JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2019, 59, 137−148

pubs.acs.org/jcim

Understanding Structure−Activity Relationships for Trypanosomal Cysteine Protease Inhibitors by Simulations and Free Energy Calculations

Lucianna H. Santos,[†,‡] Birgit J. Waldner,[§] Julian E. Fuchs,[§,⊥] Glaécia A. N. Pereira,[‡,∥,@] Klaus R. Liedl,[§] Ernesto R. Caffarena,[†] and Rafaela S. Ferreira*[,‡]

[†]Grupo de Biofísica Computacional e Modelagem Molecular, Programa de Computação Científica (PROCC), Fundação Oswaldo Cruz, Av. Brasil 4365, Rio de Janeiro, RJ 21040-360, Brazil

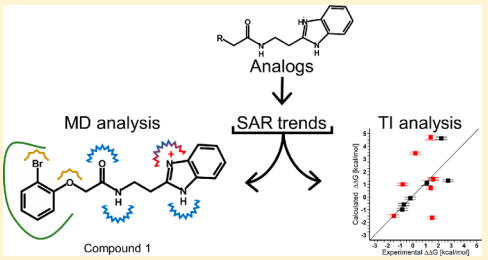
[‡]Laboratório de Modelagem Molecular e Planejamento de Fármacos, Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, Belo Horizonte, MG 31270-901, Brazil

[§]Institute of General, Inorganic and Theoretical Chemistry, and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innrain 82, Innsbruck, Tyrol 6020, Austria

[∥]CAPES Foundation, Ministry of Education of Brazil, Brasília, DF Brazil

Supporting Information

ABSTRACT: The protozoan cysteine proteases cruzain in *Trypanosoma cruzi* and rhodesain in *Trypanosoma brucei* are therapeutic targets for Chagas disease and Human African Trypanosomiasis (HAT), respectively. A benzimidazole series was previously characterized as potent noncovalent competitive cruzain and rhodesain inhibitors with activity against trypanosomes. Common structure−activity relationships (SAR) trends and structural modifications leading to selectivity against each enzyme were described. However, some of these trends could not be understood based on the reported binding mode of lead compound **1**. Therefore, we employed microsecond molecular dynamics simulations and free energy calculations to understand qualitative SAR trends and to quantitatively recapitulate them. Simulations revealed the most stable protein−ligand interactions and provided insights concerning enzyme selectivity. Calculated relative binding free energies of compound **1** analogs exhibited deviations of 1.1 and 2.2 kcal/mol from the experimental values for cruzain and rhodesain, respectively. These data encourage prospective thermodynamic integration (TI) studies to optimize this series and facilitate the prioritization of compounds for synthesis.



INTRODUCTION

Trypanosoma cruzi and *Trypanosoma brucei* are parasitic protozoa responsible for the life-threatening illnesses Chagas disease and Human African Trypanosomiasis (HAT), respectively. Chagas disease is globally prevalent with an estimation of 8 million people affected in 21 countries and only two drugs (benznidazole and nifurtimox) available for treatment.[1] HAT is exclusive endemic on the African continent with a prediction of 60 million people at risk of infection and four drugs (suramin, pentamidine, melarsoprol, and eflornithine), and one drug combination (nifurtimox/eflornithine) available.[1] Overall, current therapies for these diseases are limited due to cost, toxicity, narrow or undetermined efficacy, and the development of drug resistance.[1−3] Therefore, there is a need to develop new and improved drugs for trypanosomiases treatment.

Due to their essential role in parasitic survival and infection, cysteine proteases are relevant targets for antiparasitic chemotherapy.[4−9] Thus, medicinal chemistry studies have focused on the major cathepsin L-like cysteine proteases cruzain[10] of *T. cruzi* and rhodesain[11] of *T. brucei*. A variety of potent inhibitors of both enzymes have been reported, including vinyl sulfones,[12−14] nitriles,[15,16] thiosemicarbazones,[17−20] ketones,[6,21,22] benzimidazoles,[23−25] and other varied scaffolds.[23,26−31] Since cruzain and rhodesain share 70% sequence identity between them, and have very similar binding sites, classes of inhibitors active against both proteases have also been described.[12,15,17,26]

In previous studies, a novel class of noncovalent benzimidazole inhibitors was shown to inhibit cruzain[23,25] and rhodesain[24] at nanomolar concentrations and have trypanocidal activity. These studies provided a structure−activity relationship (SAR) for this class against both enzymes, highlighting common features and modifications that lead to inhibitor selectivity to each enzyme. For cruzain, QSAR studies were also reported and highlighted functional groups which most contribute to potency.[32] The binding mode of the lead

Received: August 20, 2018
Published: December 11, 2018

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

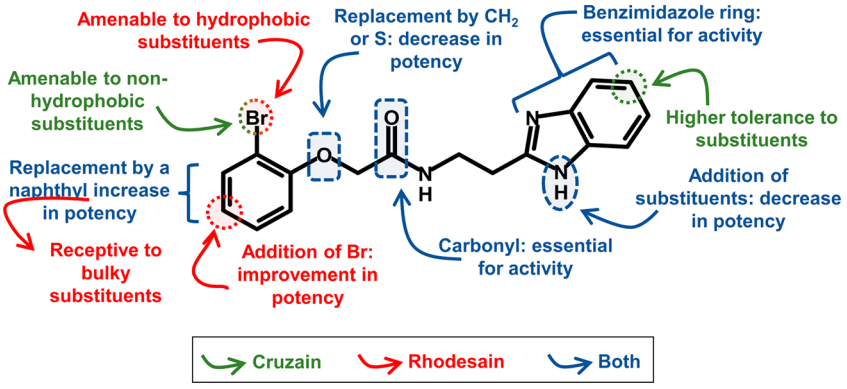


Figure 1. Summary of SAR trends for the benzimidazole series against cruzain and rhodesain based on compound **1** structure.

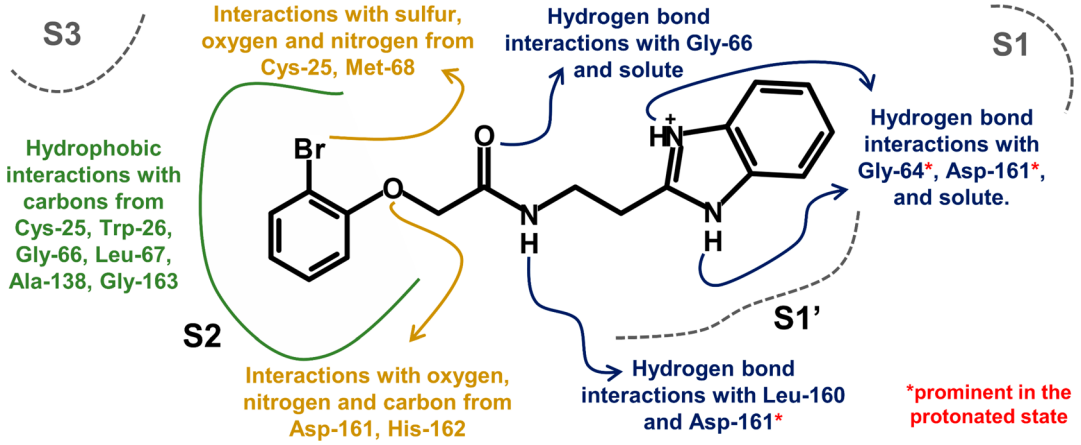


Figure 2. Frequent interactions of compound **1** in the active site of rhodesain and cruzain according to MD analysis.

compound **1** (Figure 1) to cruzain has been solved by X-ray crystallography,[23] providing a basis to rationalize the SAR obtained. However, several experimentally observed SAR trends could not be explained just based on this information. For example, the benzimidazole ring was found to be essential for enzyme inhibition, despite being mostly solvent-exposed. Therefore, we attempt to comprehend the molecular basis of some of the SAR trends (Figure 1) and observed inhibitor selectivity between these enzymes and the benzimidazole series using computational methods.

Computational methods can help guide different stages in drug development. Techniques such as molecular docking and molecular dynamics allow one to describe the intra- and intermolecular interactions most relevant for potency and provide insights toward optimization of a scaffold. Free energy calculations have been increasingly used to provide more quantitative predictions and to prioritize compounds for synthesis, based on their estimated potency.[33,34] Here, we first examined the dynamic behavior of lead compound **1** by Molecular Dynamics (MD), while its energy behavior was examined by the Molecular Mechanics Poisson−Boltzmann Surface Area (MMPB/SA) method. With this strategy, we could describe compound **1** interactions and behavior beyond the crystallographic binding mode. Also, to better understand the thermodynamics underlying the binding process between these inhibitors and both enzymes, Thermodynamic Integration (TI)[35] was employed. TI calculations were performed for compound **1** analogs that embodied different structural modifications that produced approximately 10-fold changes in potency. The computational alchemical free energies calculations were in good agreement with experimentally determined IC_{50} values.

RESULTS

Binding Stability and Molecular Interactions of Compound 1 from MD Simulations. Compound **1** was the lead compound in this series and the only one for which the binding mode has been experimentally determined. Therefore, we chose this molecule as a prototype for our MD simulations. Computer simulations were performed in different complex configurations for the enzymes. For cruzain, the available high-resolution structure of this complex[23] was employed as a starting configuration. Whereas, in the absence of a crystal structure of compound **1** bound to rhodesain, we initially performed simulations with the equivalent binding mode. However, around 250 ns of the simulation the protonated state of **1** became unstable and detached from the active site, unlike its correspondent state in cruzain (see Figure S1). Given the instability of this pose, next, we performed molecular dynamics (MD) of this complex starting from a putative pose of the ligand in the binding site (see the Supporting Information). Interestingly, the RMSD between the initial putative position of the ligand in rhodesain and compound **1** in the crystallographic configuration was 3.9 Å. This high RMSD was due to the different positioning of the benzimidazole ring.

Taking into account the presence of an ionizable nitrogen in the benzimidazole ring, we investigated the protonation state

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

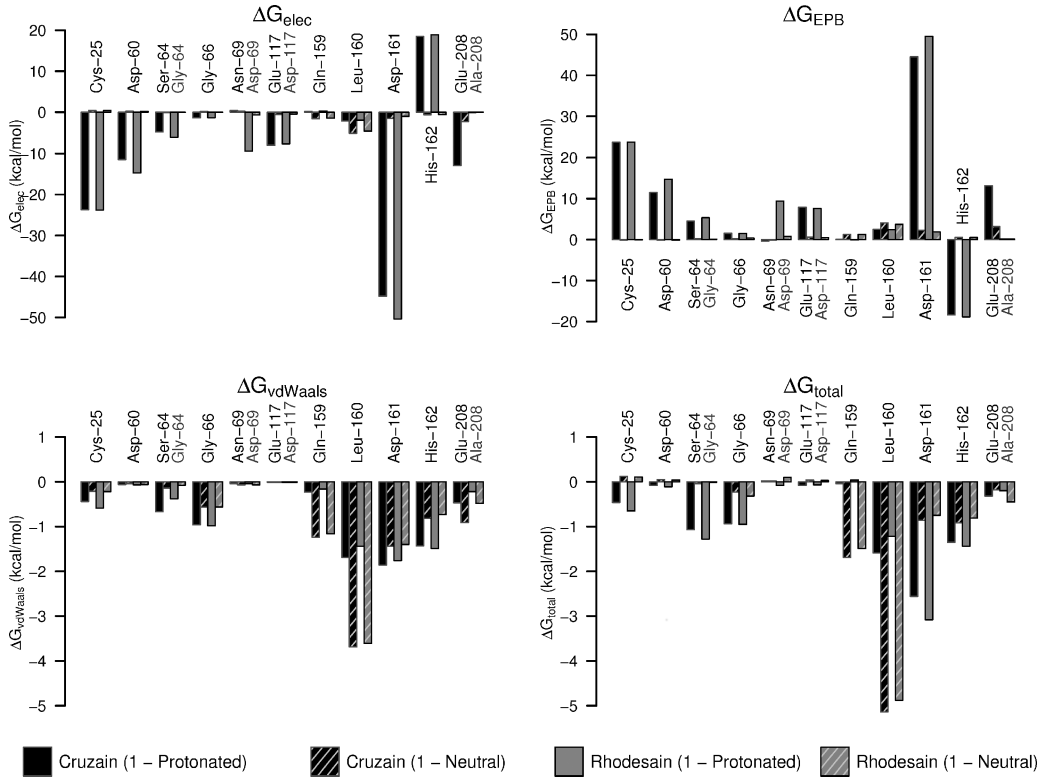


Figure 3. Per-residue energy decomposition of the absolute free energy from MMPB/SA. Residue names in black correspond to cruzain, while the ones in gray correspond to rhodesain positions.

of the lead compound **1**. Programs Marvin[36] and Epik[37] indicated the coexistence of two possible ligand states (protonated and neutral) at the assay pH value of 5.5, and both were included in the simulations (see the Supporting Information). In general, the protonated compound **1** produced configurations closer to its crystallographic position than the neutral one, and it was more stable in comparison to the starting configuration, suggesting better stability of this state in complex with the enzymes (Figures S1, S2, and S3).

Significant changes in the hydrogen bond profile were observed for the different protonation states of **1**. When the ligand was protonated, its linker region was more stably bound to the enzymes, through hydrogen bonds with Gly-66 and Asp-161 backbone atoms, as commonly observed in crystallographic complexes with cruzain[13,21−23,38,39] and rhodesain.[11,13,40] The extra hydrogen atom in the benzimidazole ring was observed to hydrogen bond to the side chains of Asp-161 and Ser-64 (Gly-64 in rhodesain), with occupancy between 45% and 80% of the simulation time. On the other hand, in simulations with **1** neutral, the benzimidazole nitrogen hydrogen-bonded mostly to water molecules (Table S1).

To gain further insight into molecular interactions between **1** and the cysteine proteases, we analyzed which close intermolecular contacts (up to 5 Å) were stable throughout the simulation. In the simulations with **1** protonated, all atoms in the bromophenyl ring and the linker stably interacted (occupancies frequently over 90%) with several protein atoms, especially in the S2 subsite, while only one carbon from the benzimidazole interacted with any protein atom with frequency over 50% (Table S2). The interactions observed in the MD simulations are summarized in Figure 2.

Comparison of the interactions of the benzimidazole ring in the two proteases indicates higher solvent exposure in simulations against cruzain, while when bound to rhodesain this ring seems to be more buried in the S1′ pocket, as indicated by more frequent hydrogen bonds and by contacts between benzimidazole carbons and the side chain of Asp-161 (Tables S1 and S2). This observation provides a possible explanation for cruzain's higher tolerance to substituents in the benzimidazole ring, a notable SAR trend we experimentally observed in a previous study.[24]

Energy Behavior of 1 from MMPB/SA. To further evaluate differences between protonation states of **1** and determine which residues most contribute to binding, we employed MMPB/SA.

General energy profiles were similar between both enzymes (Table S3). When the ligand was in the protonated form, the electrostatic energy was negative for both cruzain (−291.6 ± 25.6 kcal/mol) and rhodesain (−383.2 ± 18.9 kcal/mol). Likewise, the solvation free energy was positive for cruzain (296.4 ± 21.9 kcal/mol) and rhodesain (385.6 ± 16.0 kcal/mol). These contributions almost neutralized each other. Similar energetic profiles were also observed for both enzymes when the neutral state of **1** was considered. As expected, in this case, both the electrostatic contribution (−18.8 ± 5.7 kcal/mol and −14.9 ± 5.1 kcal/mol for cruzain and rhodesain, respectively) and desolvation penalties were much lower than observed for the protonated ligand. The van der Waals component was the major contributor to the interaction energies (−31.4 ± 2.7 kcal/mol and −30.6 ± 2.5 kcal/mol for cruzain and rhodesain, respectively).

Per-residue energy decomposition (Figure 3) showed that the contribution from residues to the total energy changed according to the protonation state of the ligand. As expected, this was observed primarily for charged residues. When the ligand was in the protonated form, residues such as Cys-25,

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

Asp-60, and Asp-161 obtained favorable electrostatic energy (−23.73, −11.50, and −44.80 kcal/mol for cruzain and −23.78, −14.75, and −50.36 kcal/mol for rhodesain) over its neutral form (0.44, 0.24, and −1.52 kcal/mol for cruzain and 0.40, 0.20, and −1.05 kcal/mol for rhodesain). However, residue electrostatic contributions were almost neutralized by desolvation penalties (23.75, −11.48, and 44.52 kcal/mol for cruzain and 23.75, 14.70, and 49.51 kcal/mol for rhodesain), leaving the van der Waals component as a critical contributor to the total energy. This component also was the major contributor to binding for the neutral state of **1**, with residues Gln-159, Leu-160, and Asp-161 providing favorable energy contributions (−1.24, −3.69, and −1.44 kcal/mol for cruzain and −1.16, −3.61, and −1.40 kcal/mol for rhodesain). Furthermore, the per-residue decomposition showed that the residue differences between rhodesain (Gly-64, Asp-69, Asp-117, and Ala-208) and cruzain (Ser-64, Asn-69, Glu-117, and Glu-208) did not grant significant differences in their contributions to binding.

Overall, the calculated absolute binding free energies, considering entropic contribution correction from normal modes analysis, favor the protonated state of ligand **1** (−8.7 ± 5.4 kcal/mol and −7.9 ± 6.8 kcal/mol for rhodesain and cruzain, respectively) against the neutral state (−6.1 ± 3.8 kcal/mol and −4.4 ± 4.5 kcal/mol for rhodesain and cruzain, respectively) for binding to both enzymes (Table S3).

Relative Free Energy of Binding from TI Simulations. Next, we turned into TI alchemical transformations, to investigate the modifications that significantly impacted potency against the enzymes. TI calculations were performed with the protonated ligand, based both on its higher stability during the simulations and on more favorable calculated absolute free energies by MMPB/SA. Aiming at quantitatively recapitulating SAR trends, we organized a workflow for both enzymes to calculate the relative binding free energy for compounds with substituents on the phenyl ring (compounds **1**, **2**, **5**, **7**, **8**, **15**, and **16**) or with a modification of the linker between the phenyl and benzimidazole rings (**24**) (Figure 4). Depending on the system, the transformations demonstrated either a higher than 10-fold change (**1** ↔ **2** and **1** ↔ **5** in cruzain; **5** ↔ **7** and **5** ↔ **8** in rhodesain; **15** ↔ **16** and **1** ↔ **24** in both) or little change (up to 4-fold changes for **5** ↔ **7** and **5** ↔ **8** in cruzain; **1** ↔ **2** and **1** ↔ **5** in rhodesain; **1** ↔ **15** in both) in potency between compounds.

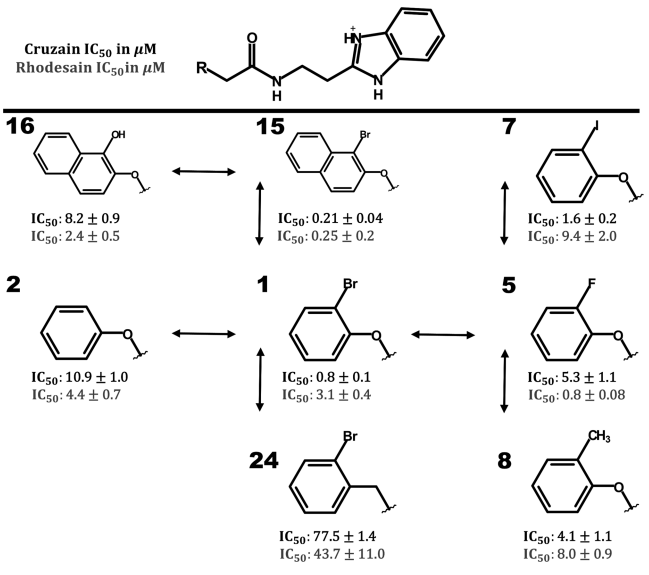


Figure 4. TI transformations workflow.

For cruzain, good results were obtained as the predicted values were within 0.2 kcal/mol of their experimental counterparts, except for two systems (**15** ↔ **16** and **1**↔ **24**) (Table 1). Lower accuracy was achieved for the rhodesain calculations; only three out of seven transformations were within 1.0 kcal/mol of their experimental values (**5** ↔ **8**, **1** ↔ **15**, and **1** ↔ **24**) (Table 1). Overall, the RMSE (eq 5) observed was ∼1.1 kcal/mol for cruzain and ∼2.2 kcal/mol for rhodesain (Figure 5).

Three alchemical transformations (**1** ↔ **2**, **1** ↔ **5**, and **5** ↔ **7**) were more accurate with cruzain than rhodesain. For example, with rhodesain, the replacement of fluorine (**5**) by iodine (**7**) at the phenyl ortho position resulted in a 12-fold higher IC_{50} value, indicating an unfavorable change ($\Delta\Delta G_{experimental}$ = 1.5 kcal/mol). Surprisingly, the calculated free energy change was −1.6 kcal/mol, suggesting a favorable transformation. On the other hand, the impact of this substitution in potency against cruzain was low and precisely captured by the calculations.

For both enzymes, two alchemical transformations (**5** ↔ **8** and **1** ↔ **15**) were highly accurate. Replacing the fluorine (**5**) by methyl (**8**) resulted in a low absolute error for cruzain (0.1 kcal/mol) and rhodesain (0.7 kcal/mol). Replacement of the phenyl with a naphthyl (**1** ↔ **15**) followed the same trend with favorable 1.5 and 1.0 kcal/mol transformations for rhodesain and cruzain, respectively, both predicted with errors within 0.2 kcal/mol.

Notwithstanding, the overall good agreement between the TI and experimental results, the free energies of the **15** ↔ **16** transformation, corresponding to the replacement of bromine with a hydroxyl in the naphthyl-substituted compound, were poorly predicted for both enzymes (absolute error of over 2.0 kcal/mol). Such large TI errors might be due to the polar nature of the hydroxyl group, as the transformation step of recharging the hydroxyl in water produced a value almost three units over the same transformation in the complex system (Table S4 and S5). Also, for rhodesain, the calculation experienced an increase in the vdW-related term (second step of TI that estimates the contribution of the solvation free energy for the transformation) of the conversion for the complex (4.5 kcal/mol) compared to the ligand in water transformation (2.4 kcal/mol) (Table S4).

To investigate whether the more significant errors obtained were due to poor sampling, we performed new calculations adding extra λ-windows for systems where the difference between calculated and experimental was between 2.0 and 4.0 kcal/mol. However, there was no improvement when more points were added to the transformation curve (data not shown). Therefore, either the issue with these systems goes beyond sampling problems or a much more extensive sampling might be needed to allow observation of phenomena such as interconversion between relevant binding modes.

Coherent with this possibility, a slow interconversion between binding modes was observed when cluster analysis was performed in the 1000 ns MD simulations of the protonated state of compound **1** complexed to both enzymes. Comparison between the most populous cluster representatives of cruzain and rhodesain showed an overlap (Figure 6a and Tables S6 and S7 and Figures S4, S5, and S6). However,

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

Table 1. Relative Free Energy of Binding of Compounds Calculated by TI and Compared to the Experimental Relative Binding Free Energy

	rhodesain			cruzain		
transformation	calculated ΔΔ*G* (kcal/mol)	experimental ΔΔ*G* (kcal/mol)	|Δ*x*|[a]	calculated ΔΔ*G* (kcal/mol)	experimental ΔΔ*G* (kcal/mol)	|Δ*x*|[a]
1 ↔ 2	3.4 ± 0.1	0.2 ± 0.3	3.2	1.4 ± 0.1	1.6 ± 0.2	0.2
1 ↔ 5	1.0 ± 0.1	−0.8 ± 0.3	1.8	1.1 ± 0.2	1.1 ± 0.3	0.0
5 ↔ 7	−1.6 ± 0.1	1.5 ± 0.4	3.1	−0.6 ± 0.1	−0.7 ± 0.3	0.1
5 ↔ 8	0.7 ± 0.1	1.4 ± 0.3	0.7	−0.1 ± 0.1	−0.2 ± 0.5	0.1
1 ↔ 15	−1.5 ± 0.1	−1.5 ± 0.9	0.0	−1.0 ± 0.1	−0.8 ± 0.3	0.2
15 ↔ 16	4.7 ± 0.2	1.4 ± 1.0	3.3	4.6 ± 0.1	2.2 ± 0.3	2.4
1 ↔ 24	1.4 ± 0.1	1.6 ± 0.4	0.2	1.3 ± 0.1	2.7 ± 0.2	1.4

[a]Absolute error between the calculated and experimental values.

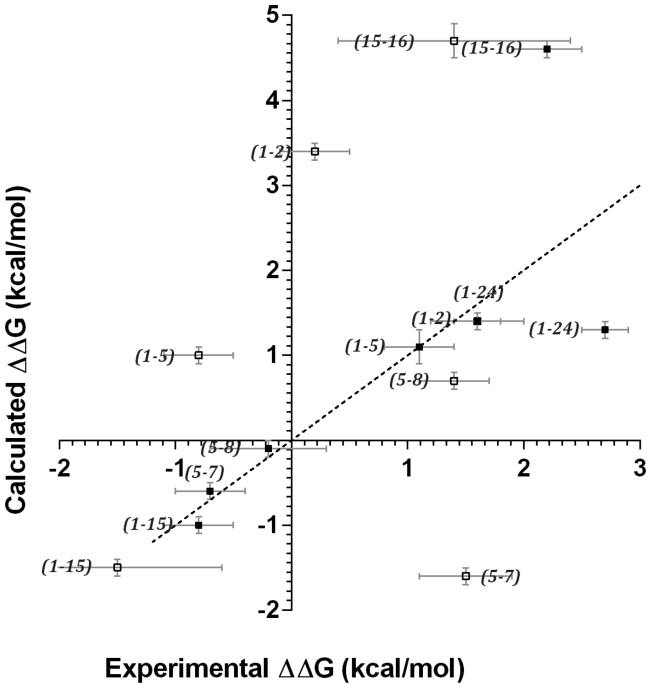


Figure 5. Comparison between the calculated relative binding free energy and experimental relative binding free energy. Cruzain and rhodesain outcomes are indicated by black and white squares, respectively. Horizontal and vertical lines display the errors in the experimental and calculated relative binding free energies, respectively.

these cluster representatives significantly differ to the crystallographic conformation of **1**, due to a different place in the benzimidazole ring (RMSD over 3.0 Å). Conformations that closely resembled the crystallographic state (RMSD less than 2.0 Å) were also sampled in the second and third cluster in cruzain and rhodesain, respectively (Figure 6b); however, their sample frequency was higher for cruzain (17% vs 1% of the simulation, Table S6). Such a difference in cluster populations is in agreement with our analysis of intermolecular interactions, which indicated higher solvent exposure of the benzimidazole ring in cruzain simulations.

In addition to the quantitative comparison between observed and predicted affinities, TI simulations also allowed analysis of interactions which were frequently observed in this compound series (Table S8). Similar to the observed for MD simulations, hydrogen bond interactions between the benzimidazole ring NH and side chains of Asp-161 and Ser/Gly-64 were noticed throughout all TI simulations. The interaction between Gly-66 and the linker carbonyl oxygen was also observed in all transformations, with higher frequencies for the rhodesain simulations. Interestingly, alchemical changes with large TI errors displayed an increase of solvent interactions with the linker carbonyl oxygen, exposed to solvent due to a more extended conformation of the ligands.

We also investigated the molecular bases for some observed SAR and selectivity trends, by analysis of the most populated clusters and close contacts between each compound and the two enzymes throughout TI transformations (Table S9). Experimentally, the transformation **1**↔**24**, in which the linker ester oxygen is replaced by a carbon atom, reduces potency against cruzain by 100-fold and against rhodesain by only 12-fold. Comparison of the most populated clusters throughout this TI transformation reveals differences in how the bromophenyl ring is accommodated within the S2 pocket of the two enzymes. Interestingly, the major clusters in cruzain represent 73−81% of the simulation, while for rhodesain the two most populated clusters represent approximately 40% of the simulation each (Figure S7). Thus, it seems that rhodesain has a higher tolerance for alterations in the position of the bromophenyl ring. It is also observed that the linker ester oxygen is in different positions when bound to the enzymes. Such a difference might, therefore, be related to its higher stabilization when binding to cruzain. In the crystallography complex cruzain-compound **1** (PDB 3KKU), the linker oxygen is in the vicinity of backbone nitrogens from His162 and Gly163 (distances of 3.2 and 3.6 Å, respectively). Similar distances are observed in the simulation of compound **1** bound to cruzain. On the other hand, these distances are bigger in simulations of compounds **1** and **24** bound to rhodesain (Table S10). Consequently, it is possible that in this system the different positioning of the bromophenyl ring decreases the importance of electrostatic interactions involving the linker oxygen from compound **1**, reducing the impact of replacing this atom by a carbon.

Another peculiar SAR trend is observed for transformation **1** ↔ **5**, in which the replacement of bromine by fluorine causes a loss of potency in cruzain and an increase in rhodesain. Here, analysis of the most populated clusters for compounds involved in the transformation reveals only a small difference in the position of the halogenated ring (Figure S8). However, analysis of the close contacts (<5 Å) between the halogens and each enzyme reveals that this difference is enough to change the chemical environment of the halogens when bound to each enzyme (Table S11). We hypothesize that the difference in SAR may be therefore a consequence of the different position of the phenyl ring when bound to the enzymes.

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

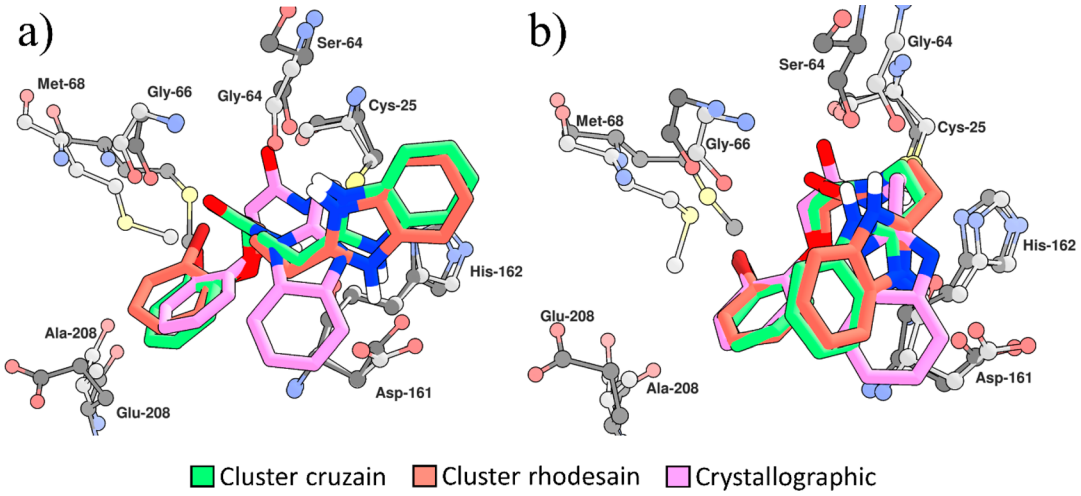


Figure 6. Outcomes from the cluster analysis of the protonated state of compound **1** in cruzain and rhodesain indicate sampling of a similar conformation space. (a) First cluster (cluster 0) representatives of both simulations overlap. However, the distance between these conformations to the crystallographic position of compound **1** is evident. (b) The second (cluster 1) and third (cluster 2) cluster representative structures of cruzain and rhodesain, respectively, showed very close conformation to the crystallographic one.

DISCUSSION

Previous studies revealed that the SAR profile and mechanism of inhibition of lead compound **1** and its analogs were in general very similar against rhodesain and cruzain. However, these studies also provided intriguing results, including SAR trends which could not be understood either based on the reported binding mode of lead compound **1** or considering docking results. Thus, we applied MD simulations to gather further structural information and better understand SAR trends. Qualitative trends were interpreted based on the analysis of stable protein−ligand interactions, while TI retrospectively recapitulated quantitative patterns.

The systems studied here are challenging cases for simulations, with ligands that contain an ambiguous protonation state. The nitrogen at position 3 in the benzimidazole ring has a predicted pK_a very close to the assay pH, and it is predicted to bind near ionizable residues, which could further cause a pK_a shift. Therefore, MD simulations included both possible protonation states of the ligand and residues Cys-25 and His-162 represented as an ionic pair.[41] On the basis of the analysis of ligand stability, absolute binding free energy predicted by MMPB/SA, and hydrogen bonding profile, the protonated state of ligand **1** displayed more favorable energy behavior for both enzymes. These results are in agreement with our expectations because the protonated ligand would facilitate electrostatic interactions due to its additional hydrogen and positive charge in an overall negatively charged binding site (Figure S9). More specifically, through the simulations, we observed that the additional hydrogen interacted with Ser-64 (cruzain), Gly-64 (rhodesain), and Asp-161 (both enzymes), while in its absence the benzimidazole ring mostly interacts with solvent molecules. Hence, our results suggest that compound **1** and its analogs are probably protonated when bound to both enzymes.

In simulations against both enzymes, the benzimidazole ring was the ligand region with higher flexibility, whereas the bromophenyl ring and the linker appear to anchor the compound in the binding site. The linker amide also seems to play an important role, through stable hydrogen bond interactions to the backbone of Gly-66 and Asp-161. Accordingly, these interactions were observed in the cruzain-compound **1** crystallographic complex,[23] and the carbonyl amide was shown to be essential for inhibition of both enzymes.[24,25] In addition, both the analysis of close contacts and MMPB/SA results indicate the importance of van der Waals contacts with the S2 pocket. A similar result was recently reported for another class of cruzain inhibitors, in which a chloroquinoline ring was found to stably interact within the S2 pocket while the substituent in this ring showed high flexibility.[42] These results are in agreement with the known importance of the S2 pocket for the potency and specificity of cysteine proteases.[43]

Despite the high sequence identity between these trypanosomal proteases, cruzain contains a Glu-208 in the bottom of the S2 pocket,[41] while in rhodesain this residue is replaced by Ala-208. This change is essential for the ability of cruzain to recognize both hydrophobic and positively change moieties in P2,[44] and it also results in a considerably more open S2 subsite in rhodesain. Consequently, we observed a higher oscillation of the bromophenyl ring in compound **1** in rhodesain simulations. This single mutation at residue 208 is also the likely explanation for one of the differences experimentally observed in the SAR between the enzymes: rhodesain seems more tolerant of modifications in naphthyl analogs.

Our simulations also provide a rationale for the importance of the benzimidazole ring in this compound series. This was a surprising result from the SAR against both enzymes, as even modification to a benzoxazole abolishes enzyme inhibition,[24,25] while the crystallographic structure indicates that this ring is very solvent-exposed. The crystal structure was determined at high resolution, with unambiguous electron density for the ligand.[23] However, crystals were obtained at pH 8.5, in which the ligand is expected to be neutral. On the other hand, at pH 5.5, simulations with both protonation states of **1** suggest the highest stability of the protonated compound, and interconversion between the crystallographic binding mode an alternative binding mode in which the benzimidazole interacts with the S1' pocket. The alternative mode resembles docking predictions described in Ferreira et al.[23] when **1** was reported and obtained by us with several docking programs (unpub-

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

lished results). In this pose, the additional hydrogen is involved in hydrogen bonds with residues from the S1′ pocket, which could not be performed by the inactive benzoxazole analog.

It has also been described that the addition of substituents in the benzimidazole ring had a lower impact on cruzain inhibition. Such a trend was observed for addition of substituents on the nitrogen or in the phenyl ring of the benzimidazole, both for electron-donating and electron-withdrawing groups.[24] In accordance with our MD results, this might be due to differences in cluster populations observed against these enzymes, as the benzimidazole ring was less solvent-exposed in rhodesain simulations, in which it occupied the S1′ subsite more frequently than in cruzain simulations. Therefore, this SAR trend seems to be mainly due to steric effects.

Having analyzed qualitative SAR trends, we moved to TI alchemical transformations. A good practice in alchemical free energy calculations is to exploit similarity between states by removing or decoupling as few atoms as possible.[45] Therefore, we created a workflow which consisted of mostly straightforward transformations starting from the lead compound **1** which resulted in significant changes in potency. In total, seven transformations systems were performed in both cruzain and rhodesain, encompassing the diversity of the analogs inhibitory potency (Figure 4).

For cruzain, estimated differences in relative binding free energy were in reasonable agreement with the experimental IC_{50} data. Even in the two cases where these values differ by over 1.0 kcal/mol, it is worth noting the correct prediction on the qualitative impact of the modifications (i.e., whether they would improve or reduce potency). Therefore, the TI outcomes for cruzain agreed with other retrospective studies in which the error of the calculated free energy of binding was around 1−2 kcal/mol relative to the experimental data.[46−49] These results are especially encouraging when we consider that the determination of ligand protonation state was not trivial and most ligands contained halogen atoms for which the interactions are still challenging for force fields.[50]

Despite similar trends that were observed in the initial simulations of **1** bound to cruzain and rhodesain, and the high residue identity between both enzymes, the precision of the TI calculations was considerably lower for rhodesain. A possible explanation for this difference in performance is the absence of a crystallographic complex. While the available high-resolution structure of compound **1** in complex to cruzain[23] was employed as a starting configuration for MD simulations, rhodesain simulations were performed from a putative pose of ligand **1** in the binding site.

Notably, in this starting configuration for rhodesain simulations, we observed a different positioning of the benzimidazole ring than found in cruzain (RMSD of 3.9 Å from the crystallographic configuration). While a cluster corresponding to the crystallographic binding mode of **1** in cruzain has also been observed in rhodesain simulations, its population represented only 1% of frames, whereas the cluster considered as the started configuration was observed 85% of the time. These results suggest a possible difference in relative contributions of multiple relevant binding modes for cruzain and rhodesain. This limitation would explain why imperfect sampling might not have allowed adequate interconversion between binding modes and resulted in more substantial errors for some transformations.[51,52]

CONCLUSION

Herein we performed MD simulations to better understand SAR and selectivity trends between a series of benzimidazoles and the trypanosomal proteases cruzain and rhodesain.

Despite the challenges faced for TI simulations with these systems, such as uncertainty on ligand protonation states, modeling of halogen interactions by force fields, and the absence of a starting crystallographic binding mode for rhodesain, the accuracy of the TI calculations is within the range currently reported in the literature. Results are better for cruzain, with most absolute errors ≤0.2 kcal/mol.

On the basis of the good overall agreement between the calculated and experimental binding free energies, the current data provide a basis for employing similar calculations in prospective studies to guide potency optimization, in an effort to generate leads for Chagas disease and HAT.

EXPERIMENTAL SECTION

Preparation of Cysteine Protease X-ray Structures for Computational Assays. Crystal structures of cruzain and rhodesain were retrieved from the Protein Data Bank,[53] being chosen based on their high resolution. In the case of cruzain, the structure (PDB code 3KKU[23]) contains the noncovalent ligand **1**, the lead compound in the series herein studied, bound to the enzyme active site. The chosen rhodesain structure (PDB code 2P86[40]) contains the irreversible pseudo peptide inhibitor K11002 bound to the active site. The reported resolution for these X-ray diffraction derived structures 1.28 and 1.16 Å, respectively.

As there is no crystal structure of compound **1** bound to rhodesain, we removed K11002 inhibitor from the binding site of PDB 2P86 and created an apo system structure before MD simulations were performed. A binary complex formed by rhodesain and compound **1** was created using molecular docking (see details in the Molecular Docking section) within the receptor active site, in which the ten most representative protein conformations obtained from a previous simulation of the apo system were considered.

Computational preparation of the proteins consisted of adding hydrogen atoms to the residues according to acidic conditions (the assay pH value of 5.5) using the Protonate3D protocol from MOE.[54] All aspartate and glutamate residues were in their deprotonated forms, except Asp-57 which was protonated due to its proximity to Asp-60. All arginine and lysine residues were in the protonated, positively charged, state. In the active site, the catalytic dyad was considered as a thiolate-imidazolium ion pair, His-NH^+/Cys-S^- (His-162-NH^+/Cys-25-S^-, cruzain numbering), commonly believed to exist in cysteine proteases.[55,56]

Determination of the protonation state of compound **1** at pH 5.5 was performed with the programs ChemAxon's MarvinSketch[36] and Schrödinger's Epik.[37] Ligands were parametrized using the Generalized Amber Force Field (GAFF)[57] and AM1-BCC[58] atomic partial charges were assigned using the Antechamber[59] program. All topologies and coordinate files for the simulations were created through the LEaP module. Both LEaP and Antechamber programs are part of the AmberTools version 14.0.[60]

Binding Site Definition. We considered as the active site of cruzain all the residues within 10 Å of compound 1 center of mass: Gln-19, Gly-23, Cys-25, Trp-26, Ala-27, Ser-29, Asp-60, Ser-61, Gly-62, Cys-63, Ser-64, Gly-65, Gly-66, Leu-67, Met-

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

68, Asn-69, Glu-117, Val-137, Ala-138, Val-139, Asp-140, Ala-141, Ser-142, Gln-159, Leu-160, Asp-161, His-162, Gly-163, Val-164, Trp-184, Thr-186, Leu-204, and Glu-208. Considering the high similarity between cruzain and rhodesain, the active site for the latter was defined as comprised of those residues that were structurally aligned to cruzain's active site. Most positions contained the same residues, although differences were found in the following positions for rhodesain: Phe-61, Gly-64, Asp-69, Asp-117, Ile-137, Thr-142, Asn-186, and Ala-208.

Molecular Docking. All docking calculations were carried out with the Glide program, version 6.6[61−63] employing the XP algorithm and using the Virtual Screening Workflow that allows ensemble docking of a ligand library against multiple rigid receptors. To determine an initial binding mode of compound **1**, ten representative structures were extracted from the apo-rhodesain molecular dynamics (MD) simulation using hierarchical clustering. The two best scoring poses were selected and submitted to additional analysis using a short ten nanoseconds (ns) molecular dynamics simulation. Outcomes were compared with cruzain MD simulations of compound **1** (see the Supporting Information).

Molecular Dynamics Simulations. MD simulations were carried out using the GPU implementation of Particle Mesh Ewald Molecular Dynamics (PMEMD) in AMBER 14.0.[60] Molecular interactions were computed using the Amber 99SB-ILDN force field[64] with the TIP3P[65] model for water molecules. Electrostatic interactions were computed using the Particle Mesh Ewald (PME)[66] method. A cutoff of 8 Å for nonbonded interactions was employed. Bonds involving hydrogen atoms were constrained with SHAKE for nonwater molecules. A truncated periodic octahedral box was used as a simulation box with a minimum distance of 12 Å between any box edge and any solute atom.

Minimization occurred with harmonic constraints on protein heavy atoms. Heating happened in the NVT ensemble from 0 to 300 K using the Langevin thermostat,[67] with a collision frequency set to 2 ps^{-1}. Lastly, an unrestrained equilibration over 1 ns in the NPT ensemble with a target pressure of 1 bar and 2 ps coupling time was applied. The details of the extensive equilibration protocol applied here can be found in Wallnoefer et al.[68] The final coordinates of the minimization and equilibration protocol were then used to complete 1000 ns at 300 K maintained by the Langevin thermostat[67] using a time step of 2 fs. Atomic coordinates and velocities were recorded every 20 ps.

Trajectories analysis were performed by cpptraj,[69] also part of the AmberTools version 14.0.[60] Hydrogen bond analysis was performed with cpptraj version 17.00 (command hbond paired with nointramol) allowing one to track only intermolecular interactions between the ligand and the enzymes. This version of cpptraj was also used to track close contacts (distance <5.0 Å) formed by the ligand (command nativecontacts). VMD[70] was used to visualize the trajectories. Root-mean-square deviations (RMSD) between heavy atoms of the ligand, extracted from simulation frames, were submitted to the Hungarian symmetry-correct heavy atom algorithm[71] using DOCK6[72] and plotted with program R.[73] To group similar conformations of the protonated state of compound 1 during the simulations, the trajectories were clustered by performing clustering analysis using the DBScan clustering algorithm[74] available through AmberTools with a cutoff value of 1.2 Å.

MMPB/SA Calculations. To further characterize the interaction of compound **1** with both enzymes, we performed MMPB/SA calculations to qualitatively determine the most significant contributors to the free energies of binding through MD simulations.[75] The Molecular Mechanics Poisson−Boltzmann Surface Area (MMPB/SA) are postprocessing end-state approaches to compute free energies of molecules in solution, characterized by the use of Poisson−Boltzmann (PB)[76] methods. In this approach, the average interaction energies between receptor and ligand are usually obtained by performing calculations on an ensemble of uncorrelated snapshots collected from an equilibrated MD or Monte Carlo simulation. The interaction energy and solvation free energy for the complex, receptor, ligand and their resulting averages were calculated using the MMPBSA.py script[75] available through the AMBER distribution.

The polar contribution to the solvation free energy was computed by solving the linearized PB equation using Parse radii and a solvent probe radius of 1.4 Å. The dielectric constant was set to 1.0 for the interior of solutes (interior of protein) and 80.0 for the solvent. The nonpolar solvation contribution was determined using a solvent-accessible surface area (SASA) term where the surface tension proportionality constant was set to 0.00542 kcal/(mol $Å^{-2}$) and the offset value was set to 0.92 kcal/mol.[75] The snapshots for all three species were obtained from a single trajectory, the so-called single trajectory approach. We extracted one snapshot every 2 ns from the 1000 ns MD simulation. The entropy correction term to the free energies was computed using the normal mode approximation using the *mmpbsa_py_nabnmode* program included with AmberTools.[60] In this method, the vibrational frequencies of normal modes can be calculated at varied local minima of the potential energy surface, and from standard formulas the vibrational entropy contribution can be approximated.[75] Since estimation by normal modes are time-consuming, we used a maximum of 200 frames from the ones selected in the PB calculation. The binding energies were also decomposed into contributions of individual residues using the MMPBSA.py[75] script by applying the so-called per-residue decomposition. However, due to the script limitations entropy correction was not included in the energy decomposition.

Thermodynamic Integration (TI). To assess the change in the biding free energy between two compounds in the benzimidazole series, relative free energy calculations were performed by thermodynamic integration (TI). In this method, simulations are usually carried out by visiting a series of nonphysical intermediate states connecting the real end states A and B.[77] The free energy difference between two states described by two potential functions V_A and V_B, corresponding to states *A* and *B*, respectively, is given by

$$\Delta G = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda; \quad V_m = (1 - \lambda_m)V_A + \lambda_m V_B \tag{1}$$

where the coupling parameter λ varies from 0 to 1,[78] and V_m represents the combined linear potential energy. The angular brackets indicate a Boltzmann-weighted ensemble average taken at a given λ-value. The integral in (eq 1) generally cannot be solved analytically. Therefore, the trapezoid rule was chosen as the numeric integration method.

As free energy is a state function, the $\Delta\Delta G$ variation can be rigorously calculated according to eq 2[77] that is based on a thermodynamic cycle:

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

$$\Delta\Delta G_{\text{theoretical}} = \Delta G_A - \Delta G_B = \Delta G_C - \Delta G_S \quad (2)$$

where $\Delta\Delta G_{\text{theoretical}}$ means the computed relative binding free energies, ΔG_A and ΔG_B stand for the binding of two different ligands to a receptor, while ΔG_C and ΔG_S are transformations from one ligand to the other bound to the host molecule or simply solvated in water, respectively.

TI simulations were carried out by the CPU implementation of PMEMD in AMBER 14.0 using a dual-topology and multistep approach,[79] keeping the same parameters utilized in previous MD simulations (see Supporting Information). For each λ_i-window, initial coordinates were optimized by 1000 steps of steepest descent algorithm. Each system was heated from 0 to 300 K over 250 ps, and an additional equilibration simulation was performed to adjust the density in the NPT ensemble for more 250 ps.

For the two systems, the ligand in solution and the ligand−receptor complex, the simulation time to each λ-window was set to 0.5 ns for switching the charge on and off. Afterward, the vdW-transformation stage was assigned to take one nanosecond for each λ-window.

In total, 21 λ-windows for each ligand transformation were performed in steps of $\Delta\lambda = 0.05$. Forward $(A \rightarrow B)$ and backward $(B \rightarrow A)$ transformations were calculated, where A and B represent the initial and final states, respectively, and results are shown as the average of both.

The statistical error can be calculated as the square root of a weighted sum of the variances corresponding to the intermediate λ-states:

$$\sigma_{\Delta G} = \sqrt{\sum_{\lambda} w_{\lambda}^2 \sigma_{\lambda}^2} \quad (3)$$

where σ_λ is the standard error of the mean for the $\partial V/\partial\lambda$ values of the λth window. To compare computational results to the TI outcomes, we used the IC_{50} values of the transforming compounds to estimate the experimental relative free energy of binding between a pair of molecules according to eq 4,

$$\Delta\Delta G_{\text{experimental}} = -kT \ln \frac{IC_{50,1}}{IC_{50,2}} \quad (4)$$

where $IC_{50,1}$ and $IC_{50,2}$ correspond to compounds 1 and 2, respectively. k is the Boltzmann constant, and T the absolute temperature.

The Root-Mean-Square-Error (RMSE) between the experimental and calculated free energies was also computed and determined by

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\Delta\Delta G_{\text{experimental}} - \Delta\Delta G_{\text{theoretical}})^2} \quad (5)$$

where N is the number of TI transformations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00557.

A description of compound 1 putative pose in rhodesain from molecular docking; information on the dynamic behavior of compound 1 (lead) from MD simulations; Figure S1 contains RMSD plots of the protonated and neutral state of the lead compound in each system using the crystallographic position; Figure S2 illustrates the RMSD plots of the protonated and neutral state of the lead compound in each system where rhodesain initial pose was achieved by molecular docking; Figure S3 shows first frame RMSD plots of the protonated and neutral state of the lead compound 1 in each system; Table S1 contains the frequency of hydrogen bonds observed throughout MD simulations of compound 1 (protonated or neutral) bound to cruzain and rhodesain; Table S2 displays the frequency of close contacts observed throughout MD simulations of compound 1 (protonated) bound to cruzain or rhodesain; Table S3 presents the averaged binding free energies decomposed in contributions and calculated by MMPBSA.py; Tables S4 and S5 displays the outcomes of all TI transformation steps of the compounds in cruzain and rhodesain; Tables S6, S7, and Figures S4, S5 and S6 exhibits cluster analysis results for the protonated and neutral state of compound 1; Table S8 presents the frequencies of hydrogen bond interactions for TI transformations; Table S9 contains the frequencies of close contacts from TI simulations; Figure S7 shows the representative clusters from TI calculation 1 ↔ 24; Table S10 displays the average distances between linker oxygen and enzymes from 1 ↔ 24 calculation; Figure S8 shows the representative clusters from TI calculation 1 ↔ 5; in Table S11, the frequency of close contacts between halogens atoms from TI transformation 1 ↔ 5 can be found; Figure S9 reveals the surface visualization of compound 1 in the negatively charged binding site colored by electrostatic potential bound to cruzain and rhodesain (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rafaelasf@icb.ufmg.br.

ORCID

Klaus R. Liedl: 0000-0002-0985-2299

Rafaela S. Ferreira: 0000-0003-3324-0601

Present Addresses

⊥J.E.F.: Department of Medicinal Chemistry, Boehringer Ingelheim RCV GmbH & Co KG, Dr. Boehringer Gasse 5−11, 1120 Vienna, Austria;

@G.A.N.P.: Sys2Diag UNR 9005 FRE 3690 CNRS/ALCEDIAG, Cap Delta, 1682 Rue de la Valsière, 34184 Montpellier, France.

Author Contributions

L.H.S. performed all docking and molecular dynamics simulations, with contributions from B.J.W., J.E.F., K.R.L., E.R.C., and R.S.F. All authors contributed to the design of experiments and discussion of results. L.H.S., G.A.N.P., E.R.C., and R.S.F. wrote the manuscript, with revisions and contributions of all authors.

Funding

The authors are grateful to the Brazilian agencies CAPES, FAPEMIG, and CNPq for funding this research. G.A.N.P. was a postdoctoral scholar supported by the grants Special Visitant Researcher (PVE A118/2013) and Edital 51/2013 Biologia Computacional (CAPES). K.R.L. was a Special Visitant Researcher (PVE A118/2013) scholar, and L.H.S. received a CAPES PDSE (99999.010357/2014-09) scholarship. R.S.F. holds a CNPq researcher fellowship (Bolsa de Produtividade

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

em Pesquisa), level 2, and has been awarded the L'Oréal-UNESCO-ABC "For Women in Science Award, Category Chemistry, Brazil, 2017" and L'Oréal-UNESCO "International Rising Talent 2018".

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

In addition to acknowledgments to the funding agencies, the computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC).

■ ABBREVIATIONS

HAT, human African trypanosomiasis; eq, equation; MD, molecular dynamics; NPT, isothermal−isobaric ensemble; NVT, canonical ensemble; PDB, Protein Data Bank; rmsd, root-mean-square deviation; TI, thermodynamic integration; MMPB/SA, molecular mechanics Poisson−Boltzmann surface area

■ REFERENCES

(1) *Disease Reference Group on Chagas Disease, Human African Trypanosomiasis and Leishmaniasis: WHO Technical Report Series (No. 975). Research Priorities for Chagas Disease, Human African Trypanosomiasis and Leishmaniasis Research Priorities for Chagas Disea*; World Health Organization, 2012.

(2) Bern, C. Antitrypanosomal Therapy for Chronic Chagas' Disease. *N. Engl. J. Med.* **2011**, *364* (26), 2527−2534.

(3) Steverding, D. The Development of Drugs for Treatment of Sleeping Sickness: A Historical Review. *Parasites Vectors* **2010**, *3* (1), 15.

(4) Sajid, M.; McKerrow, J. H. Cysteine Proteases of Parasitic Organisms. *Mol. Biochem. Parasitol.* **2002**, *121* (1), 159.

(5) McKerrow, J. The Proteases and Pathogenicity of Parasitic Protozoa. *Annu. Rev. Microbiol.* **1993**, *47* (1), 821−853.

(6) Sajid, M.; Robertson, S. A.; Brinen, L. S.; McKerrow, J. H. Cruzain. In *Cysteine Proteases of Pathogenic Organisms*; Springer, 2011; pp 100−115.

(7) da Silva, E. B.; do Nascimento Pereira, G. A.; Ferreira, R. S. Trypanosomal Cysteine Peptidases: Target Validation and Drug Design Strategies. *A Compr. Anal. Parasite Biol. From Metab. to Drug Discovery* **2016**, *7*, 121−145.

(8) Vieira, R. P.; Santos, V. C.; Ferreira, R. S. Structure-Based Approaches Targeting Parasite Cysteine Proteases. *Curr. Med. Chem.* **2017**, *24*, 1.

(9) Ferreira, L. G.; Andricopulo, A. D. Targeting Cysteine Proteases in Trypanosomatid Disease Drug Discovery. *Pharmacol. Ther.* **2017**, *180*, 49−61.

(10) Cazzulo, J.-J.; Stoka, V.; Turk, V. The Major Cysteine Proteinase of Trypanosoma Cruzi: A Valid Target for Chemotherapy of Chagas Disease. *Curr. Pharm. Des.* **2001**, *7* (12), 1143−1156.

(11) Kerr, I. D.; Wu, P.; Marion-Tsukamaki, R.; Mackey, Z. B.; Brinen, L. S. Crystal Structures of TbCatB and Rhodesain, Potential Chemotherapeutic Targets and Major Cysteine Proteases of Trypanosoma Brucei. *PLoS Neglected Trop. Dis.* **2010**, *4* (6), No. e701.

(12) Jaishankar, P.; Hansell, E.; Zhao, D. M.; Doyle, P. S.; McKerrow, J. H.; Renslo, A. R. Potency and Selectivity of P2/P3-Modified Inhibitors of Cysteine Proteases from Trypanosomes. *Bioorg. Med. Chem. Lett.* **2008**, *18* (2), 624−628.

(13) Kerr, I. D.; Lee, J. H.; Farady, C. J.; Marion, R.; Rickert, M.; Sajid, M.; Pandey, K. C.; Caffrey, C. R.; Legac, J.; Hansell, E.; McKerrow, J. H.; Craik, C. S.; Rosenthal, P. J.; Brinen, L. S. Vinyl Sulfones as Antiparasitic Agents and a Structural Basis for Drug Design. *J. Biol. Chem.* **2009**, *284* (38), 25697−25703.

(14) Bryant, C.; Kerr, I. D.; Debnath, M.; Ang, K. K. H.; Ratnam, J.; Ferreira, R. S.; Jaishankar, P.; Zhao, D. M.; Arkin, M. R.; McKerrow, J. H.; Brinen, L. S.; Renslo, A. R. Novel Non-Peptidic Vinylsulfones Targeting the S2 and S3 Subsites of Parasite Cysteine Proteases. *Bioorg. Med. Chem. Lett.* **2009**, *19* (21), 6218−6221.

(15) Mott, B. T.; Ferreira, R. S.; Simeonov, A.; Jadhav, A.; Ang, K. K. H.; Leister, W.; Shen, M.; Silveira, J. T.; Doyle, P. S.; Arkin, M. R.; McKerrow, J. H.; Inglese, J.; Austin, C. P.; Thomas, C. J.; Shoichet, B. K.; Maloney, D. J. Identification and Optimization of Inhibitors of Trypanosomal Cysteine Proteases: Cruzain, Rhodesain, and TbCatB. *J. Med. Chem.* **2010**, *53* (1), 52−60.

(16) Giroud, M.; Kuhn, B.; Saint-Auret, S.; Kuratli, C.; Martin, R. E.; Schuler, F.; Diederich, F.; Kaiser, M.; Brun, R.; Schirmeister, T.; Haap, W. 2H-1,2,3-Triazole-Based Dipeptidyl Nitriles: Potent, Selective, and Trypanocidal Rhodesain Inhibitors by Structure-Based Design. *J. Med. Chem.* **2018**, *61* (8), 3370−3388.

(17) Greenbaum, D. C.; Mackey, Z.; Hansell, E.; Doyle, P.; Gut, J.; Caffrey, C. R.; Lehrman, J.; Rosenthal, P. J.; McKerrow, J. H.; Chibale, K. Synthesis and Structure-Activity Relationships of Parasiticidal Thiosemicarbazone Cysteine Protease Inhibitors against Plasmodium Falciparum, Trypanosoma Brucei, and Trypanosoma Cruzi. *J. Med. Chem.* **2004**, *47* (12), 3212−3219.

(18) Fonseca, N. C.; Da Cruz, L. F.; Da Silva Villela, F.; Do Nascimento Pereira, G. A.; De Siqueira-Neto, J. L.; Kellar, D.; Suzuki, B. M.; Ray, D.; De Souza, T. B.; Alves, R. J.; Júnior, P. A. S.; Romanha, A. J.; Murta, S. M. F.; McKerrow, J. H.; Caffrey, C. R.; De Oliveira, R. B.; Ferreira, R. S. Synthesis of a Sugar-Based Thiosemicarbazone Series and Structure-Activity Relationship versus the Parasite Cysteine Proteases Rhodesain, Cruzain, and Schistosoma Mansoni Cathepsin B1. *Antimicrob. Agents Chemother.* **2015**, *59* (5), 2666−2677.

(19) Espíndola, J. W. P.; De Oliveira Cardoso, M. V.; De Oliveira Filho, G. B.; Oliveira E Silva, D. A.; Moreira, D. R. M.; Bastos, T. M.; De Simone, C. A.; Soares, M. B. P.; Villela, F. S.; Ferreira, R. S.; De Castro, M. C. A. B.; Pereira, V. R. A.; Murta, S. M. F.; Sales Junior, P. A.; Romanha, A. J.; Leite, A. C. L. Synthesis and Structure-Activity Relationship Study of a New Series of Antiparasitic Aryloxyl Thiosemicarbazones Inhibiting Trypanosoma Cruzi Cruzain. *Eur. J. Med. Chem.* **2015**, *101*, 818−835.

(20) Mallari, J. P.; Shelat, A.; Kosinski, A.; Caffrey, C. R.; Connelly, M.; Zhu, F.; McKerrow, J. H.; Guy, R. K. Discovery of Trypanocidal Thiosemicarbazone Inhibitors of Rhodesain and TbcatB. *Bioorg. Med. Chem. Lett.* **2008**, *18* (9), 2883−2885.

(21) Choe, Y.; Brinen, L. S.; Price, M. S.; Engel, J. C.; Lange, M.; Grisostomi, C.; Weston, S. G.; Pallai, P. V.; Cheng, H.; Hardy, L. W.; Hartsough, D. S.; McMakin, M.; Tilton, R. F.; Baldino, C. M.; Craik, C. S. Development of α-Keto-Based Inhibitors of Cruzain, a Cysteine Protease Implicated in Chagas Disease. *Bioorg. Med. Chem.* **2005**, *13* (6), 2141−2156.

(22) Brak, K.; Kerr, I. D.; Barrett, K. T.; Fuchi, N.; Debnath, M.; Ang, K.; Engel, J. C.; McKerrow, J. H.; Doyle, P. S.; Brinen, L. S.; Ellman, J. A. Nonpeptidic Tetrafluorophenoxymethyl Ketone Cruzain Inhibitors as Promising New Leads for Chagas Disease Chemotherapy. *J. Med. Chem.* **2010**, *53* (4), 1763−1773.

(23) Ferreira, R. S.; Simeonov, A.; Jadhav, A.; Eidam, O.; Mott, B. T.; Keiser, M. J.; McKerrow, J. H.; Maloney, D. J.; Irwin, J. J.; Shoichet, B. K. Complementarity between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors. *J. Med. Chem.* **2010**, *53* (13), 4891−4905.

(24) Pereira, G. A. N.; Santos, L. H.; Villela, F. S.; Dessoy, M. A.; Dias, L. C.; Andricopulo, A. D.; Costa, M. A. F.; Martins, L. C.; Wang, S. C.; Nagem, R. A. P.; Caffrey, C. R.; Liedl, K. R.; Caffarena, E. R.; Ferreira, R. S. Unpublished results, Benzimidazoles as a Novel Class of Potent Rhodesain Inhibitors, 2018.

(25) Ferreira, R. S.; Dessoy, M. A.; Pauli, I.; Souza, M. L.; Krogh, R.; Sales, A. I. L.; Oliva, G.; Dias, L. C.; Andricopulo, A. D. Synthesis, Biological Evaluation, and Structure-Activity Relationships of Potent Noncovalent and Nonpeptidic Cruzain Inhibitors as Anti-Trypanosoma Cruzi Agents. *J. Med. Chem.* **2014**, *57* (6), 2380−2392.

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

(26) Chen, Y. T.; Lira, R.; Hansell, E.; McKerrow, J. H.; Roush, W. R. Synthesis of Macrocyclic Trypanosomal Cysteine Protease Inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18* (22), 5860−5863.
(27) Wiggers, H. J.; Rocha, J. R.; Fernandes, W. B.; Sesti-Costa, R.; Carneiro, Z. A.; Cheleski, J.; da Silva, A. B. F.; Juliano, L.; Cezari, M. H. S.; Silva, J. S.; McKerrow, J. H.; Montanari, C. A. Non-Peptidic Cruzain Inhibitors with Trypanocidal Activity Discovered by Virtual Screening and In Vitro Assay. *PLoS Neglected Trop. Dis.* **2013**, *7* (8), No. e2370.
(28) Latorre, A.; Schirmeister, T.; Kesselring, J.; Jung, S.; Johé, P.; Heilos, A.; Engels, B.; Krauth-Siegel, R. L.; Dirdjaja, N.; Hellmich, U. A.; et al. Dipeptidyl Nitroalkenes as Covalent Reversible Inhibitors of Cysteine Proteases. *ACS Med. Chem. Lett.* **2016**, *7* (12), 1073−1076.
(29) Previti, S.; Ettari, R.; Cosconati, S.; Amendola, G.; Chouchene, K.; Wagner, A.; Hellmich, U. A.; Ulrich, K.; Krauth-Siegel, R. L.; Wich, P. R.; Schmid, I.; Schirmeister, T.; Gut, J.; Rosenthal, P. J.; Grasso, S.; Zappalà, M. Development of Novel Peptide-Based Michael Acceptors Targeting Rhodesain and Falcipain-2 for the Treatment of Neglected Tropical Diseases (NTDs). *J. Med. Chem.* **2017**, *60* (16), 6911−6923.
(30) Braga, S. F. P.; Martins, L. C.; da Silva, E. B.; Sales Júnior, P. A.; Murta, S. M. F.; Romanha, A. J.; Soh, W. T.; Brandstetter, H.; Ferreira, R. S.; de Oliveira, R. B. Synthesis and Biological Evaluation of Potential Inhibitors of the Cysteine Proteases Cruzain and Rhodesain Designed by Molecular Simplification. *Bioorg. Med. Chem.* **2017**, *25* (6), 1889−1900.
(31) Giroud, M.; Dietzel, U.; Anselm, L.; Banner, D.; Kuglstatter, A.; Benz, J.; Blanc, J. B.; Gaufreteau, D.; Liu, H.; Lin, X.; Stich, A.; Kuhn, B.; Schuler, F.; Kaiser, M.; Brun, R.; Schirmeister, T.; Kisker, C.; Diederich, F.; Haap, W. Repurposing a Library of Human Cathepsin L Ligands: Identification of Macrocyclic Lactams as Potent Rhodesain and Trypanosoma Brucei Inhibitors. *J. Med. Chem.* **2018**, *61* (8), 3350−3369.
(32) Pauli, I.; Ferreira, L. G.; De Souza, M. L.; Oliva, G.; Ferreira, R. S.; Dessoy, M. A.; Slafer, B. W.; DIas, L. C.; Andricopulo, A. D. Molecular Modeling and Structure-Activity Relationships for a Series of Benzimidazole Derivatives as Cruzain Inhibitors. *Future Med. Chem.* **2017**, *9* (7), 641−657.
(33) Kuhn, B.; Tichý, M.; Wang, L.; Robinson, S.; Martin, R. E.; Kuglstatter, A.; Benz, J.; Giroud, M.; Schirmeister, T.; Abel, R.; Diederich, F.; Hert, J. Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors. *J. Med. Chem.* **2017**, *60* (6), 2485−2497.
(34) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50* (7), 1625−1632.
(35) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93* (7), 2395−2417.
(36) *Marvin Sketch 15.8.31*; Chemaxon, 2015.
(37) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for PKa Prediction and Protonation State Generation for Drug-like Molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21* (12), 681−691.
(38) Brinen, L. S.; Hansell, E.; Cheng, J.; Roush, W. R.; McKerrow, J. H.; Fletterick, R. J. A Target within the Target: Probing Cruzain's P1′ Site to Define Structural Determinants for the Chagas' Disease Protease. *Structure* **2000**, *8* (8), 831−840.
(39) Chen, Y. T.; Brinen, L. S.; Kerr, I. D.; Hansell, E.; Doyle, P. S.; McKerrow, J. H.; Roush, W. R. In Vitro and in Vivo Studies of the Trypanocidal Properties of WRR-483 against Trypanosoma Cruzi. *PLoS Neglected Trop. Dis.* **2010**, *4* (9), No. e825.
(40) Marion, R.; Hansell, E.; Caffrey, C.; Roush, W. R.; Brinen, L. S. Unpublished results, The High Resolution Crystal Structure of Rohedsain, the Major Cathepsin L Protease from T. Brucei Rhodesiense, Bound to Inhibitor K11002, 2008.
(41) McGrath, M. E.; Eakin, A. E.; Engel, J. C.; McKerrow, J. H.; Craik, C. S.; Fletterick, R. J. The Crystal Structure of Cruzain: A Therapeutic Target for Chagas' Disease. *J. Mol. Biol.* **1995**, *247* (2), 251−259.
(42) Martins, L. C.; Torres, P. H. M.; de Oliveira, R. B.; Pascutti, P. G.; Cino, E. A.; Ferreira, R. S. Investigation of the Binding Mode of a Novel Cruzain Inhibitor by Docking, Molecular Dynamics, Ab Initio and MM/PBSA Calculations. *J. Comput.-Aided Mol. Des.* **2018**, *32* (5), 591−605.
(43) Novinec, M.; Lenarcic, B. Papain-like Peptidases: Structure, Function, and Evolution. *Biomol. Concepts* **2013**, *4* (3), 287−308.
(44) Gillmor, S. A.; Craik, C. S.; Fletterick, R. J. Structural Determinants of Specificity in the Cysteine Protease Cruzain. *Protein Sci.* **1997**, *6* (8), 1603−1611.
(45) Shirts, M. R.; Mobley, D. L. An Introduction to Best Practices in Free Energy Calculations. *Methods Mol. Biol.* **2013**, *924*, 271−311.
(46) Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. Parallelized-over-Parts Computation of Absolute Binding Free Energy with Docking and Molecular Dynamics. *J. Chem. Phys.* **2006**, *125* (8), 084901.
(47) Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P. Trypsin-Ligand Binding Free Energies from Explicit and Implicit Solvent Simulations with Polarizable Potential. *J. Comput. Chem.* **2009**, *30* (11), 1701−1711.
(48) Michel, J.; Essex, J. W. Hit Identification and Binding Mode Predictions by Rigorous Free Energy Simulations. *J. Med. Chem.* **2008**, *51* (21), 6654−6664.
(49) Wang, Q.; Edupuganti, R.; Tavares, C. D. J.; Dalby, K. N.; Ren, P. Using Docking and Alchemical Free Energy Approach to Determine the Binding Mechanism of EEF2K Inhibitors and Prioritizing the Compound Synthesis. *Front. Mol. Biosci.* **2015**, *2*, 9.
(50) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. Principles and Applications of Halogen Bonding in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2013**, *56* (4), 1363−1388.
(51) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical Free Energy Calculations for Drug Discovery. *J. Chem. Phys.* **2012**, *137* (23), 230901.
(52) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46* (1), 531−558.
(53) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−242.
(54) Labute, P. Protonate3D: Assignment of Ionization States and Hydrogen Coordinates to Macromolecular Structures. *Proteins: Struct., Funct., Genet.* **2009**, *75* (1), 187−205.
(55) Cstorer, A.; Ménard, R. Catalytic Mechanism in Papain Family of Cysteine Peptidases. *Methods Enzymol.* **1994**, *244* (C), 486−500.
(56) Otto, H.-H.; Schirmeister, T. Cysteine Proteases and Their Inhibitors. *Chem. Rev.* **1997**, *97* (1), 133−172.
(57) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.
(58) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21* (2), 132−146.
(59) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25* (2), 247−260.
(60) Case, D. A.; Babin, V.; Berryman, J.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H. *Amber*; 2014.
(61) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49* (21), 6177−6196.
(62) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148

Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739−1749.
(63) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750−1759.
(64) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78* (8), 1950−1958.
(65) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(66) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(67) Adelman, S. A. Quantum Generalized Langevin Equation Approach to Gas/Solid Collisions. *Chem. Phys. Lett.* **1976**, *40* (3), 495−499.
(68) Wallnoefer, H. G.; Handschuh, S.; Liedl, K. R.; Fox, T. Stabilizing of a Globular Protein by a Highly Complex Water Network: A Molecular Dynamics Simulation Study on Factor Xa. *J. Phys. Chem. B* **2010**, *114* (21), 7405−7412.
(69) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084−3095.
(70) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33−38.
(71) Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36* (15), 1132−1156.
(72) Allen, W. J.; Rizzo, R. C. Implementation of the Hungarian Algorithm to Account for Ligand Symmetry and Similarity in Structure-Based Design. *J. Chem. Inf. Model.* **2014**, *54* (2), 518−529.
(73) Team, R. C. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013 (http://www.R-project.org/).
(74) Ester, M., Kriegel, H. P., Sander, J., Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Kdd*; 1996; Vol. *96*, pp 226−231.
(75) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314−3321.
(76) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol. Inf.* **2012**, *31* (2), 114−122.
(77) Pearlman, D. A. Free Energy Calculations: Methods for Estimating Ligand Binding Affinities. *Free Energy Calculations in Rational Drug Design*; Kluwer Academic/Plenum Publishers; Reddy, R., Erion, M. D., Eds.; 2001; pp 9−35.
(78) Brandsdal, B. O.; Österberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Free Energy Calculations and Ligand Binding. *Adv. Protein Chem.* **2003**, *66*, 123−158.
(79) Kaus, J. W.; Pierce, L. T.; Walker, R. C.; McCammon, J. A. Improving the Efficiency of Free Energy Calculations in the Amber Molecular Dynamics Package. *J. Chem. Theory Comput.* **2013**, *9* (9), 4131−4139.

DOI: 10.1021/acs.jcim.8b00557
J. Chem. Inf. Model. 2019, 59, 137−148