

## **Intramolecular interactions enhance the potency of gallinamide A analogs against *Trypanosoma cruzi***

Elany Barbosa Da Silva<sup>1#</sup>, Vandna Sharma<sup>1#</sup>, Lilian Hernandez-Alvarez<sup>1,2</sup>, Arthur H. Tang<sup>4,5</sup>, Alexander Stoye<sup>4,5</sup>, Anthony J. O'Donoghue<sup>1</sup>, William H. Gerwick<sup>1,3</sup>, Richard J. Payne<sup>4,5</sup>, James H. McKerrow<sup>1</sup>, Larissa M. Podust<sup>1,\*</sup>.

<sup>1</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, Center for Discovery and Innovation in Parasitic Diseases, University of California San Diego, La Jolla, California

<sup>2</sup>Departamento de Física, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista Julio de Mesquita Filho, São José do Rio Preto, São Paulo, Brazil.

<sup>3</sup>Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California

<sup>4</sup>School of Chemistry, The University of Sydney, Sydney, NSW 2006, Australia.

<sup>5</sup>Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science.

# These authors contributed equally

Running title: X-ray structures of the gallinamide-cruzain complexes

## ABSTRACT

Gallinamide A, a metabolite of the marine cyanobacterium *Schizothrix* sp., selectively inhibits cathepsin L-like cysteine proteases. We evaluated potency of gallinamide A and 23 synthetic analogs against intracellular *Trypanosoma cruzi* amastigotes and the cysteine protease, cruzain. We determined the co-crystal structures of cruzain with gallinamide A and two synthetic analogs at  $\sim 2\text{\AA}$ . SAR data revealed that the N-terminal end of gallinamide A is loosely bound and weakly contributes in drug-target interactions. At the C-terminus, the intramolecular  $\pi$ - $\pi$  stacking interactions between the aromatic substituents at P1' and P1 restrict the bioactive conformation of the inhibitors, thus minimizing the entropic loss associated with target binding. Molecular dynamics simulations showed that in the absence of an aromatic group at P1, the substituent at P1' interacts with tryptophan-184. The P1-P1' interactions had no effect on anti-cruzain activity whereas anti-*T. cruzi* potency increased by  $\sim 5$ -fold, likely due to an increase in solubility/permeability of the analogs.

## INTRODUCTION

The World Health Organization recognizes Chagas disease, caused by the protozoan parasite *Trypanosoma cruzi*, as one of the most prevalent neglected tropical diseases in the World. In Latin America, Chagas disease is endemic, and about 8 million people are estimated to be infected worldwide<sup>1-4</sup>. The global healthcare and economic burden of Chagas disease is substantial<sup>5-7</sup>, while the clinically available treatments are limited to two drugs, benznidazole and nifurtimox. Both drugs have been recently FDA-approved for use in children of 2 to 12 years old (benznidazole) and from birth to 18 years old (nifurtimox)<sup>8-10</sup>, however, neither has been approved for adults. While benznidazole and nifurtimox have significant efficacy in the acute phase, both drugs suffer from the liabilities of low efficacy in chronically infected patients and adults and may cause serious adverse reactions leading to premature termination of treatment<sup>11</sup>. Thus, development of new anti-Chagas drugs with improved efficacy and less toxicity is a priority.

The goal of developing safer and more efficacious drugs for the treatment of Chagas disease led to an investigation of natural secondary metabolites from plants and cyanobacteria for activity against *T. cruzi*<sup>12-14</sup>. Cyanobacteria produce a large variety of bioactive molecules attractive for pharmaceutical applications in various fields of anti-infective drug discovery, including antibacterial, antifungal, antiviral, and antiprotozoal agents<sup>15-17</sup>. Tropical filamentous marine cyanobacteria have emerged as an important source of biologically active secondary metabolites<sup>18</sup>. The natural product gallinamide A, isolated from marine cyanobacteria discovered at a tropical reef near Piedras Gallinas on the Caribbean coast of Panama, was first identified as an inhibitor of *Plasmodium falciparum* growth<sup>19</sup>. Additional studies showed that the mechanism of action for this compound was due to inhibition of the food vacuole cysteine proteases that are collectively known as falcipains<sup>20,21</sup>. *P. falciparum* falcipain enzymes are cathepsin L-like proteases and therefore additional studies showed that gallinamide A is also a potent and highly selective inhibitor of human cathepsin L<sup>22</sup>. The unique linear lipopeptide structure, and potential applicability to proteases of medical relevance, provoked interest in establishing chemical syntheses of gallinamide A and structural analogs<sup>20,23-26</sup>. Indeed, gallinamide A has become an attractive starting point for developing selective inhibitors

targeting human cathepsin L and other Clan CA cysteine proteases in human pathogens<sup>21,23,24,27</sup>. Based on comparison of the protein sequence alignment and substrate specificity, cathepsin L is similar to cysteine proteases from unicellular parasites that are responsible for neglected tropical diseases, such as falcipains from *P. falciparum* and cruzain from *T. cruzi*<sup>28–30</sup>. The cysteine protease cruzain is a promising drug target because this enzyme is essential in *T. cruzi* at all stages of the life cycle<sup>31–33</sup>.

Previously, fifteen gallinamide A analogs were synthesized and tested against cruzain and human cathepsin L<sup>23</sup>, while thirty-five additional analogs were synthesized and tested against falcipain-2 and -3 and *P. falciparum*.<sup>21,27</sup> Recently, both sets of analogs were evaluated against SARS-CoV-2 and analogs **19** and **23** demonstrated efficacy against viral infection in VeroE6 cells by inhibiting host cathepsin L in the picomolar range<sup>24</sup>. Analog numbering in this work is according to Ashhurst et al.<sup>24</sup>. The gallinamide A analogs that exhibited activity against human cathepsin L also showed potent inhibition of cruzain and intracellular *T. cruzi* amastigotes, making the gallinamide scaffold a promising lead for anti-Chagas drug development<sup>23</sup>. Specifically, the parent gallinamide A natural product inhibited cruzain with IC<sub>50</sub> of 0.26 nM and intracellular *T. cruzi* amastigotes with EC<sub>50</sub> of 15 nM<sup>23</sup>.

In the absence of an experimental structure of a gallinamide-cathepsin L complex, molecular docking in human cathepsin L was instrumental in predicting a putative gallinamide A binding pose and guiding analog design for effective inhibition of the human protease<sup>22,23</sup>. Here we report the first X-ray structures of the cruzain complexes with gallinamide A and analogs that will be instrumental for developing this class of cysteine protease inhibitors as anti-Chagas agents and for other biomedical applications. We also evaluated two series of synthetic gallinamide A analogs (23 compounds) against cruzain and intracellular *T. cruzi* amastigotes. The SAR data combined with the molecular dynamics (MD) simulations, including MM-GBSA free energy decomposition and thermodynamic integrations (TI), identified hot spots and structural determinates underlying the selective inhibition of the Clan CA cysteine proteases by gallinamide A and analogs. This work lays the foundation for the design of natural product analogs with improved activities and drug-like properties.

## RESULTS

### X-ray structure analysis of the cruzain-gallinamide A complex

#### *Cruzain-gallinamide A interactions*

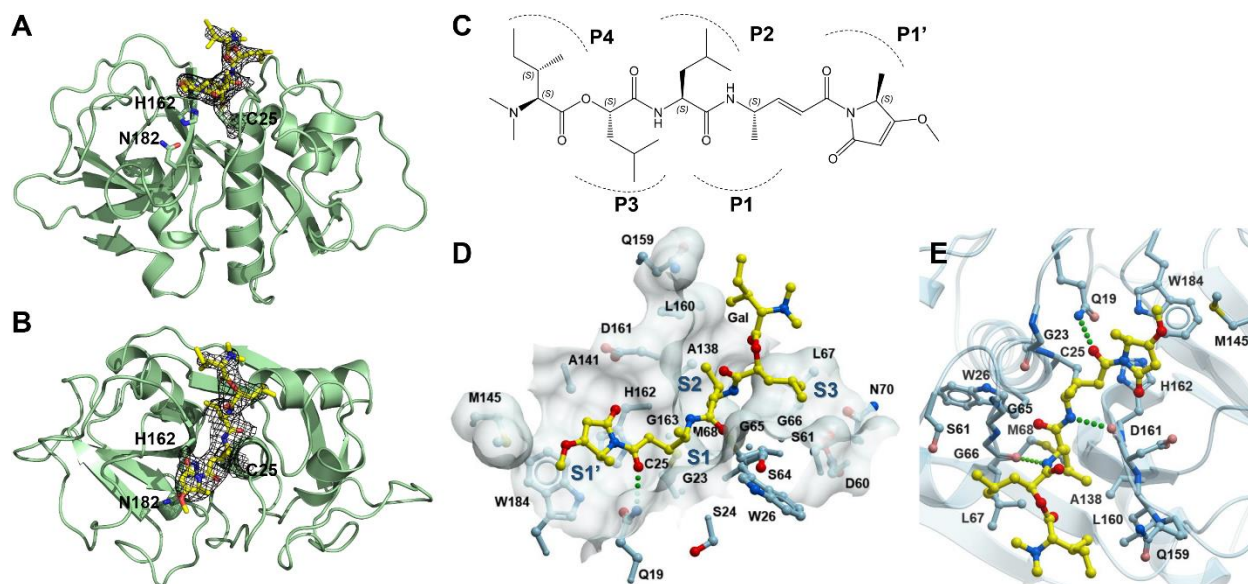
The cruzain-gallinamide A co-crystal structure containing six cruzain molecules in an asymmetric unit (chains A-F) was determined to a resolution of 2.2 Å. Data collection and refinement statistics are provided in **Table 1**. In all chains, electron density was present in the active side cleft, suggesting that gallinamide A was bound in an extended linear conformation via a covalent bond to the C25 thiol (**Fig. 1A, B**). Gallinamide A spans the S1', S1, S2, and S3 pockets of the cruzain active site (**Fig. 1C, D**). Binding of gallinamide A occurs via H-bonding, aromatic stacking interactions and hydrophobic interactions. Three H-bonds are formed between the inhibitor backbone and cruzain (**Fig. 1E**). The

amide NH atoms of P1 and P2 H-bond to carbonyl oxygen atoms of D161 and G66, respectively, while the carbonyl oxygen of P1' H-bonds to the NH of the Q19 side chain.

**Table 1.** Data collection and refinement statistics

<b>Data collection</b>			
Inhibitor (ID)	Gallinamide A (GN9)	<b>23</b> (83K)	<b>24</b> (83E)
PDB ID	7JUU	7S19	7S18
Space group	P 43 21 2	P 32 2 1	I 41 2 2
Cell dimensions			
<i>a, b, c</i> (Å)	139.85, 139.85, 163.15	95.93, 95.93, 68.85	99.50, 99.50, 85.37
$\alpha, \beta, \gamma$ (°)	90.0, 90.0, 90.0	90.00, 90.00, 120.00	90.00, 90.00, 90.00
Molecules in AU	6	1	1
Wavelength	1.11587	1.11587	1.11587
Resolution (Å)	2.20	2.08	2.14
$R_{\text{sym}}$ (%)	33.0 (882.6)	42.0 (555.8)	6.8 (557.3)
$I / \sigma I$	11.83 (0.43)	8.37 (0.54)	21.31 (0.89)
Completeness (%)	99.9 (99.4)	96.6 (83.08)	99.9 (70.36)
Redundancy	26.6 (26.8)	19.7 (16.6)	25.2 (26.7)
<b>Crystallization conditions</b>	0.8 M K-phosphate, pH 9.7; 0.01 M betaine hydrochloride	0.8 M K-phosphate, pH 8.0; 0.1 M CaCl <sub>2</sub> , 0.01 M betaine hydrochloride	3.2 M NaCl, 0.1 M sodium citrate, pH 5.3; 0.01 M sarcosine.
<b>Refinement statistics</b>			
No. reflections	78166	20511	12135
$R_{\text{work}} / R_{\text{free}}$ (%)	20.1/26.7	20.0/ 24.7	19.7/ 28.9
No. atoms			
Protein	9552	1592	1589
Inhibitor	252	57	60
Solvent	233	79	3
Wilson plot B	57.3	37.1	74.3
Mean B value	64.6	39.3	87.4
B-factors			
Protein	64.9	39.5	88.6
Inhibitor	91.4	46.5	94.3
Solvent	53.7	39.8	75.2
R.m.s deviations			
Bond lengths (Å)	0.013	0.019	0.013
Bond angles (°)	1.816	2.173	1.898
<b>Ramachandran statistics</b>			
Preferred (%)	95.0	95.8	92.5
Allowed (%)	5.0	3.8	6.6
Outliers (%)	0.0	0.5	0.9

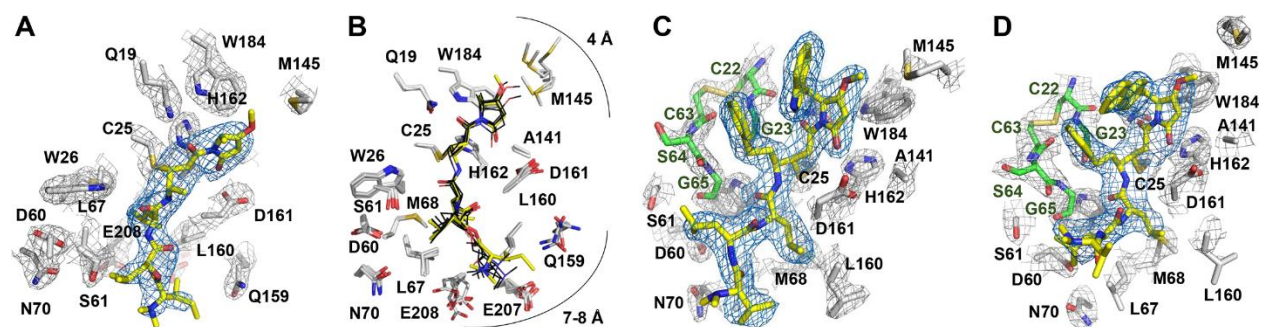
<sup>1</sup>Data for the highest resolution shell are shown in parentheses.



**Figure 1. Structure of the cruzain-gallinamide A complex.** **A, B.** View of the cruzain-gallinamide A complex in two orthogonal orientations. The view in B was obtained by 90° rotation of molecule in A toward the viewer. Cruzain (chain A) is shown as a green ribbon, gallinamide A is in yellow sticks, and the C25-H162-N182 acid-base-nucleophile triad is in green sticks. Electron density  $2F_o - F_c$  omit map (black mesh) is countered at 2.0  $\sigma$ . **C.** Chemical structure of gallinamide A. The side chains within gallinamide A are labeled according to the Schechter–Berger nomenclature<sup>34</sup>, with the first, second and third residues (from left to right) labeled as P4, P3 and P2, and fourth and fifth residues labelled as P1 and P1', respectively. **D, E.** The substrate-binding site is shown in two different orientations as a semitransparent surface (**D**) built by the amino acid residues shown in light blue and a ribbon (**E**); bound gallinamide A is in yellow balls-and-sticks. H-bonds between cruzain and gallinamide A are shown in green dots. Heteroatoms in all images are color-coded, oxygen in red, nitrogen in blue, sulfur in dark yellow.

The cyclic methylmethoxypyrrolidinone (MMP) moiety at P1' of gallinamide A is inserted between D161 and H162 in the S1' pocket and is stabilized via aromatic stacking interactions with W184 and H162 (**Fig. 2**). This orientation places the enone pharmacophore in close proximity to the thiol group of the acid-base-nucleophile triad C25-H162-N182 favoring formation of the irreversible covalent bond between C25 and gallinamide A via a Michael-addition-like reaction<sup>35</sup>. The electron density for the MMP is well defined with the exception of the 3-methoxy group due to conformational ambiguity of the latter (**Fig. 2A**). M145 in proximity of the 3-methoxy substituent of the MMP also adopts multiple conformations, suggesting a loose contact with MMP. The P2 and P3 leucine moieties of gallinamide A bind in the S2 and S3 pockets, while the *N,N*-dimethylisoleucine moiety (*N,N*-Me<sub>2</sub>-L-Ile) at P4 is exposed to the bulk solvent (**Figs. 1D, 2A**). The closest side chains to the *N,N*-Me<sub>2</sub>-L-Ile moiety include E207, E208, and Q159, and are located 7-8 Å away (**Fig. 2B**). Lack of interactions between the *N,N*-Me<sub>2</sub>-L-Ile residue and cruzain leads to binding ambiguity of the gallinamide A N-terminus and is

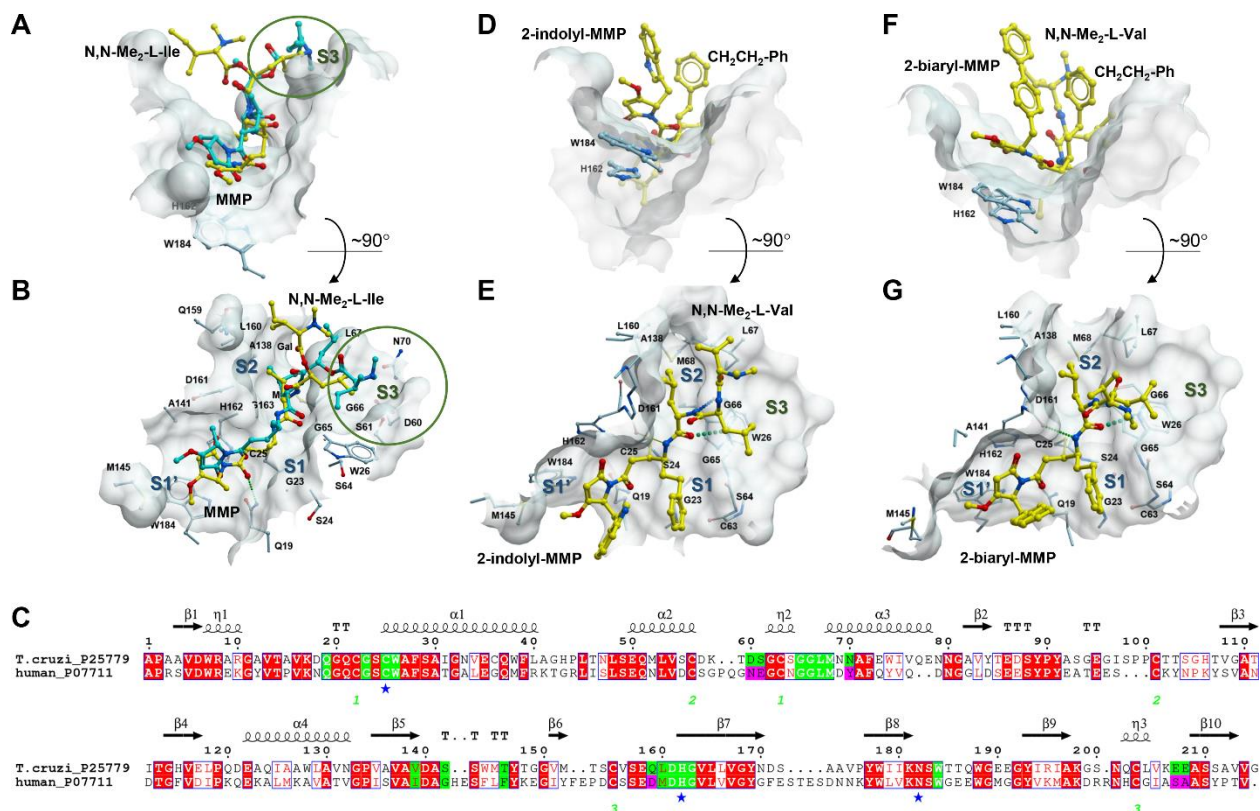
consistent with the flat SAR at P4 position of the gallinamide analogs reported elsewhere<sup>23</sup>.



**Figure 2. Cruzain-gallinamide A analog interactions.** Fragments of the electron density  $2F_o - F_c$  omit map for gallinamide A (**A**), analog **23** (**C**) and analog **24** (**D**) are countered at  $2.5 \sigma$  (blue mesh). Fragments of the electron density  $2F_o - F_c$  map for cruzain are countered at  $2 \sigma$  (gray mesh). Inhibitors are highlighted in yellow, amino acid side chains are in gray, G23, C22, C63, S64 and G65 are in green; the heteroatoms are color-coded: oxygen in red, nitrogen in blue, sulfur in dark yellow. **B.** Gallinamide A binding site is shown as an overlap between the six protein chains, A-F. Gallinamide A in chain A is in yellow, in chains B-F is in black.

### *Similarity and differences between the gallinamide A docking and X-ray poses*

The major differences between the computationally predicted docking (in the related cathepsin L protease) and experimental X-ray binding poses of gallinamide A is in binding of the *N,N*-Me<sub>2</sub>-L-Ile moiety. In the cruzain-gallinamide A X-ray structure, *N,N*-Me<sub>2</sub>-L-Ile is exposed to the bulk solvent instead of binding in the S3 pocket, as it was predicted by molecular docking in cathepsin L (compare cyan and yellow poses in **Fig. 3A, B**)<sup>22,23</sup>. While there is a possibility of having artifacts from docking, this discrepancy may also be a result of structural differences between cathepsin L and cruzain. Despite the high sequence identity (46%) and structural similarity between cathepsin L and cruzain, there are a few non-conservative amino acid substitutions in the substrate-binding site. The S3 pocket harbors three major substitutions, D60→N, S61→E and N70→Y, that distinguish the substrate binding sites of cathepsin L from cruzain (**Fig. 3C**). Three other substitutions, Q159→D, M145→L and L160→M are located in S1' and S2 pockets. In addition, a pair of consecutive glutamate residues, E207 and E208, is absent in cathepsin L, and the equivalent positions are occupied by serine and alanine, respectively.



**Figure 3. Binding modes of gallinamide A and analogs.** **A, B.** Gallinamide A binding poses in cruzain (yellow) and cathepsin L (cyan) are shown in the context of the cruzain binding site (semi-transparent surface). Encircled in green is the S3 pocket harboring non-conservative amino acid substitutions D60→N, S61→E and N70→Y. **C.** Sequence alignments between cruzain and human cathepsin L. Residues constituting the substrate-binding site are in green; non-conservative substitutions between cruzain and cathepsin L are highlighted in magenta. Blue stars indicate the catalytic acid-base-nucleophile triad C25-H162-N182. Six cysteine residues involved in formation of the disulfide bonds are numbered in green. Residue numbering is according to cruzain. **D, E.** 2-indolyl-MMP analog **23** binding in cruzain. **F, G.** 2-biaryl analog **24** binding in cruzain. Inhibitors carbon atoms are in yellow, oxygen in red, nitrogen in blue.

## Biological activity of the gallinamide A analogs

### Activity of gallinamide A analogs against cruzain and *T. cruzi* amastigotes

A total of 23 gallinamide A analogs from two series were evaluated against both *T. cruzi* cruzain and intracellular *T. cruzi* amastigotes (**Tables 2, 3**). The majority of the analogs screened exhibited inhibitory activity against cruzain with IC<sub>50</sub> values in the low nanomolar to picomolar range with twenty analogs being more potent than the parent gallinamide A natural product. In the *T. cruzi* intracellular amastigote assay, seven analogs were equipotent to gallinamide A's inhibitory activity, while others possessed inferior activity to

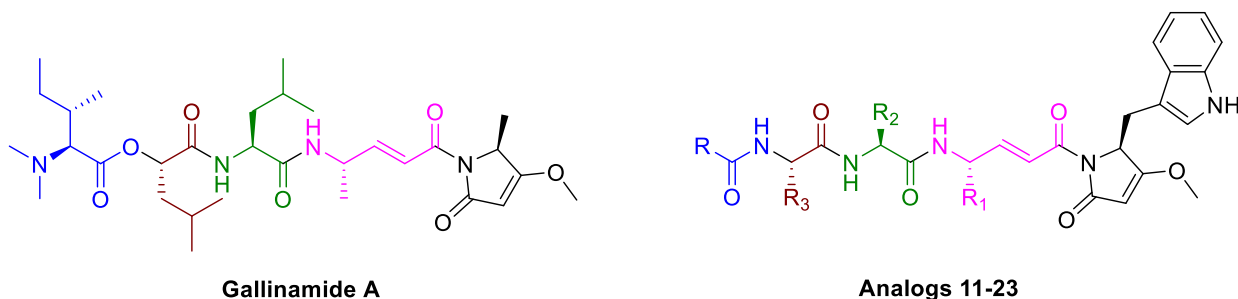
the natural product, presumably due to differences in cell permeability. Finally, the LD<sub>50</sub> toward the murine host cells was >0.6 μM for all tested compounds, yielding a selectivity index (SI) of 6,000 for the most potent analogs.

The first analog series assessed was composed of the thirteen 2-indolyl-MMP analogs **11-23** (Table 2) featuring three amino acid modules within the linear chain plus two different P4 functionalities, 4-*N*-methylpiperidine (4-(*N*-Me)piperidine) or *N,N*-Me<sub>2</sub>-L-Val, at the pseudo-N-terminus. At the C-terminal end, an indole functionality (derived from L-tryptophan) was appended to the MMP ring via a methylene bridge. All substituents at P1, P2 and P3 were introduced on the common 2-indolyl-MMP scaffold. A majority (85%) of the compounds in this series showed higher activity against cruzain than gallinamide A, with **14**, **17**, and **23** being >15-fold more potent. Two analogs, **14** and **23**, having N-terminal R=*N,N*-Me<sub>2</sub>-L-Val retained potency against *T. cruzi* amastigotes, while **17** and other analogs (**16**, **18** and **20**) having more hydrophobic R=4-(*N*-Me)piperidine at the N-terminus lost potency against *T. cruzi*. Importantly, the potent activity against the cruzain target by the analogs carrying the 4-(*N*-Me)piperidine substituent (**16**, **17**, **18** and **20**) is consistent with exposure of the N-terminal moiety observed in the crystal structures.

The hydrophobic CH<sub>2</sub>-cyclohexyl substituent was explored at P2 and P3 positions alone and in combination with the L-leucine side chain (compounds **11-14**). CH<sub>2</sub>-cyclohexyl was tolerated in both positions with the most potent configuration achieved in **14**, where R<sub>2</sub>=L-leucine and R<sub>3</sub>=CH<sub>2</sub>-cyclohexyl. A bulky aromatic substituent CH<sub>2</sub>CH<sub>2</sub>-Ph was introduced in **23** at P1 where it replaced L-alanine of the parental scaffold. In **23**, R<sub>1</sub>=CH<sub>2</sub>CH<sub>2</sub>-Ph was combined with R<sub>2</sub>=R<sub>3</sub>=L-leucine that did not affect activity against the cruzain target (compared to **11**) but enhanced approximately 5-fold the potency of **23** against *T. cruzi* amastigotes. Interestingly, the same CH<sub>2</sub>CH<sub>2</sub>-Ph group at P2 or P3 positions reduced activity against *T. cruzi* amastigotes (compounds **15**, **19** and **21**). A substitution pattern with R<sub>2</sub>=CH<sub>2</sub>CH<sub>2</sub>-Ph and R<sub>3</sub>=L-leucine largely eliminated activity of **15** against *T. cruzi*.



**Table 2.** Structure–activity relationships of gallinamide A 2-indolyl-MMP analogs **11-23**



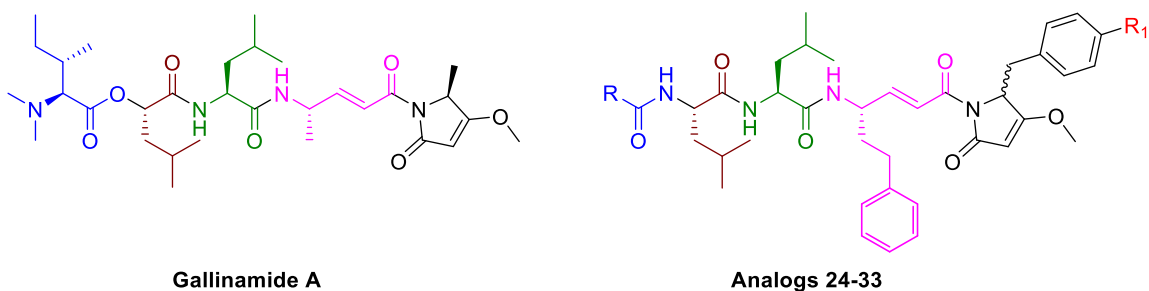
Analog	R	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	Cruzain IC <sub>50</sub> (nM) <sup>a</sup>	<i>T. cruzi</i> EC <sub>50</sub> (nM) <sup>b</sup>
<b>Gallinamide A</b>	<i>N,N</i> -Me <sub>2</sub> -L-Ile	CH <sub>3</sub>	L-leucine	L-leucine	1.3 ± 0.1	14.7 ± 2.3
<b>11</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	CH <sub>3</sub>	L-leucine	L-leucine	0.11 ± 0.0025	85 ± 25
<b>12</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	CH <sub>3</sub>	CH <sub>2</sub> -cyclohexyl	CH <sub>2</sub> -cyclohexyl	0.74 ± 0.095	30 ± 5
<b>13</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	CH <sub>3</sub>	CH <sub>2</sub> -cyclohexyl	L-leucine	0.34 ± 0.11	58 ± 19
<b>14</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	CH <sub>3</sub>	L-leucine	CH <sub>2</sub> -cyclohexyl	0.081 ± 0.013	12 ± 0.15
<b>15</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> -Ph	L-leucine	0.49 ± 0.089	>600
<b>16</b>	4-( <i>N</i> -Me)piperidine	CH <sub>3</sub>	CH <sub>2</sub> -cyclohexyl	L-leucine	1.35 ± 0.15	>600
<b>17</b>	4-( <i>N</i> -Me)piperidine	CH <sub>3</sub>	L-leucine	CH <sub>2</sub> -cyclohexyl	0.086 ± 0.0061	>600
<b>18</b>	4-( <i>N</i> -Me)piperidine	CH <sub>3</sub>	CH <sub>2</sub> -cyclohexyl	CH <sub>2</sub> -cyclohexyl	4.95 ± 1.25	196 ± 5
<b>19</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	CH <sub>3</sub>	CH <sub>2</sub> CH <sub>2</sub> -Ph	CH <sub>2</sub> -cyclohexyl	1.14 ± 0.04	40 ± 3
<b>20</b>	4-( <i>N</i> -Me)piperidine	CH <sub>3</sub>	L-leucine	L-leucine	0.19 ± 0.02	>600
<b>21</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	CH <sub>3</sub>	L-leucine	CH <sub>2</sub> -Ph	0.85 ± 0.07	56 ± 4
<b>22</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	Cyclohexyl	L-leucine	CH <sub>2</sub> -cyclohexyl	0.71 ± 0.03	18 ± 6
<b>23</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	CH <sub>2</sub> CH <sub>2</sub> -Ph	L-leucine	L-leucine	0.09 ± 0.0091	18 ± 3
<b>E-64</b>	-	-	-	-	29.73 ± 9.16	-
<b>Benznidazole</b>	-	-	-	-	-	1910 ± 700

<sup>a</sup>IC<sub>50</sub> values represent the average of two independent experiments performed in triplicate with at least 8 compound concentrations. Errors are given by the ratio between the standard deviation and the square root of the number of measurements.

<sup>b</sup>EC<sub>50</sub> are the average of two independent biological replicates performed in duplicate, +/- standard error of the mean. The LD<sub>50</sub> to the murine host cells was >0.6 μM for all tested compounds.

Introduction of biaryl-substituted MMP moieties at the C-terminus was explored by evaluating analogs **24-33** (Table 3). These compounds possessed either 4-(*N*-Me)piperidine or *N,N*-Me<sub>2</sub>-L-Val at the pseudo-N-terminus and CH<sub>2</sub>CH<sub>2</sub>-Ph at the P1 site<sup>24,27</sup>. All but one of the analogs of this series exhibited more potent activity against cruzain compared to gallinamide A. However, only three analogs, **24**, **28** and **30**, retained the anti-*T. cruzi* activity. Similar to the 2-indolyl-MMP series, analogs carrying R=4-(*N*-Me)piperidine (compounds **25**, **27**, **29**, **31** and **33**) had anti-parasitic activity inferior to the counterparts carrying R=*N,N*-Me<sub>2</sub>-L-Val. Furthermore, analogs having more hydrophobic R<sub>1</sub>'=4-CF<sub>3</sub>-Ph and R<sub>1</sub>'=Naph experienced an approximately 4-5-fold decrease in anti-*T. cruzi* activity.

**Table 3.** Structure–activity relationships of gallinamide A 2-biaryl-MMP analogs **24–33**.



Analog	R	R <sub>1</sub>	Cruzain IC <sub>50</sub> (nM) <sup>a</sup>	<i>T. cruzi</i> EC <sub>50</sub> (nM) <sup>b</sup>
Gallinamide A	<i>N,N</i> -Me <sub>2</sub> -L-Ile	CH <sub>3</sub>	1.3 ± 0.1	14.7 ± 2.3
<b>24</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	Ph	0.12 ± 0.01	13 ± 2
<b>25</b>	4-( <i>N</i> -Me)piperidine	Ph	0.40 ± 0.1	222 ± 2
<b>26</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	4-CF <sub>3</sub> Ph	0.68 ± 0.1	78 ± 8
<b>27</b>	4-( <i>N</i> -Me)piperidine	4-CF <sub>3</sub> Ph	0.79 ± 0.1	420 ± 61
<b>28</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	4-OCH <sub>3</sub> Ph	0.28 ± 0.01	19 ± 4
<b>29</b>	4-( <i>N</i> -Me)piperidine	4-OCH <sub>3</sub> Ph	0.26 ± 0.04	261 ± 90
<b>30</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	4-CNPh	0.11 ± 0.01	19 ± 7
<b>31</b>	4-( <i>N</i> -Me)piperidine	4-CNPh	0.13 ± 0.009	423 ± 100
<b>32</b>	<i>N,N</i> -Me <sub>2</sub> -L-Val	Naph	0.69 ± 0.03	59 ± 30
<b>33</b>	4-( <i>N</i> -Me)piperidine	Naph	1.56 ± 0.7	521 ± 173
E-64	-	-	29.73 ± 9.16	-
Benznidazole	-	-	-	1910 ± 700

<sup>a</sup>IC<sub>50</sub> values represent the average of two independent experiments which were determined based on at least 8 compound concentrations in triplicate. Errors are given by the ratio between the standard deviation and the square root of the number of measurements.

<sup>b</sup>EC<sub>50</sub> are the average of two independent biological replicates in duplicate, +/- standard error of the mean. The LD<sub>50</sub> to the murine host cells was >0.6 μM for all tested compounds.

### X-ray structure analysis of the cruzain-23 and cruzain-24 complexes

We also determined the co-crystal structures of cruzain covalently bound to gallinamide A analogs **23** from the 2-indolyl-MMP series (**Table 2**) and **24** from the 2-biaryl-MMP series (**Table 3**). Both analogs are among the most potent cruzain and *T. cruzi* inhibitors from this collection. Unlike cruzain-gallinamide A, the co-crystal structure of each analog contained one molecule in the asymmetric unit. Electron density for analogs was well defined with the exception of the exposed N-terminal *N,N*-dimethylvaline moiety (*N,N*-Me<sub>2</sub>-L-Val) (**Fig. 2C, D**), particularly in **24** where it pointed away from the cruzain surface (**Fig. 3D-G**). Binding ambiguity of *N,N*-Me<sub>2</sub>-L-Val resembles that of *N,N*-Me<sub>2</sub>-L-Ile in the parent gallinamide A natural product. The P2 and P3 leucine moieties interact in the S2 and S3 pockets, respectively (**Fig. 3D-G**).

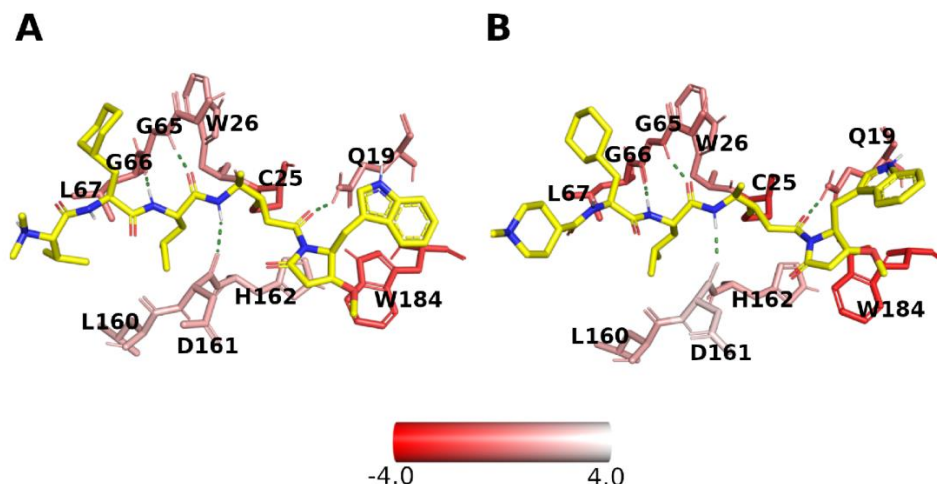
Interestingly, the indolyl moiety of **23** and the biaryl moiety of **24** established intramolecular  $\pi$ - $\pi$  stacking interactions with the P1 substituent, CH<sub>2</sub>CH<sub>2</sub>-Ph in the

molecule (**Fig. 3D, F**). While CH<sub>2</sub>CH<sub>2</sub>-Ph made surface contacts with the backbone atoms of G23, C22, C63, S64 and G65 (shown in green in **Fig. 2C, D**), neither indolyl nor biaryl moieties interacted with the protein. Intramolecular interactions of the CH<sub>2</sub>CH<sub>2</sub>-Ph at P1 with the indolyl or biaryl moieties at P1' may reduce the hydrophobic surface, increase solubility in aqueous medium, and positively impact permeability and potency of **23** and **24**. Furthermore, these intramolecular interactions may also stabilize the biologically active conformation and reduce the entropic loss associated with the binding to the target. We speculate that the role of the aromatic MMP substituents is responsible for restricting the flexibility of the analogs to lock them into their biologically active conformation. Given that no amino acid side chains are involved, these intramolecular interactions are expected to broadly enhance activity against Clan CA cysteine proteases. The **23** and **24** co-crystal structures were determined in different space groups that minimize any impact of crystal packing interactions on the inhibitor binding poses.

## Molecular modeling

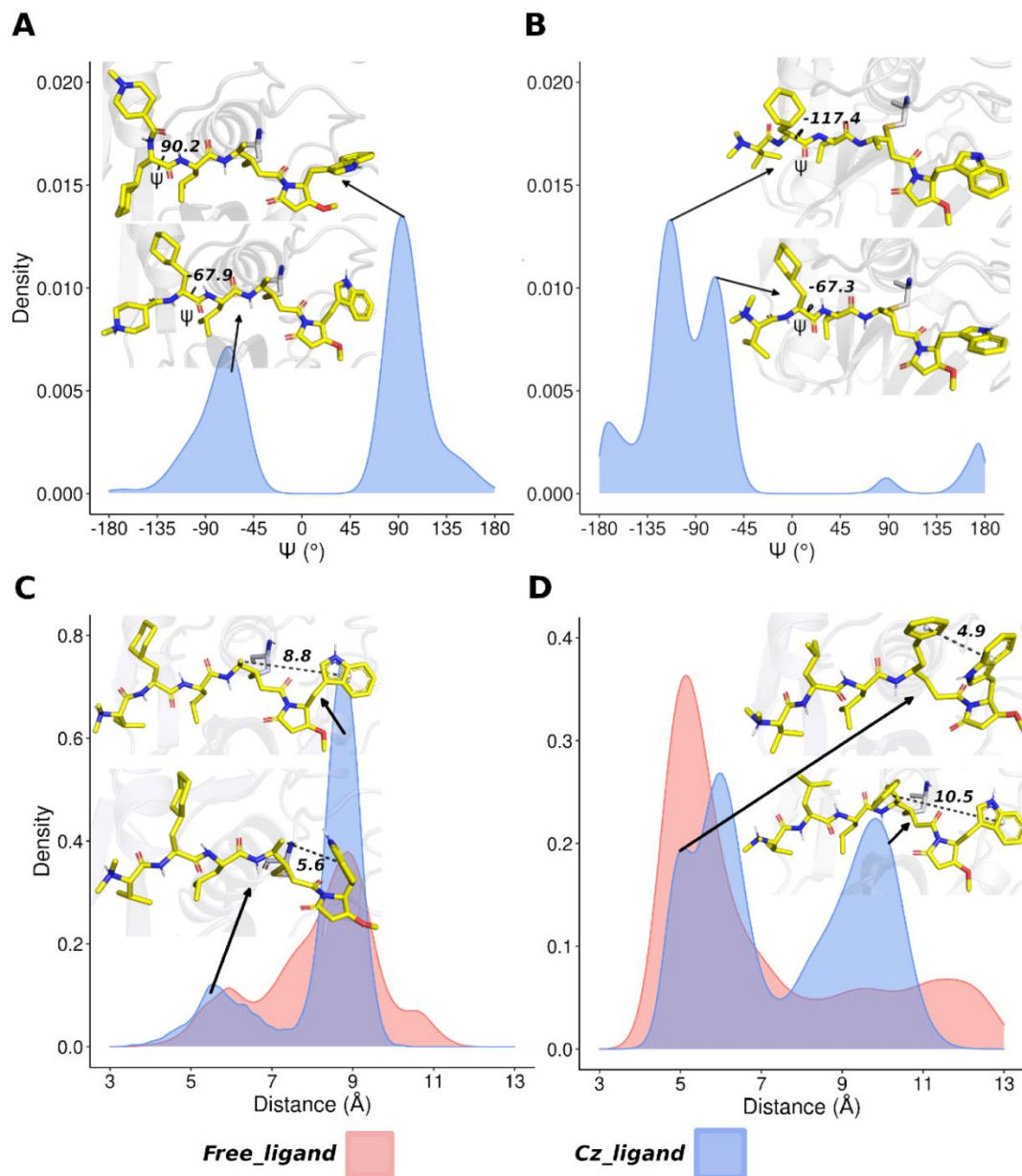
### *Preferential binding modes of 14, 17 and 23 revealed by MD simulations and per-residue free energy decomposition*

To explore dynamics of cruzain-ligand interactions *in silico*, we performed MD simulations for the top hits of the 2-indolyl-MMP series, compounds **14**, **17** and **23**, for which cruzain inhibitory activity exceeded 15-fold that of gallinamide A. A critical structural difference that distinguishes **14** and **17** from **23** is the absence of the CH<sub>2</sub>CH<sub>2</sub>-Ph moiety at P1 that excludes the stabilizing effect of the intramolecular interactions between the indolyl and CH<sub>2</sub>CH<sub>2</sub>-Ph moieties in **14** and **17**. We also calculated the per-residue free energy decomposition ( $\Delta G_{res}$ ) employing the structure ensemble generated for each complex simulation. As shown in **Fig. 4**, residues W184, L160 and L67 established non-polar interactions with the 2-indolyl-MMP, L-leucine, CH<sub>2</sub>-cyclohexyl and *N,N*-Me<sub>2</sub>-L-Val moieties at P1', P2, P3 and P4 positions, respectively. In resemblance to the conformation adopted by the L-leucine at P3 of gallinamide A, the CH<sub>2</sub>-cyclohexyl moiety of **14** and **17** project towards and interact with L67. The MD simulations showed that residues Q19, G66 and D161 are establishing H-bonds with the backbone of the gallinamide A analogs, which is in agreement with the interactions observed from the X-ray crystal structures<sup>36-38</sup>.



**Figure 4. Representative structures calculated from the MD simulations of cruzain bound to 14 (A) and 17 (B).** 14 and 17 are shown in yellow sticks. The cruzain binding site residues are depicted as sticks and colored according to the per-residue energy contributions ( $\Delta G_{res}$ ). The  $\Delta G_{res}$  values are expressed in kcal/mol. Hydrogen bonds are displayed as green-dashed lines, and the main interacting residues are labeled.

As the electron density corresponding to the P4 moiety of gallinamide A and the analogs was not well-defined (**Fig. 3**), we computationally assessed the dynamics of the P4 fragment along the course of the MD simulations. Note that compounds **14** and **23** have the *N,N*-Me<sub>2</sub>-L-Val group at P4 site while compound **17** contains the 4-(*N*-Me)piperidine group at this position. To describe the changes in the orientation of the P4 moiety within the cruzain binding site, we monitored over time the dihedral angle  $\psi$  (between the equivalent C and CA atoms) of the P4-P3 linkage calculated for each of the three complexes. We observed that the cyclohexyl ring at P3 established more stable interactions with S3 when the P4 position was occupied by *N,N*-Me<sub>2</sub>-L-Val (**Fig. 5A**). In contrast, with 4-(*N*-Me)piperidine at P4, two well-defined populations of the cyclohexyl ring at P3 sampled by  $\psi$  angle were observed. One population bound within S3 while the other one was exposed to the solvent (**Fig. 5B**). These results suggest that the 4-(*N*-Me)piperidine group placed in the P4 position may destabilize interactions of cyclohexyl moiety at S3.

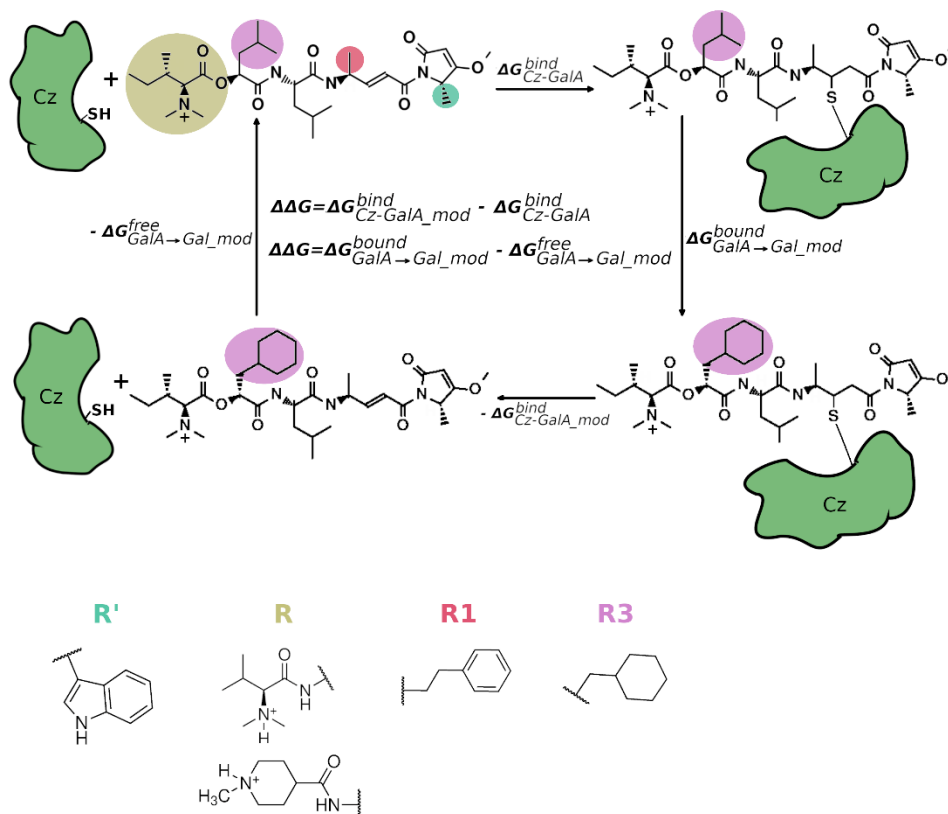


**Figure 5. Binding conformations of 14, 17 and 23 revealed by MD simulations.**  $\Psi$  values of the P3 moiety sampled along the MD simulation trajectory are plotted for cruzain-bound **14** (**A**) and **17** (**B**). Distribution of distances between the center of mass of phenyl and indole rings along the MD simulation trajectory are plotted for **14** (**C**) and **23** (**D**), cruzain-bound (blue) and free in a water box (red). The representative structure of each  $\psi$  angle and each distance population is depicted above the respective peak. Gallinamide A analogs are shown in yellow sticks. The  $\psi$  angles and distances are represented by dashed lines labelled with the corresponding values.

The MD simulations showed that the predicted interactions of the indolyl substituent in **14** and **17** differ from those observed in the co-crystal structure of the **23**-cruzain complex (compare **Fig. 4A** and **4B** with **Fig. 2C**). In the absence of the CH<sub>2</sub>CH<sub>2</sub>-Ph group at P1, the indole ring of **14** and **17** mainly interacts with W184 (**Fig. 5C**) that is consistent with the critical role of W184 in the interactions with these analogs (see energy values of W184 in **Fig. 4**). Molecular dynamic simulations of the **23**-cruzain complex showed that the indole group of **23** may occur in both conformations: (1) interacting with W184 and (2) establishing intramolecular  $\pi$ - $\pi$  stacking interactions with the CH<sub>2</sub>CH<sub>2</sub>-Ph moiety at P1 (**Fig. 5D**). When free in solution, **23** is mostly forming intramolecular  $\pi$ - $\pi$  stacking interactions as observed in the co-crystal structure of the **23**-cruzain complex (**Fig. 5D**). Given that **14** and **23** have the same activity against cruzain (**Table 2**), this might suggest that the gain in binding enthalpy due to intermolecular interactions of the indole ring in **14** is comparable to the entropy loss associated with the intramolecular interactions of the indole ring in **23**.

*Prediction of relative binding affinities of the gallinamide A analogs for cruzain by thermodynamic integration free energy calculations*

Because in each experimentally tested gallinamide A derivative more than one moiety was substituted, we computationally assessed the contribution of each individual modification for the affinity for cruzain. For this purpose, we performed  $\Delta\Delta G_{calc}$  calculations of cruzain-gallinamide A analogs relative to the cruzain-gallinamide A complex, employing TI free energy calculations. The thermodynamic cycle involving the alchemical transformations of the inhibitor free in solution and covalently bound to the enzyme is shown in **Fig. 6**. The results of computation are shown in **Table 4**, where positive values indicate that the native complex possesses higher affinity. We conclude that the individual inclusion of CH<sub>2</sub>-cyclohexyl at R<sub>3</sub>, and CH<sub>2</sub>CH<sub>2</sub>-Ph at R<sub>2</sub> significantly enhanced affinity for cruzain. Addition of the indole ring at R' also increased the compound affinity, albeit not as much as the aforementioned modifications. Conversely, 4-(N-Me)piperidine at P4 resulted in a weaker binding compared to the N,N-Me<sub>2</sub>-L-Ile present in gallinamide A. Finally, the incorporation of N,N-Me<sub>2</sub>-L-Val at P4 does not seem to change the affinity of the derivative. These results are in agreement with the X-ray structures showing ambiguous N,N-Me<sub>2</sub>-L-Val binding (**Fig. 3C, D**) and experimental IC<sub>50</sub> values shown in **Tables 2 and 3**. This suggests that the per-residue free energy decomposition computational approach could be very useful for evaluating new chemical modifications prior to chemical synthesis of the analogs.



**Figure 6. Thermodynamic cycles used to calculate the  $\Delta\Delta G_{calc}$  values for the covalent complexes.** Thermodynamic cycle involves the gallinamide A and analogs free in solution and covalently-bonded to cruzain (in green). Each moiety that was chemically-modified within the gallinamide A structure is colored differentially; corresponding chemical modifications are shown below the cycle diagram. The  $\Delta\Delta G_{calc}$  values were calculated according to the equation depicted inside the cycle. NB: Cz = *T. cruzi* cruzain.

**Table 4. Per-residue free energies differences between the complexes of cruzain-gallinamide A and cruzain-gallinamide A analogs**

Chemically-modified group	$\Delta\Delta G_{Coulomb}^a$ (kcal/mol)	$\Delta\Delta G_{vdw}^a$ (kcal/mol)	$\Delta\Delta G_{calc}^{a,b}$ (kcal/mol)
Indolyl pyrrolidinone at P'	$0.65 \pm 0.11^c$	$-2.79 \pm 0.08$	$-2.13 \pm 0.14$
CH <sub>2</sub> -cyclohexyl at P3	$-2.36 \pm 0.09$	$-0.85 \pm 0.17$	$-3.21 \pm 0.19$
CH <sub>2</sub> CH <sub>2</sub> -Ph at P2	$-2.76 \pm 0.07$	$-0.28 \pm 0.09$	$-3.04 \pm 0.11$
4-( <i>N</i> -Me)piperidine at P4	$1.31 \pm 0.14$	$0.39 \pm 0.11$	$1.70 \pm 0.18$
<i>N,N</i> -Me <sub>2</sub> -L-Val at P4	$-1.03 \pm 0.08$	$0.86 \pm 0.14$	$-0.17 \pm 0.16$

<sup>a</sup> $\Delta\Delta G_X = \Delta G_X(\text{gallinamide A}) - \Delta G_X(\text{derivative})$ , where X stands for Coulomb (charge-related transformations), vdw (van der Waals transformations), and calc (calculated relative free energy).

<sup>b</sup> $\Delta\Delta G_{calc} = \Delta\Delta G_{Coulomb} + \Delta\Delta G_{vdw}$

<sup>c</sup>Final standard error of mean calculated by error propagation

## DISCUSSION

The inhibitory potential of the indolyl-substituted MMP moiety in gallinamide A analogs was first recognized in the studies on the falcipain proteases from *P. falciparum*. The most potent *P. falciparum* falcipain-2 and -3 inhibitor ( $IC_{50}^{\text{falcipain-2}} = 12.0$  nM,  $IC_{50}^{\text{falcipain-3}} = 66.7$  nM,  $EC_{50}^{\text{P. falciparum}} = 9.7$  nM) carried an 2-indolyl-MMP, similar to the analogs **11-23**<sup>21</sup>. One of these analogs was able to cure murine malaria infection in mice at 25 mg/kg and delayed parasitemia at  $\geq 6.25$  mg/kg<sup>27</sup>. We show in this work that gallinamide A and a range of synthetic analogs are also potent inhibitors of *T. cruzi* cruzain ( $IC_{50}$  of 1 nM) and intracellular *T. cruzi* amastigotes ( $EC_{50}$  of 15 nM)<sup>23</sup>.

The co-crystal structures of the gallinamide drug–target complexes determined as part of this work will aid in hit-to-lead optimization of this class of protease inhibitors to improve metabolic stability, solubility and permeability, while maintaining high potency and selectivity. The structures demonstrated that the N-terminal P4 moiety makes loose contacts with cruzain and may be used to improve solubility and permeability of the next generation of the anti-Chagas gallinamide A analogs. Exposure of the P4 moiety in cruzain contrasts to its proposed binding in the S3 pocket of cathepsin L as predicted by molecular docking<sup>23</sup>. The discrepancy between the docking and crystallographic poses may be a docking artifact or the result of non-conservative amino acid substitutions in S3 of cathepsin L affecting pocket shape and electronic properties. Determination of the experimental cathepsin L-gallinamide A structure would be instrumental to resolving this ambiguity and may also have additional applications for the development of SARS-CoV-2 entry inhibitors<sup>24</sup>.

Activity data for the P2 and P3 substituents (**Tables 2, 3**) is consistent with earlier studies by others aimed at the development of cruzain inhibitors. Specifically, bulky groups at P3, such as carboxybenzyl or 3-pyridinyl, are present in known potent irreversible cruzain inhibitors<sup>39–41</sup>. Similarly, the CH<sub>2</sub>-cyclohexyl group at P3 is well tolerated in the gallinamide analogs. The L-phenylalanine and to a lesser extent L-leucine and L-valine are commonly present at P2<sup>42,43</sup>. These groups were embedded within potent gallinamide A analogs used in this study.

SAR trends observed for the P1 site suggest a preference for a center with S-configuration<sup>40,44,45</sup>. This preference was also observed when gallinamide A analogs were tested against cruzain and cathepsin L. D-Phe at the P1 site results in  $\geq 100$ -fold reduction in activity<sup>23</sup>. Inhibitors with a homophenylalanine at P1, which tends to be more resistant to metabolism, showed high potency<sup>40</sup>. The  $\beta$ -branched alkyl substituents, particularly L-Val, were also beneficial for activity<sup>44</sup>.

The most striking observation made in this work is that CH<sub>2</sub>CH<sub>2</sub>-Ph at P1 is involved in intramolecular  $\pi$ - $\pi$  stacking interactions with the aromatic indolyl or biaryl substituents of MMP at P1'. Neither indolyl nor biaryl moieties contact the protein target, suggesting that in the context of **23** and **24** their role is in restricting the flexible inhibitor into its bioactive conformation and minimizing the entropic loss associated with binding to the target. The conformational restriction (rigidification) of a flexible ligand is a commonly used strategy in drug design that enhances potency, improves selectivity and reduces drug metabolism<sup>46</sup>. The same rigidification principle, either via non-covalent intramolecular



interactions or via introducing a covalent linkage, may be explored in designing peptidomimetic inhibitors targeting cathepsin L-like proteases. The gallinamide A and analogs co-crystal structures determined in this work will also have utility in assessing the contributions of the individual substituents for the binding affinity of gallinamide A analogs to Clan CA cysteine proteases. Combining this structural data with the per-residue free energy decomposition approach provides the possibility of validation of new chemical modifications computationally prior to chemical synthesis of future gallinamide A analogs.

## EXPERIMENTAL SECTION

### Chemicals

Gallinamide A and its analogs were synthesized and characterized previously by the HPLC analysis to be >95%.<sup>24,27</sup> Analog numbering is according to<sup>24</sup>. Dulbecco's modified Eagle medium was from Thermo Scientific (Paisley, UK); penicillin–streptomycin from Omega Scientific (Tarzana, USA), fetal bovine serum from Sigma (St. Louis, USA) and benznidazole (*N*-benzyl-2-nitro-1*H*-imidazole-1-acetamide) were from Sigma (St. Louis, USA).

### Zymogen procruzain expression and activation

The pET21a vector (Novagen/EMD) containing the gene sequence encoding the C-terminally truncated His6-tagged zymogen procruzain ( $\Delta$ c; GenBank entry M84342.1) was used for protein expression. Procruzain was expressed in ArcticExpress (DE3) RIL cells (Agilent) and purified as described by Silva *et al*<sup>47</sup>. Procruzain was activated at 0.1 M sodium acetate, pH 5.3, 10 mM EDTA, 0.9 M NaCl, and 5 mM dithiothreitol (DTT). Then DTT was removed using NAP-5 desalting columns (GE Healthcare) and the activation buffer without reducing agents. Covalent reversible inhibitor, S-methyl methanethiosulfonate (MMTS) or gallinamide A, was added to protect the catalytic thiol group of C25 and prevent self-degradation of cruzain. The cruzain-gallinamide A complex was directly used for crystallography studies. Covalently inhibited cruzain tolerated a short-term storage at -80 °C. The detailed purification and activation protocols are provided in Supplementary Information S1.

Modifications to the protocol were made to obtain the cruzain-gallinamide A analog complexes. Following procruzain activation as described above, DTT was removed using HiPrep 26/10 desalting column (GE Healthcare) saturated with the activation buffer with no DTT and supplemented with MMTS. Eluted protein fractions were immediately tested for activity. Fractions containing the MMTS-cruzain complex were immediately pooled, concentrated to ~3 mg/mL using Amicon Ultra centrifugal filter (3 kDa molecular weight cut-off) (Millipore) and loaded onto the size-exclusion Superdex 75 10/300 GL (GE Healthcare) column saturated with 2 mM Bis-Tris, pH 5.8. To prevent cruzain autodegradation, excess MMTS was added to each fraction immediately upon elution from the column. Homogenously purified MMTS-cruzain complex was further concentrated to 5 mg/mL and stored at -80 °C for inhibitor exchange and co-crystallization.

### Enzymatic activity assay

Cruzain activity was measured by monitoring the cleavage of the fluorescent substrate, Z-Phe-Arg-aminomethylcoumarin (Z-FR-AMC), using a Synergy HTX (Biotek) fluorimeter. A 10 mM stock solution of Z-FR-AMC was prepared in dimethyl sulfoxide (DMSO). All assays were performed in black flat-bottom 96-well plates (Costar, catalog 3915), in 200  $\mu$ l of 0.1 M sodium acetate, pH 5.5, supplemented with 1 mM DTT, 0.01% Triton X-100, 0.5 nM cruzain, and 2.5  $\mu$ M of Z-FR-AMC<sup>48</sup>. Prior to addition of the substrate, enzyme was incubated with the test compounds for 10 minutes. Following the substrate addition, fluorescent signal was recorded. Enzymatic activity was calculated from the initial rates of the reaction. Two independent experiments were performed in triplicate. Half-maximal inhibitory concentration (IC<sub>50</sub>) was determined by nonlinear regression analysis of the initial velocity vs. inhibitor concentration plot using Prism<sup>49</sup>. At least eight inhibitor concentrations were used to build each curve. DMSO and trans-Epoxy succinyl-L-leucylamido(4-guanidino)butane (E64) were used as negative and positive controls, respectively.

### Crystallization, data collection and structure determination

For cruzain-gallinamide A complex, gallinamide A-bound cruzain was concentrated to 10 mg/mL and the activation buffer was exchanged to 2 mM Bis-Tris pH 5.8 using a centrifugal filter with molecular weight cut-off of 3 kDa (Millipore). For cruzain-gallinamide A analog complexes, 1.2 molar excess of the analog in DMSO was added to cruzain-MMTS complex followed by adding 10-molar excess of DTT. The reaction mix was incubated for 60 minutes at room temperature for inhibitor exchange. Screening of crystallization conditions was performed using commercial high-throughput screening kits available in deep-well format from Hampton Research (Aliso Viejo, CA) or Qiagen (Germantown, MD), a nanoliter drop-setting Mosquito robot (TTP LabTech, Melbourn, UK) operating with 96-well plates, and a hanging drop crystallization protocol. For diffraction quality, crystals were further optimized in 96-well plates configured using the Dragonfly robot (TTP LabTech, Melbourn, UK) and the Designer software (TTP LabTech, Melbourn, UK). Crystals were obtained at 23 °C from the conditions specified in **Table 1**.

Diffraction data were collected remotely at beamline 8.3.1, Advanced Light Source, Lawrence Berkeley National Laboratory. Data indexing, integration, and scaling were conducted using XDS<sup>50</sup>. Cruzain structure (PDB ID 3KKU)<sup>51</sup> was used as a molecular replacement model. The PHASER and REFMAC5 modules of the CCP4 software<sup>52</sup> suite were used to solve and refine the structure. COOT software<sup>53</sup> was used for the model building. Data collection and refinement statistics are shown in **Table 1**.

### *T. cruzi* cell-based assay

Antiparasitic activity and cytotoxicity of compounds were determined as described by Boudreau *et al*<sup>23</sup>. Mouse myoblasts, cell line C2C12 (ATCC CRL-1772), were maintained in Dulbecco's modified Eagle medium (DMEM - Invitrogen) containing 4.5 g/l glucose, supplemented with 1% penicillin-streptomycin 10,000 U/mL (Invitrogen) and 5% fetal bovine serum (Sigma) at 37 °C with 5% CO<sub>2</sub>. *T. cruzi* CA-1/72 strain was maintained by

weekly coinfection with C2C12 cells. To assess the anti-parasitic activity of the compounds, cells and parasites were seeded in 384-well black clear-bottom plates at  $1 \times 10^5$  parasites/mL and  $2 \times 10^4$  C2C12/mL density in 50  $\mu$ L of DMEM media per well. Compounds were evaluated at 10 serially diluted concentrations (from 600 nM to 1.2 nM). Plates were then fixed with 4% formaldehyde for at least 2 h and stained with 0.5  $\mu$ g/mL of 4',6-diamidino-2-phenylindole (DAPI) for at least 4 hours prior to reading. Plates were imaged and analyzed with the automated ImageXpress MicroXL microscope (Molecular Devices). Infection levels (parasites per host cell) were normalized to positive control (50  $\mu$ M benznidazole) and negative control (DMSO). EC<sub>50</sub> (antiparasitic activity) and CC<sub>50</sub> (host cell cytotoxicity) values were determined from the dose-response curves using Prism software (GraphPad Software, La Jolla, CA)<sup>49</sup>. Two independent assays were performed, each in duplicate.

### Parametrization of gallinamide A and analogs for computational studies

To model the covalently bound equivalents, the enone group of each compound was reduced to the *trans*-isomer. Further, the compounds were subjected to geometry optimization with single-point (SP) calculations using Gaussian 09 package<sup>54</sup> with Hartree-Fock level of 163 theory, 6-31G(d) basis and Merz-Kollman (MK) scheme<sup>55</sup>. To obtain the atom-centered partial charges, the electrostatic potential (ESP) derived from quantum mechanical calculations were fitted using the RESP algorithm implemented in Antechamber20<sup>56</sup>. The compounds parameters were obtained from the Generalized Amber Force-Field 2 (GAFF2) force field<sup>56</sup>. Finally, the charges obtained through RESP fits underwent certain modifications, so that all equivalent atoms outside the thermodynamic integration (TI) regions across the studied compounds had the same charge, and those involved in alchemical transformations had a 0 or +1 net charge.

### MD simulations

The starting structure of cruzain-inhibitor complexes were obtained by transforming gallinamide A in the co-crystal structure determined in this work into each individual analog using Avogadro program<sup>57</sup>. Protonation states of cruzain residues were determined at pH=5.5 using the PDB2PQR server<sup>58</sup>. Systems setup was performed with tleap program from AmberTools20<sup>56</sup>, and a covalent bond was built between the appropriate C atom of enone moiety of each ligand to the S atom of the catalytic cysteine. AMBER14SB force field (ff14SB)<sup>59</sup> was employed for protein parameters. All complexes were solvated with explicit TIP3P water molecules<sup>60</sup> in an octahedral box extending at least 10 Å from the solute surface. Systems were neutralized by replacing water molecules with Na<sup>+</sup> and Cl<sup>-</sup> counterions, depending on their net charges.

All simulations were conducted with pmemd.cuda of Amber20<sup>56</sup>. 50,000 energy minimization (EM) steps were performed using a combination of steepest descent and conjugate gradient procedures. The equilibration procedure was carried out in two sequential steps, i.e., the NVT and NPT ensembles. During the NVT equilibration, the heating was performed employing a linear temperature gradient from 10 to 298 K. The subsequent NPT equilibration was performed at a constant temperature of 298 K. Both equilibration steps were performed for 500 ps with the solute heavy atoms restrained with a 10.0 kcal·mol<sup>-1</sup>·Å<sup>-2</sup> restraint constant. The Berendsen barostat and thermostat were

employed in each equilibration step<sup>61</sup>. Finally, the non-constrained MD simulation was performed at constant pressure and temperature (1 atm and 298 K, respectively), employing the Berendsen barostat<sup>61</sup> and the Langevin thermostat<sup>62</sup>, respectively. The particle mesh Ewald (PME) method was used to handle long-range electrostatic interactions<sup>63</sup> and a distance cutoff of 10 Å was employed. The MD integration step was set to 2 fs and one snapshot was sampled every 10,000 steps. Three replicas of 200 ns were simulated in all cruzain-compound complexes by assigning different random velocities to the systems' atoms during the respective heating steps.

### Trajectory analyses

The cpptraj program of AmberTools20 package<sup>56</sup> was employed to analyze all MD trajectories and also for calculating the  $\psi$  angle (formed between the planes defined by O-C-CA-CB atoms) which describes the dynamics of P3 and P4 positions of gallinamide A analogs. Hydrogen bonds established between gallinamide A analogs and cruzain were calculated with the default geometric definition of cpptraj, i.e., a distance cutoff  $\leq 3.0$  Å between acceptor and donor atom, and the acceptor-hydrogen-donor angle  $\geq 135^\circ$ <sup>56</sup>.

### Per-residue free energy decomposition with Molecular Mechanics Generalized Born Surface Area (MM-GBSA)

A per-residue effective free energy decomposition was carried out in order to determine the most important residues involved in cruzain-ligand interactions<sup>64,65</sup>. Effective binding free energies ( $\Delta G_{eff}$ ), which do not include the contribution of configurational entropy, were conducted for complexes of cruzain with the top hits of gallinamide A analogs using MMPBSA.py<sup>66</sup>. The single-trajectory approximation was used for this calculation, employing the MD simulations of aforementioned complexes. The GB-neck2 model (igb=8) with mbondi3 radii was used for estimating the polar solvation energy ( $\Delta G_{GB}$ )<sup>56</sup>. A salt concentration of 0.1 M and the default dielectric constant value ( $\epsilon=1$ ) were set. A per-residue effective free energy decomposition was carried out in order to determine the most important residues involved in cruzain-ligand interactions<sup>64,66</sup>. Energetically-relevant residues, i.e., hot-spots, at the interfaces of the studied complexes were predicted using the energy decomposition protocol implemented in MMPBSA.py program.

### TI Free Energy Calculations

To assess the contribution of each chemical modification to the affinity for cruzain, we perform rigorous alchemical transformation and relative free energy calculations ( $\Delta\Delta G_{calc}$ ). For this purpose, the new chemical modifications present in three best hits (compounds **14**, **17**, and **23** from the series of 2-indolyl-MMP analogs) were independently analyzed. TI  $\Delta\Delta G_{calc}$  calculations were performed using the thermodynamic cycle reported by Hernandez-Gonzalez *et al.* for covalent ligands of cysteine proteases<sup>67</sup>. The resulting binding free energy of gallinamide A and analogs was derived from the sum of the various steps in which each ligand is desolvated and subsequently introduced into the cruzain binding site. The calculations were performed for each ligand in solution and within the enzyme binding site. Simulations of cruzain complexes and solvated free ligands were prepared in the same manner as the conventional MD simulations. The representative structure of each studied complex,

obtained from the RMSD-clustering analysis of the equilibrium MD simulations, was selected for TI free energy calculations.

The discharge, charge, and van der Waals steps of the ligand in solution and in complex form were slowly decoupled along the simulation by changing the  $\lambda$ -parameter from 0 to 1 with a stride of 0.05. A total of 21 windows were used for the decoupling process. For each window, a short EM, a 1 ns NVT heating process followed by a 12 ns production run was performed in the NPT ensemble. The pressure in the NPT MD simulations was controlled by means of the Monte Carlo barostat<sup>56</sup>. The timestep used for all simulations was 2 fs and the PME algorithm was used to treat long-range electrostatic interactions with a cut-off of 10 Å. All MD simulations were performed using pmemd.cuda from Amber20<sup>56</sup>. The results were analyzed using the `alchemical_analysis.py` python tool<sup>68</sup>. Charge-related alchemical transformations involving charged moieties were conducted following a single box/dual system approach in order to keep the simulation box neutral at all  $\lambda$  values<sup>69</sup>. The protein-ligand complex and the free ligand were placed in an 73x73x78 Å cuboid box with a minimal distance of 16 Å between. Their interaction was prevented by applying a harmonic restraint ( $k=50 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$ ) to the C $\alpha$  atom of A30 in cruzain and the CA atom of the moiety placed at P2 position in the free ligand. Errors were propagated to calculate the total standard error of means (SEMs) of the final  $\Delta\Delta G_{calc}$  values.

## ASSOCIATED CONTENT

### Supporting Information

The detailed protocols for cruzain expression, purification and activation are provided in the Supporting Information S1.

### Accession Codes

Coordinates and structure factors of the cruzain-gallinamide complexes are available in the Protein Data Bank (PDB) under accession codes 7JUU, 7S19 and 7S18. Authors will release the atomic coordinates and experimental data upon article publication.

## AUTHOR INFORMATION

### Corresponding Authors

Larissa M. Podust

Skaggs School of Pharmacy and Pharmaceutical Sciences, Center for Discovery and Innovation in Parasitic Diseases, University of California San Diego, La Jolla, California, United States

E-mail address: [lpodust@health.ucsd.edu](mailto:lpodust@health.ucsd.edu)

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGEMENTS

We thank the staff members of beamline 8.3.1, James Holton, George Meigs and Kathryn Burnett, at the Advanced Light Source at Lawrence Berkeley National Laboratory, for assistance with data collection; Jair Lage Siqueira-Neto for access to the BSL2 instrumentation and facilities.

## FUNDING

This work was supported in part by the UCSD start-up fund to L.M.P., Sao Paulo Research Foundation Agency (FAPESP) (2018/25311-2 and 2018/03911-8 to L.H.A.), the National Health and Medical Research Council of Australia (APP1174941 to R.J.P.), National Institute of Health (R21AI27505 to W.H.G.).

## ABBREVIATIONS USED

Benzimidazole, *N*-benzyl-2-nitro-1*H*-imidazole-1-acetamide; DAPI, 4',6-diamidino-2-phenylindole; DMEM, Dulbecco's modified Eagle medium; DMSO, dimethyl sulfoxide; DTT, dithiothreitol; EM, energy minimization; ESP, electrostatic potential; E-64, *trans*-epoxysuccinyl-L-leucylamido(4-guanidino)butane; ff14SB, AMBER14SB force field; GAFF2, Generalized Amber Force-Field 2; IC<sub>50</sub>, half-maximal inhibitory concentration; MD, molecular dynamics; MK, Merz-Kollman; MM-GBSA, Molecular Mechanics Generalized Born Surface Area; MMP, methylmethoxypyrrolidinone; MMTS, *S*-methyl methanethiosulfonate; 4-(*N*-Me)piperidine, 4-(*N*-methyl)piperidine; *N,N*-Me<sub>2</sub>-L-Ile, *N,N*-dimethylisoleucine; *N,N*-Me<sub>2</sub>-L-Val, *N,N*-dimethylvaline; PDB, Protein Data Bank; PME, particle mesh Ewald; SEMs, standard error of means; SI, selectivity index; SP, single-point; TI, thermodynamic integrations;  $\Delta\Delta G_{calc}$ , relative free energy calculations;  $\Delta G_{eff}$ , effective binding free energies;  $\Delta G_{GB}$ , polar solvation energy;  $\Delta G_{res}$ , per-residue free energy decomposition; Z-FR-AMC, Z-Phe-Arg-aminomethylcoumarin.

## REFERENCES

- (1) WHO. Chagas disease (American trypanosomiasis) <https://www.who.int/chagas/epidemiology/en/> (accessed Mar 11, 2021).
- (2) Pérez-Molina, J. A.; Molina, I. Chagas Disease. *Lancet* **2018**, *391* (10115), 82–94. [https://doi.org/10.1016/S0140-6736\(17\)31612-4](https://doi.org/10.1016/S0140-6736(17)31612-4).
- (3) Franco-Paredes, C.; Villamil-Gómez, W. E.; Schultz, J.; Henao-Martínez, A. F.; Parra-Henao, G.; Rassi, A.; Rodríguez-Morales, A. J.; Suarez, J. A. A Deadly Feast: Elucidating the Burden of Orally Acquired Acute Chagas Disease in Latin America – Public Health and Travel Medicine Importance. *Travel Med. Infect. Dis.* **2020**, *36* (January), 101565. <https://doi.org/10.1016/j.tmaid.2020.101565>.
- (4) Martins-Melo, F. R.; Carneiro, M.; Ribeiro, A. L. P.; Bezerra, J. M. T.; Werneck, G. L. Burden of Chagas Disease in Brazil, 1990–2016: Findings from the Global Burden of Disease Study 2016. *Int. J. Parasitol.* **2019**, *49* (3–4), 301–310. <https://doi.org/10.1016/j.ijpara.2018.11.008>.
- (5) Kalil-Filho, R. Globalization of Chagas Disease Burden and New Treatment Perspectives. *J. Am. Coll. Cardiol.* **2015**, *66* (10), 1190–1192. <https://doi.org/10.1016/j.jacc.2015.07.024>.
- (6) Stanaway, J. D.; Roth, G. The Burden of Chagas Disease: Estimates and Challenges. *Glob. Heart* **2015**, *10* (3), 139. <https://doi.org/10.1016/j.gheart.2015.06.001>.
- (7) Lee, B. Y.; Bacon, K. M.; Bottazzi, M. E.; Hotez, P. J. Global Economic Burden of Chagas Disease: A Computational Simulation Model. *Lancet Infect. Dis.* **2013**, *13* (4), 342–348. [https://doi.org/10.1016/S1473-3099\(13\)70002-1](https://doi.org/10.1016/S1473-3099(13)70002-1).
- (8) FDA. U.S. FDA approves Chemo Group’s benznidazole to treat children with Chagas disease <https://dndi.org/press-releases/2017/fda-approves-benznidazole-chagas-children/> (accessed Feb 12, 2021).
- (9) Administration, U. S. F. & D. FDA approves first U.S. treatment for Chagas disease <https://www.fda.gov/news-events/press-announcements/fda-approves-first-us-treatment-chagas-disease>.
- (10) Administration, U. S. F. & D. FDA Approves LAMPIT (Nifurtimox) for the Treatment of Chagas Disease in Pediatric Patients [https://www.ascpt.org/Resources/ASCPT-News/View/articleid/25834?dnnprintmode=true&mid=16095&SkinSrc=\[G\]Skins%2F\\_default%2FNo+Skin&ContainerSrc=\[G\]Containers%2F\\_default%2FNo+Container](https://www.ascpt.org/Resources/ASCPT-News/View/articleid/25834?dnnprintmode=true&mid=16095&SkinSrc=[G]Skins%2F_default%2FNo+Skin&ContainerSrc=[G]Containers%2F_default%2FNo+Container).
- (11) Bermudez, J.; Davies, C.; Simonazzi, A.; Pablo, J.; Palma, S. Acta Tropica Current Drug Therapy and Pharmaceutical Challenges for Chagas Disease. *Acta Trop.* **2016**, *156*, 1–16. <https://doi.org/10.1016/j.actatropica.2015.12.017>.
- (12) Izumi, E.; Ueda-Nakamura, T.; Dias Filho, B. P.; Veiga Júnior, V. F.; Nakamura, C.

- V. Natural Products and Chagas' Disease: A Review of Plant Compounds Studied for Activity against *Trypanosoma Cruzi*. *Nat. Prod. Rep.* **2011**, *28* (4), 809–823. <https://doi.org/10.1039/c0np00069h>.
- (13) de Morais, M. C.; de Souza, J. V.; da Silva Maia Bezerra Filho, C.; Dolabella, S. S.; de Sousa, D. P. Trypanocidal Essential Oils: A Review. *Molecules* **2020**, *25* (19), 1–19. <https://doi.org/10.3390/molecules25194568>.
- (14) Barros, L. M.; Duarte, A. E.; Morais-Braga, M. F. B.; Waczuk, E. P.; Vega, C.; Leite, N. F.; De Menezes, I. R. A.; Coutinho, H. D. M.; Rocha, J. B. T.; Kamdem, J. P. Chemical Characterization and Trypanocidal, Leishmanicidal and Cytotoxicity Potential of *Lantana Camara L.* (Verbenaceae) Essential Oil. *Molecules* **2016**, *21* (2). <https://doi.org/10.3390/molecules21020209>.
- (15) Demay, J.; Bernard, C.; Reinhardt, A.; Marie, B. Natural Products from Cyanobacteria: Focus on Beneficial Activities. *Mar. Drugs* **2019**, *17* (6), 1–49. <https://doi.org/10.3390/md17060320>.
- (16) Niedermeyer, T. H. ors. J. Anti-Infective Natural Products from Cyanobacteria. *Planta Med.* **2015**, *81* (15), 1309–1325. <https://doi.org/10.1055/s-0035-1546055>.
- (17) Shishido, T. K.; Popin, R. V.; Jokela, J.; Wahlsten, M.; Fiore, M. F.; Fewer, D. P.; Herfindal, L.; Sivonen, K. Dereplication of Natural Products with Antimicrobial and Anticancer Activity from Brazilian Cyanobacteria. *Toxins (Basel)*. **2019**, *12* (1), 1–17. <https://doi.org/10.3390/toxins12010012>.
- (18) Uzair, B.; Tabassum, S.; Rasheed, M.; Rehman, S. F. Exploring Marine Cyanobacteria for Lead Compounds of Pharmaceutical Importance. *Sci. World J.* **2012**, *2012*, 1–10. <https://doi.org/10.1100/2012/179782>.
- (19) Linington, R. G.; Clark, B. R.; Trimble, E. E.; Almanza, A.; Ureña, L.-D.; Kyle, D. E.; Gerwick, W. H. Antimalarial Peptides from Marine Cyanobacteria: Isolation and Structural Elucidation of Gallinamide A. *J. Nat. Prod.* **2009**, *72* (1), 14–17. <https://doi.org/10.1021/np8003529>.
- (20) Stolze, S. C.; Deu, E.; Kaschani, F.; Li, N.; Florea, B. I.; Richau, K. H.; Colby, T.; Van Der Hoorn, R. A. L.; Overkleeft, H. S.; Bogoyo, M.; Kaiser, M. The Antimalarial Natural Product Symplostatin 4 Is a Nanomolar Inhibitor of the Food Vacuole Falcipains. *Chem. Biol.* **2012**, *19* (12), 1546–1555. <https://doi.org/10.1016/j.chembiol.2012.09.020>.
- (21) Conroy, T.; Guo, J. T.; Elias, N.; Cergol, K. M.; Gut, J.; Legac, J.; Khatoon, L.; Liu, Y.; McGowan, S.; Rosenthal, P. J.; Hunt, N. H.; Payne, R. J. Synthesis of Gallinamide A Analogues as Potent Falcipain Inhibitors and Antimalarials. *J. Med. Chem.* **2014**, *57* (24), 10557–10563. <https://doi.org/10.1021/jm501439w>.
- (22) Miller, B.; Friedman, A. J.; Choi, H.; Hogan, J.; McCammon, J. A.; Hook, V.; Gerwick, W. H. The Marine Cyanobacterial Metabolite Gallinamide A Is a Potent and Selective Inhibitor of Human Cathepsin L. *J. Nat. Prod.* **2014**, *77* (1), 92–99. <https://doi.org/10.1021/np400727r>.



- (23) Boudreau, P. D.; Miller, B. W.; McCall, L. I.; Almaliti, J.; Reher, R.; Hirata, K.; Le, T.; Siqueira-Neto, J. L.; Hook, V.; Gerwick, W. H. Design of Gallinamide A Analogs as Potent Inhibitors of the Cysteine Proteases Human Cathepsin L and Trypanosoma Cruzi Cruzain. *J. Med. Chem.* **2019**, *62* (20), 9026–9044. <https://doi.org/10.1021/acs.jmedchem.9b00294>.
- (24) Ashhurst, A. S.; Tang, A. H.; Fajtová, P.; Yoon, M. C.; Aggarwal, A.; Bedding, M. J.; Stoye, A.; Beretta, L.; Pwee, D.; Drelich, A.; Skinner, D.; Li, L.; Meek, T. D.; McKerrow, J. H.; Hook, V.; Tseng, C.-T.; Larance, M.; Turville, S.; Gerwick, W. H.; O'Donoghue, A. J.; Payne, R. J. Potent Anti-SARS-CoV-2 Activity by the Natural Product Gallinamide A and Analogues via Inhibition of Cathepsin L. *J. Med. Chem.* **2021**. <https://doi.org/10.1021/acs.jmedchem.1c01494>.
- (25) Conroy, T.; Guo, J. T.; Hunt, N. H.; Payne, R. J. Total Synthesis and Antimalarial Activity of Symplostatin 4. *Org. Lett.* **2010**, *12* (23), 5576–5579. <https://doi.org/10.1021/ol1024663>.
- (26) Conroy, T.; Guo, J. T.; Linington, R. G.; Hunt, N. H.; Payne, R. J. Total Synthesis, Stereochemical Assignment, and Antimalarial Activity of Gallinamide A. *Chem. - A Eur. J.* **2011**, *17* (48), 13544–13552. <https://doi.org/10.1002/chem.201102538>.
- (27) Stoye, A.; Juillard, A.; Tang, A. H.; Legac, J.; Gut, J.; White, K. L.; Charman, S. A.; Rosenthal, P. J.; Grau, G. E. R.; Hunt, N. H.; Payne, R. J. Falcipain Inhibitors Based on the Natural Product Gallinamide A Are Potent in Vitro and in Vivo Antimalarials. *J. Med. Chem.* **2019**, *62* (11), 5562–5578. <https://doi.org/10.1021/acs.jmedchem.9b00504>.
- (28) Toman, N. P.; Kamenik, A. S.; Santos, L. H.; Hofer, F.; Liedl, K. R.; Ferreira, R. S. Profiling Selectivity of Chagasin Mutants towards Cysteine Proteases Cruzain or Cathepsin L through Molecular Dynamics Simulations. *J. Biomol. Struct. Dyn.* **2020**, *0* (0), 1–13. <https://doi.org/10.1080/07391102.2020.1796797>.
- (29) Vieira, R. P.; Santos, V. C.; Ferreira, R. S. Structure-Based Approaches Targeting Parasite Cysteine Proteases. *Curr. Med. Chem.* **2017**, *25*, 1–18. <https://doi.org/http://dx.doi.org/10.2174/0929867324666170810165302>.
- (30) Cazzulo, J. J.; Cazzulo Franke, M. C.; Martínez, J.; Franke de Cazzulo, B. M. Some Kinetic Properties of a Cysteine Proteinase (Cruzipain) from Trypanosoma Cruzi. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* **1990**, *1037* (2), 186–191. [https://doi.org/10.1016/0167-4838\(90\)90166-D](https://doi.org/10.1016/0167-4838(90)90166-D).
- (31) Eakin, a E.; Mills, a a; Harth, G.; McKerrow, J. H.; Craik, C. S. The Sequence, Organization, and Expression of the Major Cysteine Protease (Cruzain) from Trypanosoma Cruzi. *J. Biol. Chem.* **1992**, *267* (11), 7411–7420.
- (32) Engel, B. J. C.; Doyle, P. S.; Hsieh, I.; Mckerrow, J. H. Cysteine Protease Inhibitors Cure an Experimental. **1998**, *188* (4), 725–734.
- (33) Bontempi, E.; Martinez, J.; Cazzulo, J. J. Subcellular Localization of a Cysteine Proteinase from Trypanosoma Cruzi. *Mol. Biochem. Parasitol.* **1989**, *33* (1), 43–47. [https://doi.org/10.1016/0166-6851\(89\)90040-6](https://doi.org/10.1016/0166-6851(89)90040-6).

- (34) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27* (2), 157–162. [https://doi.org/10.1016/S0006-291X\(67\)80055-X](https://doi.org/10.1016/S0006-291X(67)80055-X).
- (35) Omotuyi, O. I. Methyl-Methoxypyrrolinone and Flavinium Nucleus Binding Signatures on Falcipain-2 Active Site. *J. Mol. Model.* **2014**, *20* (8). <https://doi.org/10.1007/s00894-014-2386-2>.
- (36) Bryant, C.; Kerr, I. D.; Debnath, M.; Ang, K. K. H.; Ratnam, J.; Ferreira, R. S.; Jaishankar, P.; Zhao, D.; 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. <https://doi.org/10.1016/j.bmcl.2009.08.098>.
- (37) 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. <https://doi.org/10.1074/jbc.M109.014340>.
- (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) Roush, W. R.; Cheng, J.; Knapp-reed, B.; Alvarez-hernandez, A.; Mckerrow, J. H.; Hansell, E.; Engel, J. C. Potent Second Generation Vinyl Sulfonamide Inhibitors of the Trypanosomal Cysteine Protease Cruzain. **2001**, *11*, 2759–2762.
- (40) Huang, L.; Lee, A.; Ellman, J. A. Identification of Potent and Selective Mechanism-Based Inhibitors of the Cysteine Protease Cruzain Using Solid-Phase Parallel Synthesis. *J. Med. Chem.* **2002**, *45* (3), 676–684. <https://doi.org/10.1021/jm010333m>.
- (41) Rocha, D. A.; Silva, E. B.; Fortes, I. S.; Lopes, M. S.; Ferreira, R. S.; Andrade, S. F. Synthesis and Structure-Activity Relationship Studies of Cruzain and Rhodesain Inhibitors. *Eur. J. Med. Chem.* **2018**, *157*, 1426–1459. <https://doi.org/10.1016/j.ejmech.2018.08.079>.
- (42) 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. <https://doi.org/10.1002/pro.5560060801>.
- (43) da Silva, E. B.; do Nascimento Pereira, G. A.; Ferreira, R. S. Trypanosomal Cysteine Peptidases: Target Validation and Drug Design Strategies. In *Comprehensive Analysis of Parasite Biology: From Metabolism to Drug Discovery*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2016; pp 121–145. <https://doi.org/10.1002/9783527694082.ch5>.
- (44) 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  $\alpha$ -Keto-Based

- Inhibitors of Cruzain, a Cysteine Protease Implicated in Chagas Disease. *Bioorg. Med. Chem.* **2005**, 13 (6), 2141–2156. <https://doi.org/10.1016/j.bmc.2004.12.053>.
- (45) Roush, W. R.; González, F. V.; McKerrow, J. H.; Hansell, E. Design and Synthesis of Dipeptidyl A',B'-Epoxy Ketones, Potent Irreversible Inhibitors of the Cysteine Protease Cruzain. *Bioorg. Med. Chem. Lett.* **1998**, 8 (19), 2809–2812. [https://doi.org/10.1016/S0960-894X\(98\)00494-6](https://doi.org/10.1016/S0960-894X(98)00494-6).
- (46) Fang, Z.; Zhan, P.; Zhang, Q.; Liu, X. Conformational Restriction: An Effective Tactic in 'Follow-on'-Based Drug Discovery. *Futur. Med. Chem.* **2014**, 6 (8), 885–901. <https://doi.org/10.4155/fmc.14.50>.
- (47) Barbosa Da Silva, E.; Dall, E.; Briza, P.; Brandstetter, H.; Ferreira, R. S. Cruzain Structures: Apocruzain and Cruzain Bound to S-Methyl Thiomethanesulfonate and Implications for Drug Design. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2019**, 75 (6), 419–427. <https://doi.org/10.1107/S2053230X19006320>.
- (48) Pereira, G. A. N.; da Silva, E. B.; Braga, S. F. P.; Leite, P. G.; Martins, L. C.; Vieira, R. P.; Soh, W. T.; Villela, F. S.; Costa, F. M. R.; Ray, D.; de Andrade, S. F.; Brandstetter, H.; Oliveira, R. B.; Caffrey, C. R.; Machado, F. S.; Ferreira, R. S. Discovery and Characterization of Trypanocidal Cysteine Protease Inhibitors from the 'Malaria Box.' *Eur. J. Med. Chem.* **2019**, 179, 765–778. <https://doi.org/10.1016/j.ejmech.2019.06.062>.
- (49) One-Way ANOVA Followed by Dunnett's Multiple Comparisons Test Was Performed Using GraphPad Prism Version 6.00 for Windows. La Jolla, California, USA.
- (50) Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, 66 (2), 125–132. <https://doi.org/10.1107/S0907444909047337>.
- (51) 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. <https://doi.org/10.1021/jm100488w>.
- (52) Collaborative Computational Project, N. 4. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1994**, 50 (5), 760–763. <https://doi.org/10.1107/S0907444994003112>.
- (53) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, 60 (12 I), 2126–2132. <https://doi.org/10.1107/S0907444904019158>.
- (54) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ort, and D. J. F. Gaussian 09, Gaussian Inc. Wallingford CT 2009.
- (55) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from

- Semiempirical Methods. *J. Comput. Chem.* **1990**, *11* (4), 431–439. <https://doi.org/10.1002/jcc.540110404>.
- (56) D.A. Case, K.B., I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kur, D. M. Y. and P. A. K. AMBER 2020. University of California San Francisco 2020.
- (57) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminform.* **2012**, *4* (1), 17. <https://doi.org/10.1186/1758-2946-4-17>.
- (58) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32* (Web Server), W665–W667. <https://doi.org/10.1093/nar/gkh381>.
- (59) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (60) Jorgensen, W. L.; Jenson, C. Temperature Dependence of TIP3P, SPC, and TIP4P Water from NPT Monte Carlo Simulations: Seeking Temperatures of Maximum Density. *J. Comput. Chem.* **1998**, *19* (10), 1179–1186. [https://doi.org/10.1002/\(SICI\)1096-987X\(19980730\)19:10<1179::AID-JCC6>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1096-987X(19980730)19:10<1179::AID-JCC6>3.0.CO;2-J).
- (61) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. <https://doi.org/10.1063/1.448118>.
- (62) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103* (11), 4613–4621. <https://doi.org/10.1063/1.470648>.
- (63) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An  $N \cdot \log(N)$  Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. <https://doi.org/10.1063/1.464397>.
- (64) Zoete, V.; Michielin, O. Comparison between Computational Alanine Scanning and Per-Residue Binding Free Energy Decomposition for Protein-Protein Association Using MM-GBSA: Application to the TCR-p-MHC Complex. *Proteins Struct. Funct. Bioinforma.* **2007**, *67* (4), 1026–1047. <https://doi.org/10.1002/prot.21395>.
- (65) Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein-Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras-Raf and Ras-RalGDS Complexes. *J. Mol. Biol.* **2003**, *330* (4), 891–913. [https://doi.org/10.1016/S0022-2836\(03\)00610-7](https://doi.org/10.1016/S0022-2836(03)00610-7).

- (66) 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. <https://doi.org/10.1021/ct300418h>.
- (67) Hernández González, J. E.; Hernández Alvarez, L.; Leite, V. B. P.; Pascutti, P. G. Water Bridges Play a Key Role in Affinity and Selectivity for Malarial Protease Falcipain-2. *J. Chem. Inf. Model.* **2020**, *60* (11), 5499–5512. <https://doi.org/10.1021/acs.jcim.0c00294>.
- (68) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the Analysis of Free Energy Calculations. *J. Comput. Aided. Mol. Des.* **2015**, *29* (5), 397–411. <https://doi.org/10.1007/s10822-015-9840-9>.
- (69) Gapsys, V.; Michielssens, S.; Peters, J. H.; de Groot, B. L.; Leonov, H. Calculation of Binding Free Energies; 2015; pp 173–209. [https://doi.org/10.1007/978-1-4939-1465-4\\_9](https://doi.org/10.1007/978-1-4939-1465-4_9).

## Table of Contents graphic

