# *Article*

# An *in silico* Study of Benzophenone Derivatives as Potential Non-Competitive Inhibitors of *Trypanosoma cruzi* and *Leishmania Amazonensis* Cysteine Proteinases

Poliany G. Freitas,*[,a] Thiago E. Castilho,[a] Leticia de Almeida,[b] Claudia M. Maciel-Rezende,[c] Luciano T. Costa,[d] Claudio Viegas Junior,[c] Marcos J. Marques,[b] Marcelo H. dos Santos[e] and Nelson J. F. da Silveira[a]

[a]*Laboratório de Modelagem Molecular e Simulação Computacional, Instituto de Química,*
[b]*Laboratório de Biologia Molecular de Microorganismos, Instituto de Ciências Biomédicas,*
[c]*Laboratório de Fitoquímica e Química Medicinal, Instituto de Química,*
*Universidade Federal de Alfenas, Rua Gabriel Monteiro da Silva, 700, 37130-000 Alfenas-MG, Brazil*

[d]*Instituto de Química, Departamento de Química Física, Universidade Federal Fluminense,*
*Outeiro de São João Batista, 24020-150 Niterói-RJ, Brazil*

[e]*Departamento de Química, Universidade Federal de Viçosa, Avenida PH Rolfs, 36571-000 Viçosa-MG, Brazil*

This study investigates the mechanisms of interaction between benzophenone derivatives and cruzain and Llacys1 (the protein expressed by cysteine protease gene isoform 1 of *L. amazonensis*) by homology modelling, docking and molecular dynamics simulation. The results predict that the same binding site in cruzain and Llacys1 is involved in complexes with benzophenone derivatives that cause non-competitive inhibition of the enzymes. The Gln residue is conserved among the enzymes, and is shown to be a key residue in the allosteric site of these cysteine proteases and in the interaction with benzophenone derivatives. The binding free energies highlight that the main energetic term contributing to the cruzain- and Llacys1-benzophenone compound interactions is the van der Waals term. Experimental results showed that benzophenone derivatives are promising potential inhibitors of cysteine proteases. Moreover, we found that two benzophenone derivatives are the most effective inhibitors of cruzain and *L. amazonensis* cysteine protease.

**Keywords:** benzophenone, *Trypanosoma cruzi*, *L. amazonensis*, cysteine protease, computational methods

## Introduction

Neglected tropical diseases (NTDs) are characterized as a wide range of infectious diseases that remain a leading cause of morbidity and mortality in the developing world and have a clear economic burden on the affected countries.[1-3] Despite the global morbidity of those diseases, only 1.3% of the 1556 drugs approved for their treatment was specifically developed to address NTDs.[4] This situation reflects the lack of interest among pharmaceutical companies in developing drugs to treat NTDs. Despite the high demand for new treatments, the potential consumers have no money to pay for medicines. Among the 17 medical conditions listed by the World Health Organization (WHO) as NTDs, there are two Trypanosomiasis conditions: leishmaniasis and Chagas disease.[3]

Classified as one of the most neglected diseases, leishmaniasis comprises several clinical manifestations caused by different species of the protozoan parasites from the genus *Leishmania*.[5,6] This complex disease has emerged in 98 countries and has a high rate of incidence, affecting 12 million people worldwide, with 350 million people at risk of infection.[7] Pentavalent antimonials remain the first line of treatment for all clinical forms of leishmaniasis, despite the variable therapeutic response and growing concerns about treatment failure and drug administration by parents.[8] Should these treatments fail, a number of other drugs may be employed, depending upon the species of *Leishmania* concerned and the resources available to the health professionals involved. Recommended secondary treatments employ a variety of drugs such as amphotericin B, pentamidine, paromomycin and miltefosine.[9] Some new strategies involve newer formulations of amphotericin B; however, they are too expensive to use in the majority

*e-mail: poliany.santos@gmail.com

of endemic countries.[10] Nonetheless, extensive toxicity complications are associated with the available drugs, and the emergence of drug resistant parasite strains is another problem related to leishmaniasis therapeutics.

American trypanosomiasis, also known as Chagas disease (CD), is a parasitic illness that results from infection by the hemoflagellate protozoan *Trypanosoma cruzi* (*T. cruzi*).[11] CD is an important public health problem not only in Latin America, where it is endemic in 21 countries, but also increasingly in other areas such as Europe, North America, Japan and Australia, mainly due to migration.[12,13] Around 6 million people are affected worldwide and approximately 7000 deaths occur annually, making CD the major cause of death from parasitic disease in Latin America and a significant contributor to the global burden of cardiovascular disease, as CD is the most important cause of infectious cardiomyopathy in the world.[14,15] Concerning therapeutics for CD, two almost 100-year old drugs are the only treatments used, namely the heterocyclic derivatives benzonidazole (**1**) and nifurtimox (**2**). However, neither of these drugs is effective during the chronic phase of the disease, and both of them cause numerous toxic side effects.[11]

In this context, both of the mentioned neglected Trypanosomiasis diseases present an urgent and continuous need for safe and effective new drugs. In the development of new treatments, it is important to identify the molecular targets of a new drug in order to avoid potential side effects. Parasite proteases have been described as promising chemotherapeutic targets.[16-18] The interest in cysteine proteases (CP) as targets derives from the recognition that they are critical to the life cycle or pathogenicity of many parasites, including *Leishmania* and *T. cruzi*. Parasite CPs play key roles in immunoevasion, enzyme activation, virulence, and tissue and cellular invasion.[19] Thus, CPs are an interesting target for the design and development of new antileishmanial and trypanocidal drug candidates. CPs are putative virulence factors of *Leishmania*,[20,21] and indeed *Leishmania* parasites that lack the multicopy cathepsin L-like genes in the CPB array (Δcpb) are considerably less effective at invading macrophages *in vitro*.[22] In 2003, the gene which expresses the *L. amazonensis* CP isoform 1 (Llacys1) was described, and its sequence is similar to CPs from other *Leishmania* species.[23] In *T. cruzi* the best characterized CP is cruzipain,[24] also known as cruzain.[25] The enzyme is expressed in the four main stages of the parasite, and is present in lysosome-related organelles. Its highest concentration is found in an epimastigote-specific prelysosomal organelle called the "reservosome".[26,27]

We have recently described the antileishmanial activity of nine alkyl-substituted benzophenones (**1a-c**, **2a-c** and **3a-c**) against promastigotes[28] and intracellular amastigotes of *L. amazonensis.* Moreover, we also highlighted the potential of benzophenones as non-competitive cysteine protease inhibitors against isoforms of cathepsin-L such as CPB and cruzain.[29] The aim of this study is to understand the interactions involved in the binding of these compounds to *T. cruzi* and *L. amazonensis* CPs using computational methods and to gain experimental insights into their pattern of inhibition. Thus, to predict the possible binding site responsible for the non-competitive inhibition of cysteine proteases by benzophenone derivatives, an *in silico* study of cruzain and Llacys1 interacting with the compounds was performed using homology, docking and molecular dynamics (MD). We identified a common binding site in cruzain and Llacys1 for compounds **1c**, **2c** and **3a**. The information in this study is expected to provide a better understanding of the mechanism of inhibition of *T. cruzi* and *L. amazonensis* cysteine proteases and to enable the design of safe and potent new drugs against neglected Trypanosomiasis diseases.

## Methodology

### Inhibitory activity against cruzain

Since the benzophenone derivatives LFQM-117 (**1c**), LFQM-120 (**2c**) and LFQM-121 (**3a**) (Figure 1), which were further evaluated *in vitro* against the amastigote form of *L. amazonensis*, showed the best relationship between potency against the parasite (50.0% inhibitory growth concentration ($IC_{50}$) = 74.4 ± 3.4; 27.4 ± 3.5 and 66.1 ± 2.4 μM) and safety (selectivity index (SI) = 6.7; 11.9 and 7.3), these compounds were selected for the cruzain inhibition assay.[29]

The concentration of cruzain obtained from *Escherichia coli*, strain DH5a containing the *T. cruzi* cruzain expression plasmid (kindly supplied by McKerrow, University of California, San Francisco) was 1.7 μM. Hydrolysis of the fluorogenic peptide substrate (10 μM carbobenzoxy-Phe-Arg-7-amino-4-methylcoumarin (Z-FR-MCA)) by the enzyme was assessed in 50 mM sodium acetate, pH 6.2 at 37 °C and 10 mM ethylenediaminetetraacetic acid (EDTA), with previous activation of the enzyme by the addition of 5 mM dithiothreitol (DTT) for 5 min. The inhibitors assayed were **1c**, **2c** and **3a**, with concentrations ranging from 1 to 200 μM. Substrate hydrolysis was monitored by measuring the fluorescence at λem = 460 nm with excitation at λex = 380 nm using a Cary Varian fluorescence spectrophotometer. The $IC_{50}$ values were determined by non-linear regression using GraFit 5.0 software (Erithacus Software Ltd.).[30]
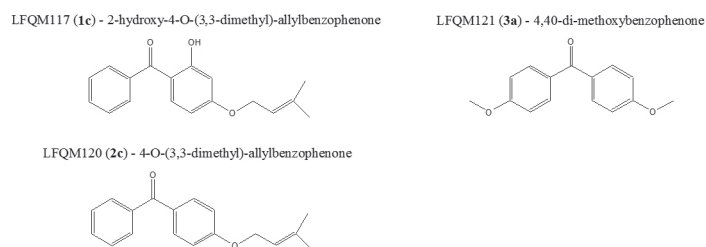
**Figure 1.** Chemical structures of benzophenones.

## Homology modelling

Since there are no crystal structures of the cysteine protease encoded by the *L. amazonensis* Llacys1 gene deposited in the Protein Data Bank[31] (PDB), a homology model was generated using the MODELLER[32] 9v11 program. The primary sequence of the cysteine proteinase target [Protein: AAP21894] consisting of 209 residues belonging to the peptidase_C1 region (residues 132-340) was obtained from the National Center for Biotechnology Information[33] (NCBI). We selected cruzain from *T. cruzi* comprising 215 residues as a template [PDB: 1EWP, resolution: 1.75 Å], with the aid of the basic local alignment search tool program[34] (BLAST). The sequence alignment between the target and the template protein was performed using the ClustalW2[35] program. We generated 100 models, which were submitted to a stereochemical analysis using the PROCHECK[36] program. The model with the highest percentage of amino acid residues in the most favorable regions of the Ramachandran plot was selected.[37,38] The best model was subjected to molecular dynamics simulation with 10 ns duration to examine relaxation of the atoms. The final structure obtained from the molecular dynamics simulation was used for binding site prediction and molecular docking analysis.

## Binding site prediction and druggability

DoGSite[39] finds active sites in proteins and evaluates their druggability. The binding pockets, volume and drug score were calculated using the DoGSiteScorer for the Llacys1 model and cruzain. DoGSite predicts druggability, reporting a value between 0 and 1. A score of zero indicates that a site is nondrugable and a score greater than zero indicates druggability; the closer the score is to 1 the greater the druggability.

## Molecular docking

Docking analysis was performed for all of the pockets found for the Llacys1 and cruzain models with the two benzophenone derivatives, using the AutoDock4[40] and Vina[41] programs. The protonated states of ionizable residues were determined based on the $pK_a$ values predicted by H++ server[42] for receptors at pH 7. Molecules were drawn in the software ChemSketch[43] and their geometries were optimized with the Gaussian09[44] program using a hybrid density functional method (B3LYP)[45] and basis set 6-31+G(d,p).[46] The molecular descriptors of compounds such as molecular weight, hydrogen donors and acceptors, log P, total polar surface area (TSPA), rotatable bonds, Lipinski violations, and oral bioavailability were obtained using the FAF-Drugs3[47] web server. The structures of cruzain, Llacys1, **1c**, **2c** and **3a** were prepared using the MGLTools[48] suite of programs. For docking simulations, a grid box of $60 \times 60 \times 60$ and $22.5 \times 22.5 \times 22.5$ points was constructed, with a grid spacing of 0.375 and 1.0 Å for AutoDock4 and Vina, respectively, in the center of the box, covering the entirety of all of the cavities found for the Llacys1 and cruzain. Details of the configuration parameters and the centers of the search space are in Table S1 (in the Supplementary Information (SI) section). Molecular docking was performed using flexible ligands and rigid receptors.

## Molecular dynamics

The Llacys1 model, Llacys1- and cruzain-compound complexes with the best docking poses were placed in a cubic periodic box and solvated by about 32000 water molecules, respectively, using a transferable intermolecular potential 3P[49] (TIP3P) model. The AMBER99sb[50] force field was used for the Llacys1 and cruzain templates. Force field parameters for the small molecules were obtained using the generalized amber force field (GAFF).[51] Amber topologies for the compounds were converted to GROMACS format using acpype.[52] The system was neutralized and the ionic strength (0.15 mol L$^{-1}$) of the medium was adjusted by adding Na$^+$ and Cl$^-$. The molecular dynamics simulations were performed using the GROMACS[53-55] software package, keeping the number of particles constant. The energy minimization of the systems was performed with the steepest descent method and number of steps of 100

with cycles until the energy converged. The pressure and temperature (NpT ensemble) were kept constant at 300 K and 1.0 bar using the velocity-rescaling[56] and Parrinello-Rahman[57] methods for temperature and pressure control, with 0.1 and 2.0 ps time coupling, respectively. Equilibration periods were 1.0 ns, and production runs were of 10 ns duration. A cutoff distance of 1.2 nm was used for Lennard-Jones interactions and particle mesh Ewald[58] (PME) was used for long-range electrostatic interactions. The bond lengths of hydrogen atoms were controlled with the LINCS[59] algorithm, and the SETTLE[60] algorithm was used to constrain the geometry of the water molecules.

Binding free energy estimation for cruzain- and Llacys1-compound complexes, based on molecular mechanics/Poisson-Boltzmann solvent-accessible surface area (MM-PBSA),[61] was calculated using the g_mmpbsa tool implemented as a subroutine of the GROMACS and APBS[62] packages. The binding free energy was estimated from 1000 snapshots extracted every 10 ps during the MD trajectories and does not include the entropy term. LigPlot+[63] was used to evaluate the interaction complexes.

## Results and Discussion

### Inhibitory activity against cruzain

The inhibition assay against cruzain was performed with the benzophenone derivatives **1c**, **2c** and **3a**. The results show potential inhibition of the enzyme (Table 1) by **1c** and **2c**, but **3a** showed no inhibition. These same compounds were also tested against papain, in which **1c** and **2c** were more active than **3a** with $IC_{50}$ values of 42.8,

**Table 1.** $IC_{50}$ values for the *in vitro* inhibitory effects of compounds **1c**, **2c** and **3a** against cruzain

| Compound | $IC_{50}$ / ($\mu$g mL$^{-1}$) |
|---|---|
| **1c** | 9.51 |
| **2c** | 10.86 |
| **3a** | no inhibition |

**Table 2.** Molecular properties of the compounds obtained from FAF-Drugs3[a]

72.1 and 253.6 $\mu$M, respectively.[29] Therefore, the methoxyl group in **3a** may not be favorable for cruzain inhibition. In addition, it was observed that **1c** and **2c** are more lipophilic than **3a** according to log P (Table 2). Thus, lipophilicity may influence the characteristics of the interactions of the compounds with the binding site. Compounds **1c**, **2c** and **3a** were used for an *in silico* assay to predict binding modes on the surface of the enzymes cruzain and Llacys1, based on the non-competitive inhibition mechanism of **1c** against cruzain and two isoforms of *Leishmania* CP, rCPB2.8 and rCPB3.0, as previously described.[29]

### Homology modelling

A 3D structure for Llacys1 (PDB coordinate file can be found in SI) was constructed by homology modelling based on the X-ray crystal structure of cruzain from *T. cruzi* (PDB: 1EWP, resolution: 1.75 Å). No homology was available for the N and C terminal regions, so only the 209 residues (amino acids 132-340 of the complete enzyme) corresponding to the peptidase_C1 region were modeled. A sequence alignment between the template and the peptidase_C1 region of Llacys1 computed by EMBOSS Needle and shown in Figure 2 exhibits 60.0% identity, 74.0% sequence similarity and 2.8% gaps.[64] The matching score computed by ClustalW2 was 61.24, being sufficient for structure prediction.[65]
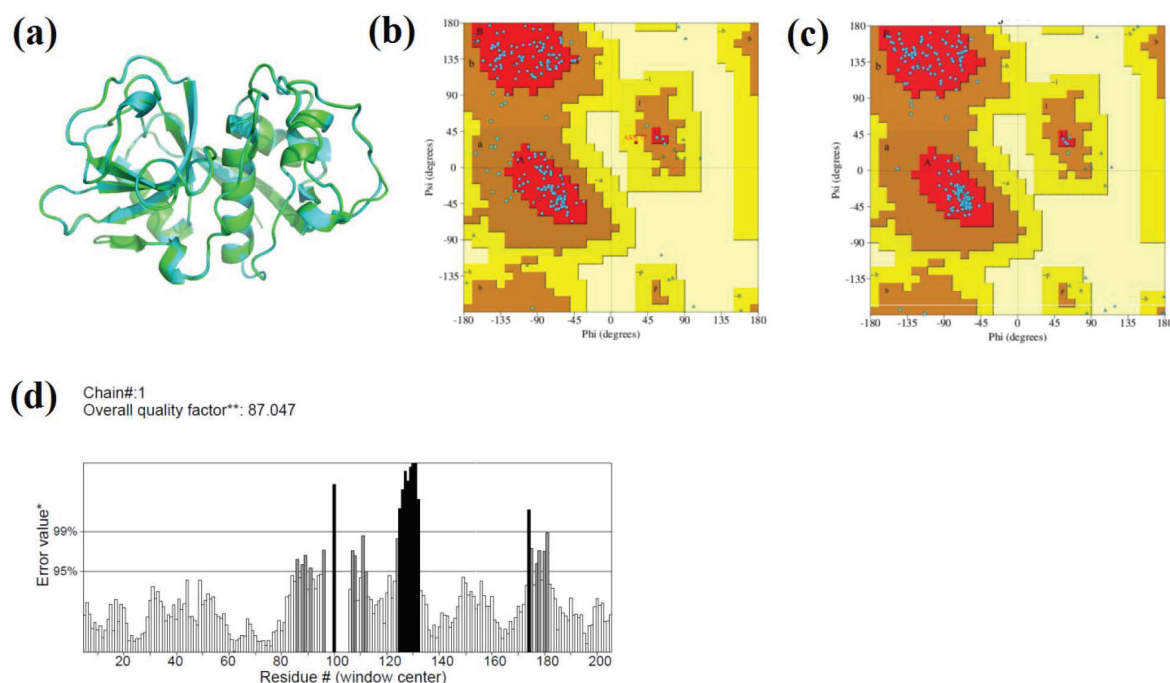


**Figure 2.** Alignment in the ClustalW2 program between the primary sequences of the enzyme cruzain of *T. cruzi* (1EWP) and cysteine protease isoform 1 from *L. amazonensis* (AAP21894.1).

One hundred models were generated and model number 82 was selected, which presented 87.9% of the residues in the most favorable regions of the Ramachandran plot,

| Compound | MW | log P | tPSA (A$^2$) | Donor HB | Accept HB | Rotatable bond | Lipinski violation (max. 4) | Oral bioavailability (Veber) |
|---|---|---|---|---|---|---|---|---|
| **1c** | 281.33 | 4.93 | 46.53 | 1 | 3 | 6 | 0 | good |
| **2c** | 266.33 | 4.73 | 26.30 | 0 | 2 | 5 | 0 | good |
| **3a** | 242.27 | 3.21 | 35.53 | 0 | 3 | 4 | 0 | good |

[a]MW: Molecular weight; log P: logarithm of the partition coefficient between *n*-octanol and water, characterizing lipophilicity; tPSA: topological polar surface area; donor HB: hydrogen bond donors; accept HB: hydrogen bond acceptors.

**Figure 3.** Model validation: (a) overlap between the Llacys1 model and cruzain of *T. cruzi* used as template; (b) Ramachandran plot of the Llacys1 model; (c) Ramachandran plot of the Llacys1 model after molecular dynamics and (d) ERRAT analysis of the Llacys1 model. *On the error axis, two lines are drawn to indicate the confidence with which it is possible to reject regions that exceed that error value. **Expressed as the percentage of the protein for which the calculated error value falls bellow the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3 Å) the average overall quality factor is around 91%.

11.6% of residues in additional allowed regions and 0.6% in the generously allowed and 0.0% disallowed regions (Figure 3). ERRAT was used to check the protein structure prediction based on known crystallographic structures, and the score found for the Llacys1 model was 87.047%. Verify 3D evaluates the profile of the three-dimensional structure; for the Llacys1 model, we found 89.0% of the residues had an average 3D-1D score ≥ 0.2. The calculated RMSD (root mean square deviation) was 0.324 Å (computed by the UCSF Chimera software MatchMaker tool)[66] taking all atoms into account. The ERRAT, Verify3D and RMSD analysis confirmed that the model was of good quality. This model was submitted for refinement using MD simulations. The relaxed model was evaluated by Procheck (Figure 2) and 93.1% of the residues were located in favorable regions of the Ramachandran plot, while 6.9% of residues were in additional allowed regions and 0.0% in the generously allowed and disallowed regions. Thus, we can conclude that the structural relaxation of the protein produced by the MD simulations was important for the refinement of the Llacys1 model.
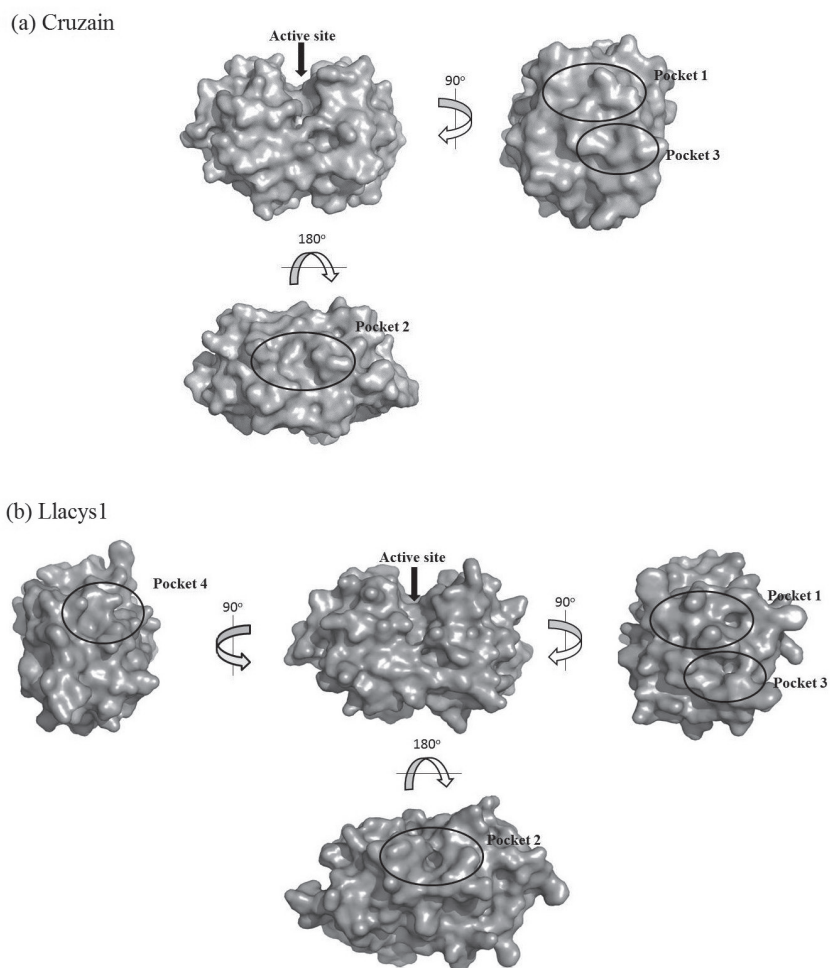
Binding site prediction and druggability

Binding site detection is used in structure-based drug design to identify where a potential drug may bind. Figure 4 shows the location of the binding sites on the surface of the protein structure and Table 3 shows the volume, depth and drug score for cruzain and Llacys1. Three pockets were found for cruzain and four pockets for Llacys1; the first two pockets presented larger volume, depth and drugability. The P_1, P_2, and P_3 pockets of cruzain and Llacys1 are similar, while the Llacys1 P_4 pocket had a lower drug score. In the docking step, all of the pockets predicted for cruzain and Llacys1 were selected.

Molecular docking

An *in silico* study was performed with cruzain and Llacys1 against compounds **1c**, **2c** and **3a**. Molecular docking analysis was performed for several pockets in both cruzain and Llacys1 against compounds using the AutoDock4 and Vina programs (Figure 5). Table 4 shows the binding energy values for the enzymes. The binding energies are better for pocket P_2 in both enzymes (Figure 5) calculated by both docking programs. Thus, we can infer that pocket P_2 is a possible allosteric site for benzophenone derivatives in cruzain and Llacys1. As well as having the best binding energies, there are also conserved residues between the two enzymes in this pocket. Thus, these two programs confirm the same fit for the compounds in pocket P_2. The binding energy

**Figure 4.** Predicted allosteric sites on (a) cruzain and (b) Llacys1.

**Table 3.** Volume, depth and druggability of the predicted allosteric sites on cruzain and Llacys1

| Receptor | Binding site | Volume / Å³ | Depth / Å | Drugscore |
|----------|-------------|-------------|-----------|-----------|
| | P_1 | 250.37 | 10.07 | 0.41 |
| Cruzain | P_2 | 200.06 | 13.25 | 0.52 |
| | P_3 | 104.13 | 9.88 | 0.36 |
| | P_1 | 224.64 | 16.67 | 0.65 |
| Llacys1 | P_2 | 173.38 | 10.16 | 0.57 |
| | P_3 | 108.22 | 8.40 | 0.21 |
| | P_4 | 100.42 | 6.71 | 0.15 |

values for cruzain in the P_2 pocket are −6.27, −6.26 and −5.17 kcal mol⁻¹ as calculated by the AutoDock4 program and −6.3, −6.1 and −5.4 kcal mol⁻¹ using Vina for compounds **1c**, **2c** and **3a**, respectively. The results are very similar for both programs, and correspond with the $IC_{50}$ values against cruzain (Table 1). However, **1c** is a little more active and **3a** presents the highest energy and no activity. In Llacys1 the binding energy values are −5.72, −5.77 and −4.58 kcal mol⁻¹ calculated using AutoDock4 and −6.4, −6.4

and −4.5 kcal mol⁻¹ using Vina for compounds **1c**, **2c** and **3a**. The energies being very close using both programs for compounds **1c** and **2c**, except for **3a** which shows higher binding energy values. The binding energies in the pocket P_2 of cruzain and Llacys1 complexes are very close when calculated by either program. The binding energies of the compounds **1c** and **2c** are probably so similar because the molecules are highly similar, being differentiated only by the presence of a hydroxyl group in the ring of compound **1c**.
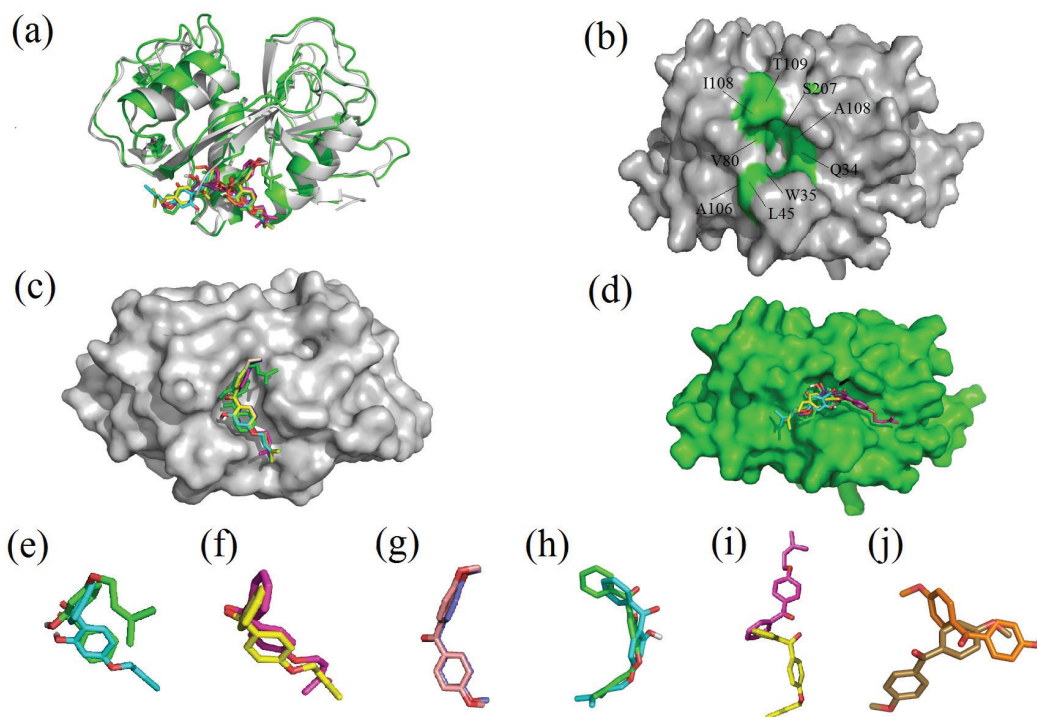
**Figure 5.** Binding poses of cruzain (gray) and Llacys (green) calculated by the AutoDock4 and Vina programs: (a) overlap of cruzain and Llacys with **1c**, **2c** and **3a** by AutoDock4 and Vina; (b) residues conserved between cruzain and Llacys1 in the P_2 pocket; (c) cruzain surface with **1c**, **2c** and **3a** by AutoDock4 and Vina; (d) Llacys1 surface with **1c**, **2c** and **3a** by AutoDock4 and Vina; (e) pose of **1c** with cruzain AutoDock4 (green) and Vina (cyan); (f) pose of **2c** with cruzain AutoDock4 (magenta) and Vina (yellow); (g) pose of **3a** with cruzain AutoDock4 (blue) and Vina (pink); (h) pose of **1c** with Llacys1 AutoDock4 (green) and Vina (cyan); (i) pose of **2c** with Llacys1 AutoDock4 (magenta) and Vina (yellow) and (j) pose of **3a** with Llacys1 AutoDock4 (orange) and Vina (brown).

**Table 4.** Binding energies as calculated by the AutoDock4 and Vina programs

| Receptor | Binding site | Compound | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1c | | 2c | | 3a | |
| | | $\Delta G_{AutoDock4}$ / (kcal mol$^{-1}$) | $\Delta G_{Vina}$ / (kcal mol$^{-1}$) | $\Delta G_{AutoDock4}$ / (kcal mol$^{-1}$) | $\Delta G_{Vina}$ / (kcal mol$^{-1}$) | $\Delta G_{AutoDock4}$ / (kcal mol$^{-1}$) | $\Delta G_{Vina}$ / (kcal mol$^{-1}$) |
| Cruzain | P_1 | −5.49 | −5.4 | −5.31 | −5.6 | −4.88 | −4.9 |
| | P_2 | −6.26 | −6.3 | −6.27 | −6.1 | −5.17 | −5.4 |
| | P_3 | −4.93 | −5.3 | −5.23 | −5.4 | −4.21 | −5.0 |
| Llacys1 | P_1 | −5.42 | −5.1 | −5.07 | −5.0 | −4.49 | −4.9 |
| | P_2 | −5.72 | −6.4 | −5.77 | −6.4 | −4.58 | −4.5 |
| | P_3 | −5.04 | −5.6 | −5.32 | −5.4 | −4.18 | −4.5 |
| | P_4 | −5.24 | −6.1 | −5.58 | −5.7 | −4.33 | −4.6 |

$\Delta G_{AutoDock4}$: binding energy calculated by AutoDock4; $\Delta G_{Vina}$: binding energy calculated by Vina.

Compound **3a** shows the greatest difference both in the binding energy values and in its chemical structure. This compound does not show good interaction with the site, nor does it inhibit cruzain. Therefore, compound **3a** may not be a good *Leishmania* inhibitor promising. Thus, pocket P_2 with compounds **1c** and **2c** was selected for molecular dynamics simulation, in order to refine the docking results taking into account possible conformational changes in the interaction site and the effects of the solvent, as well as to possibly improve the results of the binding energy calculations taking into account the flexibility of both proteins and benzophenones.

Absorption, distribution, metabolism and excretion (ADME) properties

For a new drug to be successfully developed, its ADME properties are very important, since unfavorable properties can cause a drug to be rejected.[67] The web server FAF-Drugs3 was used to determine the ADME properties

(Table 2) of the compounds. All of the compounds showed good oral bioavailability, and conformed to Lipinski's rule of five by meeting the following criteria (RO5): molecular weight ≤ 500; calculated log P (*n*-octanol/water) ≤ 5; number of hydrogen bond donors (HBD) ≤ 5; number of hydrogen bond acceptors (HBA) ≤ 10.[68] Therefore, compounds **1c**, **2c**, and **3a** have favorable ADME properties.

### Molecular dynamics

To understand the interactions between the compounds and their targets, MD simulations were performed for a 10 ns duration. First, we evaluated the structural stability of the protease-inhibitor complex using the RMSD of the carbon alpha receptor relative to the initial structure, whose pose was obtained by better energy docking. The RMSD calculation is an important measure for the analysis of molecular dynamics simulations in biological systems. Using the RMSD it is possible to assess the structural stability of the protein, providing a measure of the deviation of a given molecular structure over time. The RMSD values (Figure 6) of cruzain in complex with **1c** and **2c** reached values up to 0.168 and 0.161 nm, with mean values of RMSD of 0.162 ± 0.001 and 0.163 ± 0.001 nm, respectively. Llacys1 in complex with compounds **1c** and **2c** reached values up to 0.159 and 0.150 nm, with mean values of 0.090 ± 0.010 and 0.100 ± 0.020. The mean values of RMSD indicate that there was little movement of the enzymes. However, the complexes with cruzain rapidly reached equilibrium and were slightly more stable than the complexes with Llacys1.

In addition, the radius of gyration of the proteins was evaluated, which means that the hydrodynamic radius of the molecular system was estimated. Using this calculation, it is possible to show processes of compression or expansion, i.e., folding or unfolding of the polypeptide chain, indicating a conformational change in the protein. Figure 6 shows that cruzain presents good stability in complex with compounds **1c** and **2c**, but after 4 ns in complex with **1c**, the enzyme becomes more compact than when in complex with **2c**, with average values of 1.630 ± 0.010 and 1.640 ± 0.010, respectively. In Llacys1, the complexes were stable after 5 ns. The enzyme complexed with **1c** became more compact than with **2c** with average values of 1.620 ± 0.010 and 1.630 ± 0.010, respectively. These results indicate that both enzymes in complex with **1c** and **2c** are more stable.
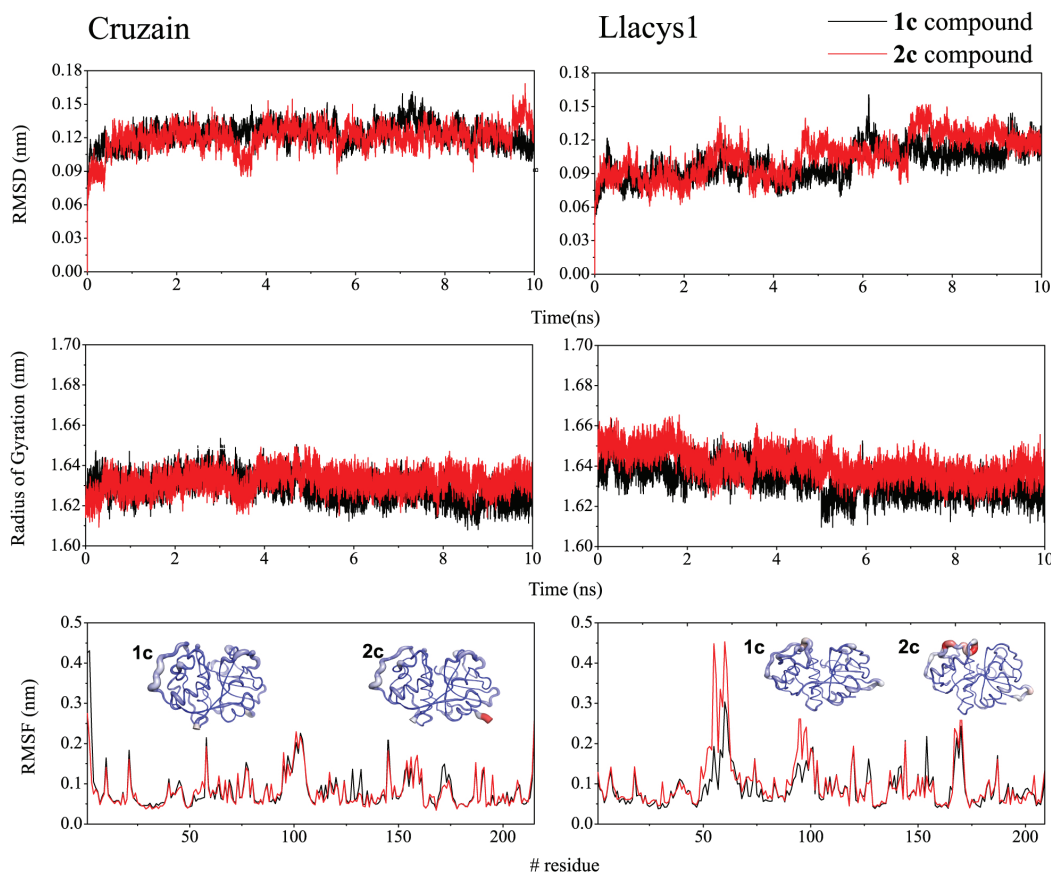


**Figure 6.** RMSD, radius of gyration and RMSF values of Llcys1-compound complexes were generated against the initial structures of the complexes during a 10 ns MD simulation period. The three-dimensional structures show the RMSF values in the regions of smallest to largest movement (blue to red).

RMSF is the root mean square deviation based on the coordinates of the amino acid residues atoms of the protein. Figure 6 shows the RMSF values of the enzymes and the areas of the protein where major fluctuations occur during the simulation. Cruzain presents mean values of RMSF of $0.085 \pm 0.053$ and $0.082 \pm 0.042$ nm for complexes with compounds **1c** and **2c**, respectively, the values are very close to each other. Llacys1 presents mean values of RMSF of $0.089 \pm 0.045$ and $0.102 \pm 0.068$ nm for complexes with compounds **1c** and **2c**, respectively. Thus, the Llacys1-**2c** complex presents regions with greater movement than in the Llacys1-**1c** complex. It is also observed that in the region of the binding site, the movement of the residues is very small, which suggests a strong affinity of this region for the ligands.
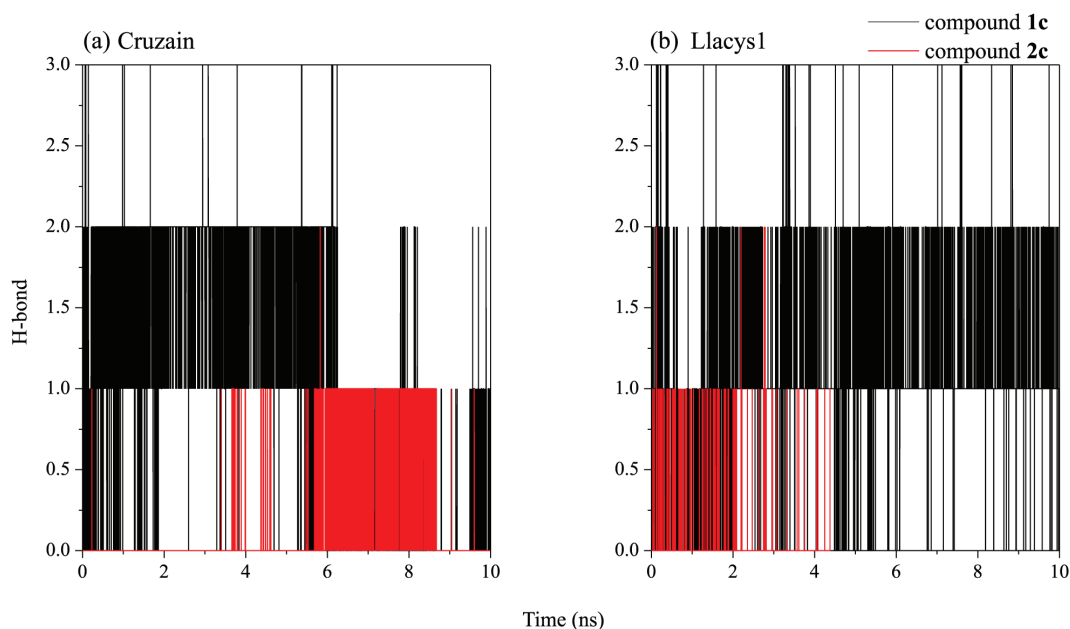
### Hydrogen bonds

Hydrogen bonding was determined at a distance of 0.35 nm and at an angle of $30°$ using the g_hbond tool of Gromacs (Figure 7). For cruzain in complex with **1c** the average value of hydrogen bonding was 2, with 53.5% occupancy for the pair **1c** (H3-hydroxyl group), Gln37 (OE1) and 74.1% occupancy for Val214 (H), **1c** (O2). For **2c** the mean binding value was zero. For Llacys1 in complex with molecule **1c** the average value of hydrogen bonds was 1, with binding between the pair **1c** (H8-hydroxyl group), Thr209 (OC2) having 57% occupancy. The value was zero for hydrogen bonding with **2c**. In both enzymes the complexes with compound **1c** have hydrogen bonds, which may be favorable for the stability of the complex. The

differences in hydrogen bonding between molecules present valuable information, because of the hydroxyl group in the ring of molecule **1c**. Its presence may explain the greater affinity of interaction resulting in the lower $IC_{50}$ value for cruzain. However, hydrogen bonding in the Llacys1-**1c** complex is not as stable as in cruzain-**1c**.

### Binding energy

The molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method has been used to estimate binding free energies both in the study of biomolecular interactions as well as for scoring functions in drug design.[60] This method, combined with MD simulations to rescore the docked complexes, lent significant support to the choice of the best ligand. The calculated free energies of the cruzain and Llacys1 complexes with the benzophenone derivatives ranged from $-7.54$ to $-13.50$ kcal mol$^{-1}$. Table 5 shows the energy components that contribute to the binding free energy, the most favorable components being the van der Waals, electrostatic and nonpolar solvation terms, while polar solvation was unfavorable. In the two compounds **1c** and **2c**, $G_{vdW}$ is much stronger than $G_{eletrostatic}$, $G_{polar}$ and $G_{SASA}$ with values greater than 20.00 kcal mol$^{-1}$. It is evident that this energy governs the strength of the interactions of the compounds with the cysteine proteases. However, $G_{eletrostatic}$ is more favorable for compound **1c**, because its hydrogen bonding is more stable than for molecule **2c** due to the chemical nature of the ligands. The binding free energy for the cruzain complexes follows the same pattern as the $IC_{50}$ values (Table 1), where compound **1c** presents lower values.



**Figure 7.** Hydrogen bond number over time in simulations of complexes with (a) cruzain and (b) Llacys1.

**Table 5.** Binding free energy results for different compounds bound with cruzain and Llacys1

| Compound | | $G_{vdW}$ / (kcal mol$^{-1}$) | $G_{eletrostatic}$ / (kcal mol$^{-1}$) | $G_{polar solv}$ / (kcal mol$^{-1}$) | $G_{SASA}$ / (kcal mol$^{-1}$) | $G_{total}$ / (kcal mol$^{-1}$) |
|---|---|---|---|---|---|---|
| Cruzain | **1c** | −22.57 ± 3.75 | −6.24 ± 0.31 | 17.78 ± 2.45 | −2.46 ± 0.29 | −13.50 ± 2.78 |
| | **2c** | −21.11 ± 2.88 | −1.06 ± 0.36 | 23.56 ± 2.55 | −2.58 ± 0.25 | −12.96 ± 2.10 |
| Llacys1 | **1c** | −22.80 ± 3.07 | −8.02 ± 1.21 | 26.16 ± 5.29 | −2.88 ± 0.34 | −7.54 ± 2.05 |
| | **2c** | −27.45 ± 3.08 | −2.09 ± 0.28 | 19.77 ± 3.82 | −2.94 ± 0.22 | −12.72 ± 2.02 |

$G_{vdW}$: van der Waals energy; $G_{eletrostatic}$: electrostatic energy; $G_{polar solv}$: polar solvation energy; $G_{SASA}$: solvent accessible surface area energy and $G_{total}$: binding energy.

The lower binding energy can be explained by the presence of two hydrogen bonds that can better stabilize the complex in the pocket on the surface of the protein. In complexes with Llacys1, the binding free energy follows the ranking of the IC$_{50}$ values (**1c** = 74.4 and **2c** = 27.4 μM) against the amastigote form of *L. amazonensis*. In this case, it is observed that there is only one hydrogen bond, with 57% occupancy, in the Llacys-**1c** complex, which is not as stable as the cruzain-**1c** complex. In addition, the contribution of van der Waals energy is greater for Llacys1-**2c** compared with the other complexes. Thus, although the compounds bind to the same site on the surface of the enzymes and have conserved amino acid residues, small differences in other residues can result in differences in the binding free energy.

The energy contribution of each residue was calculated to verify the difference between the complexes. This energy decomposition analysis (Figure 8) shows major contributions from the hydrophobic residues Leu40, Val213 and Val214

for cruzain in complexes with **1c** and **2c**. The polar residue Trp35, and the hydrophobic residues Ile108 and Ser38 make major contributions in Llacys1 in complexes with **1c** and **2c**. The energy contribution is unfavorable for the hydrophobic residue Ala212 in cruzain, and for the His40 and Lys107 residues for Llacys1. These results verify that the Val residues are important for interactions in the cruzain-**1c** complex. These interactions in combination with hydrogen bonds may account for the greater activity of **1c** against cruzain, as the Val residues are not conserved in Llacys1.

Figure 9 shows interactions in the cruzain- and Llacys1- complexes calculated by LIGPLOT from the final structure derived from the MD simulations. In all poses, van der Waals contacts were observed between protein residues, except for complexes with ligand **1c** where hydrogen bonding was also significant. In addition, there was a high affinity of the conserved residue Gln34/Gln37 for compounds **1c** and **2c** in both enzymes, showing it to be a key residue in the interaction.
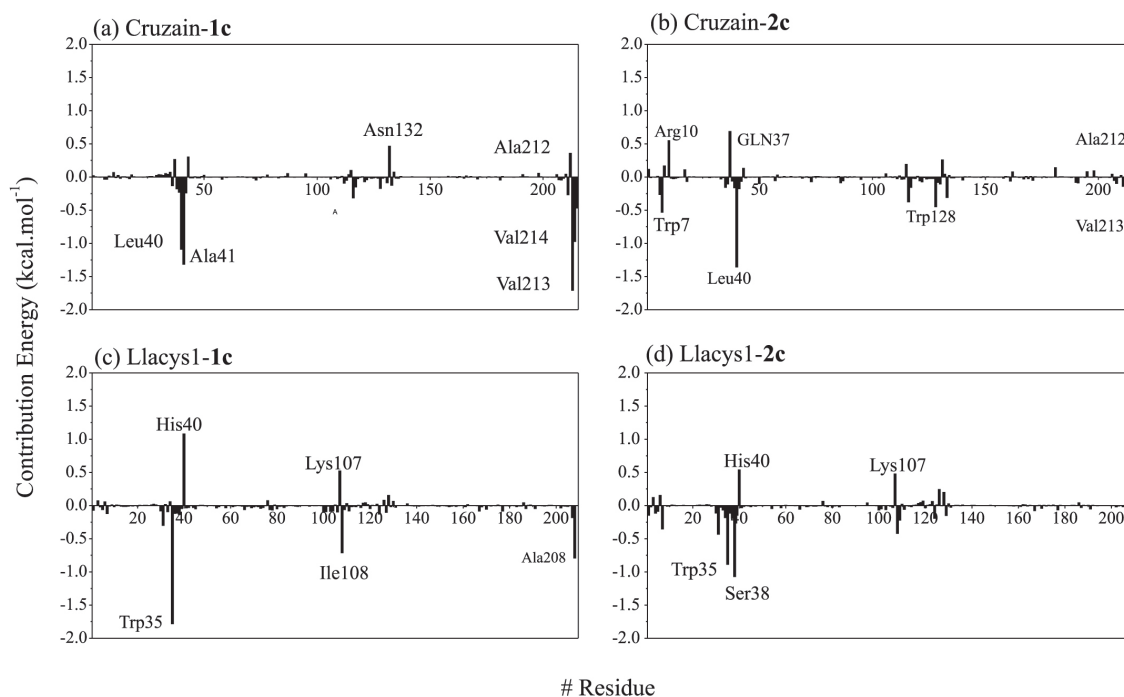


**Figure 8.** *Per*-residue decomposition free energy of cruzain and Llacys1 complexes with compounds (a, c) **1c** and (b, d) **2c** in simulations.
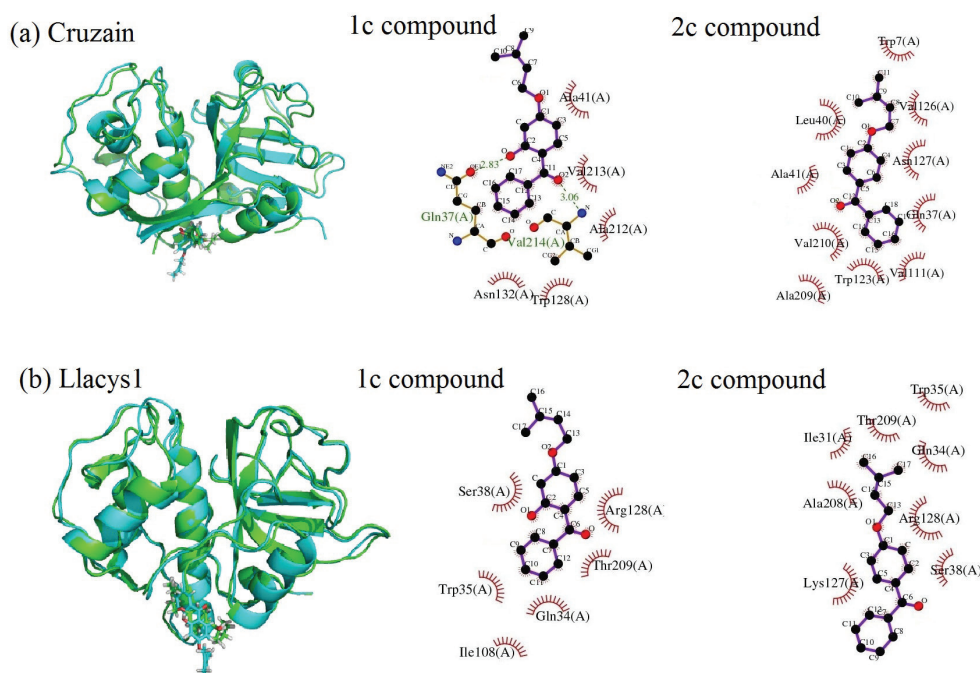
**Figure 9.** (Left) Binding poses of (a) cruzain- and (b) Llacys1-complexes obtained after 10 ns MD simulations. **1c** (cyan) and **2c** (green) are rendered in stick form. Schematic representation of the final structure of MD. (Center, right) The interaction of cruzain and Llacys1 with the compounds calculated by LIGPLOT. Thatched semi-circles indicate van der Waals contacts between hydrophobic protein residues and the compounds **1c** and **2c**. Hydrogen bonds are shown as green dashed lines.

## Conclusions

We investigated the mechanism of interaction of Llacys1 and cruzain with benzophenones by homology modelling, docking and molecular dynamics analysis. These enzymes belong to the C1 family of cysteine proteases and present a high degree of sequence identity with each other. An enzymatic inhibition assay of cruzain showed that compounds **1c** and **2c** act as potent inhibitors, however, **3a** showed no inhibition. To better understand how these derivatives bind to the proteases, we searched for binding sites on cruzain and Llacys1. We found a similar binding site on the surface of cruzain and Llacys1 by computational methods. The *in silico* results corroborated the experimental data, with compounds **1c** and **2c** being the strongest inhibitors of cruzain and the amastigote form of *L. amazonensis*, respectively. These results may contribute to the development of novel inhibitors for the treatment of leishmaniasis and Chagas disease, with the benzophenone derivatives **1c** and **2c** as potential candidates for prototype drugs.

## Supplementary Information

Supplementary information is available free of charge at http://jbcs.org.br as a PDF file.

## References

1. Conteh, L.; Engels, T.; Molyneux, D. H.; *Lancet* **2010**, *375*, 239.

2. Mathers, C. D.; Ezzati, M.; Lopez, A. D.; *PLoS Neglected Trop. Dis.* **2007**, *1*, e114.

3. Njoroge, M.; Njuguna, N. M.; Mutai, P.; Ongarora, D. S.; Smith, P. W.; Chibale, K.; *Chem. Rev.* **2014**, *114*, 11138.

4. Chirac, P.; Torreele, E.; *Lancet* **2006**, *367*, 1560.

5. Alvar, J.; Yactayo, S.; Bern, C.; *Trends Parasitol.* **2006**, *22*, 552.

6. Kedzierski, L.; *Hum. Vaccines* **2011**, *7*, 1204.

7. Moreno, J.; Alvar, J.; *Trends Parasitol.* **2002**, *18*, 399.

8. Chakravarty, J.; Sundar, S.; *J. Global Infect. Dis.* **2010**, *2*, 167.

9. Croft, S. L.; Yardley, V.; *Curr. Pharm. Des.* **2002**, *8*, 319.

10. Mohamed-Ahmed, A. H. A.; Brocchini, S.; Croft, S. L.; *Curr. Opin. Infect. Dis.* **2012**, *25*, 695.

11. Sueth-Santiago, V.; Decote-Ricardo, D.; Morrot, A.; Freire-de-Lima, C. G.; Lima, M. E.; *World J. Biol. Chem.* **2017**, *8*, 57.

12. Gascon, J.; Bern, C.; Pinazo, M. J.; *Acta Trop.* **2010**, *115*, 22.

13. Jackson, Y.; Pinto, A.; Pett, S.; *Trop. Med. Int. Health* **2014**, *19*, 212.

14. Cucunuba, Z. M.; Okuwoga, O.; Basanez, M. G.; Nouvellet, P.; *Parasites Vectors* **2016**, *9*, 42.

15. Martins-Melo, F. R.; Ramos Jr., A. N.; Alencar, C. H.; Heukelbach, J.; *J. Infect. Dev. Countries* **2012**, *6*, 817.

16. McKerrow, J. H.; Engel, J. C.; Caffrey, C. R.; *Bioorg. Med. Chem.* **1999**, *7*, 639.

17. McKerrow, J. H.; *Int. J. Parasitol.* **1999**, *29*, 833.

18. Selzer, P. M.; Pingel, S.; Hsieh, I.; Ugele, B.; Chan, V. J.; Engel, J. C.; Bogyo, M.; Russell, D. G.; Sakanari, J. A.; McKerrow, J. H.; *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 11015.

19. Sajid, M.; McKerrow, J. H.; *Mol. Biochem. Parasitol.* **2002**, *120*, 1.

20. Mottram, J. C.; Coombs, G. H.; Alexander, J.; *Curr. Opin. Microbiol.* **2004**, *7*, 375.

21. Mahmoudzadeh-Niknam, H.; McKerrow, J. H.; *Exp. Parasitol.* **2004**, *106*, 158.

22. Frame, M. J.; Mottram, J. C.; Coombs, G. H.; *Parasitology* **2000**, *121*, 367.

23. Lasakosvitsch, F.; Gentil, L. G.; dos Santos, M. R.; da Silveira, J. F.; Barbieri, C. L.; *Int. J. Parasitol.* **2003**, *33*, 445.

24. Cazzulo, J. J.; Franke, M. C. C.; Martinez, J.; de Cazzulo, B. M. F.; *Biochim. Biophys. Acta* **1990**, *1037*, 186.

25. Eakin, A. E.; Mills, A. A.; Harth, G.; McKerrow, J. H.; Craik, C. S.; *J. Biol. Chem.* **1992**, *267*, 7411.

26. Soares, M. J.; Souto-Padron, T.; de Souza, W.; *J. Cell Sci.* **1992**, *102*, 157.

27. Souto-Padron, T.; Campetella, O. E.; Cazzulo, J. J.; de Souza, W.; *J. Cell Sci.* **1990**, *96*, 485.

28. Maciel-Rezende, C. M.; de Almeida, L.; Costa, E. D.; Pires, F. R.; Alves, K. F.; Viegas Jr., C.; Dias, D. F.; Doriguetto, A. C.; Marques, M. J.; dos Santos, M. H.; *Bioorg. Med. Chem.* **2013**, *21*, 3114.

29. de Almeida, L.; Alves, K. F.; Maciel-Rezende, C. M.; Jesus, L. D. O. P.; Pires, F. R.; Junior, C. V.; Izidoro, M. A.; Júdice, W. A. S.; dos Santos, M. H.; Marques, M. J.; *Biomed. Pharmacother.* **2015**, *75*, 93.

30. Leatherbarrow, R. J.; *GraFit Version 5.0*, Erithacus Software Ltd., Horley, UK, 2002.

31. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E.; *Nucleic Acids Res.* **2000**, *28*, 235.

32. Šali, A.; Blundell, T. L.; *J. Mol. Biol.* **1993**, *234*, 779.

33. http://www.ncbi.nlm.nih.gov/protein/30142572?report=graph&sat=4&itemID=8, accessed in February 2015.

34. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J.; *Nucleic Acids Res.* **1997**, *25*, 3389.

35. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G.; *Bioinformatics* **2007**, *23*, 2947.

36. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M.; *J. Appl. Crystallogr.* **1993**, *26*, 283.

37. Saraiva, L. A.; Veloso, M. P.; Camps, I.; da Silveira, N. J. F.; *Mol. Inf.* **2011**, *30*, 219.

38. da Silveira, N. J.; Arcuri, H. A.; Bonalumi, C. E.; de Souza, F. P.; Mello, I. M.; Rahal, P.; Pinho, J. R.; de Azevedo, W. F.; *BMC Struct. Biol.* **2005**, *5*, 1.

39. Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M.; *Bioinformatics* **2012**, *28*, 2074.

40. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J.; *J. Comput. Chem.* **2009**, *30*, 2785.

41. Trott, O.; Olson, A. J.; *J. Comput. Chem.* **2010**, *31*, 455.

42. Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A.; *Nucleic Acids Res.* **2005**, *33*, W368.

43. Spessard, G. O.; *J. Chem. Inf. Model.* **1998**, *38*, 1250.

44. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Gaussian 09,* Gaussian, Inc., Wallingford, CT, USA, 2009.

45. Lee, C.; Yang, W.; Parr, R. G.; *Phys. Rev. B* **1988**, *37*, 785.

46. Ditchfield, R.; Hehre, W. J.; Pople, J. A.; *J. Chem. Phys.* **1971**, *54*, 724.

47. Lagorce, D.; Sperandio, O.; Baell, J. B.; Miteva, M. A.; Villoutreix, B. O.; *Nucleic Acids Res.* **2015**, *43*, W200.

48. Sanner, M. F.; *J. Mol. Graphics Modell.* **1999**, *17*, 57.

49. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.; *J. Chem. Phys.* **1983**, *79*, 926.

50. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C.; *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712.

51. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A.; *J. Comput. Chem.* **2004**, *25*, 1157.

52. da Silva, A. W. S.; Vranken, W. F.; *BMC Res. Notes* **2012**, *5*, 1.

53. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E.; *J. Chem. Theory Comput.* **2008**, *4*, 435.

54. van der Spoel, D.; Hess, B.; *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 710.

55. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E.; *Bioinformatics* **2013**, *29*, 845.

56. Bussi, G.; Donadio, D.; Parrinello, M.; *J. Chem. Phys.* **2007**, *126*, 014101.

57. Parrinello, M.; Rahman, A.; *J. Appl. Phys.* **1981**, *52*, 7182.

58. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G.; *J. Chem. Phys.* **1995**, *103*, 8577.

59. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M.; *J. Comput. Chem.* **1997**, *18*, 1463.

60. Miyamoto, S.; Kollman, P. A.; *J. Comput. Chem.* **1992**, *13*, 952.

61. Kumari, R.; Kumar, R.; Lynn, A.; Consort, O. S. D. D.; *J. Chem. Inf. Model.* **2014**, *54*, 1951.

62. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A.; *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037.

63. Laskowski, R. A.; Swindells, M. B.; *J. Chem. Inf. Model.* **2011**, *51*, 2778.

64. Rice, P.; Longden, I.; Bleasby, A.; *Trends Genet.* **2000**, *16*, 276.

65. Méndez-Cuesta, C. A.; Méndez-Lucio, O.; Castillo, R.; *J. Mol. Graphics Modell.* **2012**, *38*, 50.

66. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E.; *J. Comput. Chem.* **2004**, *25*, 1605.

67. Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L.; *Br. J. Pharmacol.* **2011**, *162*, 1239.

68. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J.; *Adv. Drug Delivery Rev.* **2001**, *46*, 3.