conclusions

In this work, we wanted to create pharmacophore models detailing the important residues for ligand binding to TcTS and utilize that information to screen for novel inhibitors. To our knowledge, no pharmacophore models of TcTS have been created previously due to a lack of known active molecules. But the method recently proposed by Loving et al. [26] allows for the generation of pharmacophore models without *a priori* knowledge of actives by using the energetic results of docked fragments to the protein. We created three e-pharmacophore hypotheses using three different conformations of TcTS (PDB 1S0I, holo, and apo) each with 15 unique features. To our knowledge, these are the first pharmacophore models ever reported for TcTS and provide a detailed description of locations important for drug inhibition.

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

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

## Authors' contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## Acknowledgements

The authors wish to acknowledge the NIH for funding the work featured in this manuscript (Grant R01 AI073674). B.M. would also like to thank members of the Roitberg group for proofreading this manuscript.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmgm.2013.08.009>.

## References

- [1] C. Chagas, Nova tripanozomíase humana: estudos sobre a morfologia eo ciclo evolutivo do *Schizotrypanum cruzi* n. gen. n. sp., agente etiológico de nova entidade morbida do homem, Mem. Inst. Oswaldo Cruz 1 (1909) 159–218.
- [2] Control of Chagas Disease, World Health Organ Tech Rep Ser. 905, 2002, i–vi, 1–109, back cover.
- [3] D.W.D.W.T. Crompton, P. Peters, Working to Overcome the Global Impact of Neglected Tropical Diseases: First WHO Report on Neglected Tropical Diseases 2010, World Health Organization, 2010.
- [4] J.A. Castro, M.M. deMecca, L.C. Bartel, Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis), Hum. Exp. Toxicol. 25 (2006) 471–479.
- [5] D.A. Leiby, R.M. Herron, E.J. Read, B.A. Lenes, R.J. Stumpf, *Trypanosoma cruzi* in Los Angeles and Miami blood donors: impact of evolving donor demographics on seroprevalence and implications for transfusion transmission, Transfusion 42 (2002) 549–555.
- [6] M.J. Nowicki, C. Chinchilla, L. Corado, L. Matsuoaka, R. Selby, F. Steurer, et al., Prevalence of antibodies to *Trypanosoma cruzi* among solid organ donors in Southern California: a population at risk, Transplantation 81 (2006) 477–479.
- [7] H. Kun, A. Moore, L. Mascola, S. Frank, L. Gena, B. Kubak, et al., Transmission of *Trypanosoma cruzi* by heart transplantation, Clin. Infect. Dis. 48 (2009) 1534–1540.
- [8] C. Riera, A. Guarro, H.E. Kassab, J.M. Jorba, M. Castro, R. Angrill, et al., Congenital transmission of *Trypanosoma cruzi* in Europe (Spain): a case report, Am. J. Trop. Med. Hyg. 75 (2006) 1078–1081.
- [9] K.S. Pereira, F.L. Schmidt, A.M.A. Guaraldo, R.M.B. Franco, V.L. Dias, L.A.C. Passos, Chagas' disease as a foodborne illness, J. Food Prot. 72 (2009) 441–446.
- [10] K.S. Pereira, F.L. Schmidt, R.L. Barbosa, A.M.A. Guaraldo, R.M.B. Franco, V.L. Dias, et al., Chapter 3 – Transmission of Chagas disease (American trypanosomiasis) by food, in: S.L. Taylor (Ed.), Advances in Food and Nutrition Research, Academic Press, Burlington, Massachusetts, USA, 2010, pp. 63–85.
- [11] S. Schenkman, M.-S. Jiang, G.W. Hart, V. Nussenzweig, A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells, Cell 65 (1991) 1117–1125.
- [12] M.F. Amaya, A. Buschiazzi, T. Nguyen, P.M. Alzari, The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T. cruzi* trans-sialidase, J. Mol. Biol. 325 (2003) 773–784.
- [13] L. Freire-de-Lima, I.A. Oliveira, J.L. Neves, L.L. Penha, F. Alisson-Silva, W.B. Dias, et al., Sialic acid: a sweet swing between mammalian host and *Trypanosoma cruzi*, Front. Immunol. 3 (2012) 356.
- [14] F. Vandekerckhove, S. Schenkman, L.P. de Carvalho, S. Tomlinson, M. Kiso, M. Yoshida, et al., Substrate specificity of the *Trypanosoma cruzi* trans-sialidase, Glycobiology 2 (1992) 541–548.
- [15] G. Paris, L. Ratier, M.F. Amaya, T. Nguyen, P.M. Alzari, A.C.C. Frasch, A sialidase mutant displaying trans-sialidase activity, J. Mol. Biol. 345 (2005) 923–934.
- [16] J. Neres, P. Bonnet, P.N. Edwards, P.L. Kotian, A. Buschiazzi, P.M. Alzari, et al., Benzoic acid and pyridine derivatives as inhibitors of *Trypanosoma cruzi* trans-sialidase, Bioorg. Med. Chem. 15 (2007) 2106–2119.
- [17] J. Neres, M.L. Brewer, L. Ratier, H. Botti, A. Buschiazzi, P.N. Edwards, et al., Discovery of novel inhibitors of *Trypanosoma cruzi* trans-sialidase from in silico screening, Bioorg. Med. Chem. Lett. 19 (2009) 589–596.
- [18] J.H. Kim, H.W. Ryu, J.H. Shim, K.H. Park, S.G. Withers, Development of new and selective *Trypanosoma cruzi* trans-sialidase inhibitors from sulfonamide chalcones and their derivatives, ChemBioChem 10 (2009) 2475–2479.
- [19] S. Arioka, M. Sakagami, R. Uematsu, H. Yamaguchi, H. Togame, H. Take-moto, et al., Potent inhibitor scaffold against *Trypanosoma cruzi* trans-sialidase, Bioorg. Med. Chem. 18 (2010) 1633–1640.
- [20] X.-Y. Meng, H.-X. Zhang, M. Mezei, M. Cui, Molecular docking: a powerful approach for structure-based drug discovery, Curr. Comput. Aided Drug Des. 7 (2011) 146–157.
- [21] Y. Chen, B.K. Shoichet, Molecular docking and ligand specificity in fragment-based inhibitor discovery, Nat. Chem. Biol. 5 (2009) 358–364.
- [22] K. Babaoglu, B.K. Shoichet, Deconstructing fragment-based inhibitor discovery, Nat. Chem. Biol. 2 (2006) 720–723.
- [23] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, et al., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, J. Med. Chem. 47 (2004) 1739–1749.
- [24] G. Marcou, D. Rognan, Optimizing fragment and scaffold docking by use of molecular interaction fingerprints, J. Chem. Inf. Model. 47 (2007) 195–207.
- [25] G. Wolber, T. Seidel, F. Bendix, T. Langer, Molecule-pharmacophore superpositioning and pattern matching in computational drug design, Drug Discov. Today 13 (2008) 23–29.
- [26] K. Loving, N.K. Salam, W. Sherman, Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation, J. Comput. Aided Mol. Des. 23 (2009) 541–554.
- [27] O.F. Guner, History and evolution of the pharmacophore concept in computer-aided drug design, Curr. Top. Med. Chem. 2 (2002) 1321–1332.
- [28] I. Massova, P. Kollman, Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, Perspect. Drug Discov. 18 (2000) 113–135.
- [29] N. Okimoto, N. Futatsugi, H. Fuji, A. Suenaga, G. Morimoto, R. Yanai, et al., High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations, Biophys. J. 98 (2010) 460a.
- [30] O. Demir, A.E. Roitberg, Modulation of catalytic function by differential plasticity of the active site: case study of *Trypanosoma cruzi* trans-sialidase and *Trypanosoma rangeli* sialidase, Biochemistry 48 (2009) 3398–3406.
- [31] J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, R.G. Coleman, ZINC: a free tool to discover chemistry for biology, J. Chem. Inf. Model. 52 (2012) 1757–1768.
- [32] M.F. Amaya, A.G. Watts, I. Damager, A. Wehenkel, T. Nguyen, A. Buschiazzi, et al., Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase, Structure 12 (2004) 775–784.

- [33] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free, *J. Comput. Chem.* 26 (2005) 1701–1718.
- [34] G.M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, *J. Comput. Aided Mol. Des.* (n.d.) 1–14.
- [35] LigPrep v2.5, Schrodinger, Inc., Portland, OR, n.d.
- [36] R.A. Friesner, R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, et al., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes, *J. Med. Chem.* 49 (2006) 6177–6196.
- [37] N.K. Salam, R. Nuti, W. Sherman, Novel method for generating structure-based pharmacophores using energetic analysis, *J. Chem. Inf. Model.* 49 (2009) 2356–2368.
- [38] S.L. Dixon, A.M. Smondyrev, S.N. Rao, PHASE: a novel approach to pharmacophore modeling and 3D database searching, *Chem. Biol. Drug Des.* 67 (2006) 370–372.
- [39] A.W. Götz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born, *J. Chem. Theory Comput.* 8 (2012) 1542–1555.
- [40] D.A. Case, T.A. Darden, T.E.I. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, et al., AMBER 12, University of California, San Francisco, 2012.
- [41] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38.
- [42] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (2004) 1157–1174.
- [43] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins* 65 (2006) 712–725.
- [44] D.J. Price, C.L. Brooks, A modified TIP3P water potential for simulation with Ewald summation, *J. Chem. Phys.* 121 (2004) 10096–10103.
- [45] J.-P. Ryckaert, G. Ciccotti, H.J. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.* 23 (1977) 327–341.
- [46] T.A. Andrea, W.C. Swope, H.C. Andersen, The role of long ranged forces in determining the structure and properties of liquid water, *J. Chem. Phys.* 79 (1983) 4576–4584.
- [47] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, M.M.P.B.S.A. py., An efficient program for end-state free energy calculations, *J. Chem. Theory Comput.* 8 (2012) 3314–3321.
- [48] A. Onufriev, D. Bashford, D.A. Case, Exploring protein native states and large-scale conformational changes with a modified generalized born model, *Proteins* 55 (2004) 383–394.
- [49] L.Y. Zhang, E. Gallicchio, R.A. Friesner, R.M. Levy, Solvent models for protein–ligand binding: comparison of implicit solvent Poisson and surface generalized born models with explicit solvent simulations, *J. Comput. Chem.* 22 (2001) 591–607.
- [50] A. Buschiazzo, M.F. Amaya, M.L. Amaya, A.C. Frasch, P.M. Alzari, The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis, *Mol. Cell* 10 (2002) 757–768.
- [51] G. Pierdominici-Sottile, N.A. Horenstein, A.E. Roitberg, Free energy study of the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. From the Michaelis complex to the covalent intermediate, *Biochemistry* 50 (2011) 10150–10158.
- [52] G. Pierdominici-Sottile, A.E. Roitberg, Proton transfer facilitated by ligand binding. An energetic analysis of the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase, *Biochemistry* 50 (2011) 836–842.
- [53] G. Rastelli, A.D. Rio, G. Degliesposti, M. Sgobba, Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA, *J. Comput. Chem.* 31 (2010) 797–810.