

Molecular Modelling Study of Heteroarylamide/Sulfonamide Compounds with Antitrypanosomal Activity

André L. P. F. Souza,^a Fábio J. B. Cardoso,^a Lucas S. Martins,^b Cláudio N. Alves,^b
José R. A. Silva ^{*,b} and Fábio A. Molfetta ^{*,a}

^aLaboratório de Modelagem Molecular, Instituto de Ciências Exatas e Naturais,
Universidade Federal do Pará, CP 11101, 60075-110 Belém-PA, Brazil

^bLaboratório de Planejamento e Desenvolvimento de Fármacos,
Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará,
CP 11101, 60075-110 Belém-PA, Brazil

According to the World Health Organization (WHO), Chagas disease (CD), whose etiological agent is the *Trypanosoma cruzi* (*T. cruzi*) parasite, affects about eight million people, mainly in Latin America. The cruzain enzyme is highlighted among the main biological targets, since it is the most abundant of the cysteine protease class from *T. cruzi* and is involved in the entire life cycle of the parasite, essential in regulating the interaction between parasite and host. The drugs available for the treatment of CD usually have strong side effects, and the nitro(triazole/imidazole)-based heteroarylamide/sulfonamide compounds (HA/S) emerge with high antitrypanosomal potential. In this study, the quantitative structure-activity relationship (QSAR) were built using partial least squares (PLS) regression, and the results were robust and adequate for predicting and proposing five new derivatives according to the statistical parameters. The docking results suggest that the best-scored HA/S derivatives showed hydrogen bonds with the residuals that comprise the catalytic region of the enzyme. The molecular dynamics (MD) simulations, performed with different methods, revealed the strong stability of the compound obtained by the QSAR model of this study, in addition to a better binding free energy value than the HA/S obtained from literature.

Keywords: Chagas disease, cruzain, QSAR, PLS, molecular docking, molecular dynamics

Introduction

The etiological agent of Chagas disease (CD), a neglected tropical disease, is the parasite *Trypanosoma cruzi* (*T. cruzi*). It is estimated that about 10 million people are infected with *T. cruzi* and more than 25 million people are in areas at risk for vector contamination, mainly in Latin America, where the disease is endemic.¹ The disease is transmitted through insect bites (*Triatoma infestans*), transfusion of contaminated blood, organ transplantation, congenital transmission and consumption of contaminated food or drink.²

There is no vaccine available to prevent the disease and the only two drugs on the market, benznidazole and nifurtimox, have serious side effects, low efficacy, and are inefficient during the chronic phase of the disease. Thus, there is an urgent need for the development of new drugs

that are safer and more effective to meet the therapeutic needs of infected individuals.^{3,4}

The cruzain enzyme, the main cysteine protease of *T. cruzi*, is a cathepsin L-like cysteine protease responsible for the major proteolytic activity, and that is involved in the hydrolysis of peptides and proteins.⁵ The enzyme is essential for parasite survival and growth and is among the most important targets for the development of candidates for new drugs for Chagas disease.⁶ Animal model studies⁷ have also validated this enzyme for the control and elimination of *T. cruzi*. Cruzain is expressed throughout the life cycle of the parasite and is important for the protozoan's development, survival and differentiation, responsible for its nutrition, immune system evasion and invasion of new tissues.¹ The validation of cruzain as a drug target relies on preclinical proof-of-concept studies,^{1,8} which showed that cruzain inhibitors significantly reduce parasite burden *in vivo*. Consequently, its inhibition is an interesting strategy for treating the disease.⁹

*e-mail: rogerio@ufpa.br; fabioam@ufpa.br

Papadopoulou *et al.*^{10,11} synthesized and tested a series of derivatives of 3-nitro-*H*-1,2,4-triazole-based arylamide and arylsulfonamide (HA/S) class of compounds using structure-activity relationship (SAR) studies. The results demonstrated significant activity in *in vitro* tests against *T. cruzi*, and analogous compounds showed higher *in vivo* activity than benznidazole (reference compound).

Considering that CD is a neglected tropical disease, in this study, quantitative structure-activity relationship (QSAR) calculations were performed to obtain a regression model capable of predicting antichagasic activities, due to the molecular and structural information of the compounds, as well as to understand the action mechanism of HA/S derivatives against *T. cruzi*. In addition, the docking methods and MD, followed by energetic analyses of affinity between the compounds and the enzyme, were used to understand the molecular behavior of these ligands at the active site of the cruzain enzyme, highlighting the fundamental importance at all stages of the parasite's life cycle.

Methodology

QSAR 2D

QSAR studies were performed using 22 selected derivatives (CP1-CP22) of HA/S selected on a previous study¹¹ and experimental biological activities were converted into their corresponding pIC_{50} (or $-\log IC_{50}$, where IC_{50} is the concentration in mol L^{-1} that inhibited 50% of biological activity) before QSAR analysis. Table 1 shows the HA/S analogues and the respective pIC_{50} values.

Two-dimensional structures were built in MarvinSketch program¹² and all geometries were initially optimized by MOPAC 2016¹³ at PM7 semi-empirical level.¹⁴ Next, these obtained structures were exported to the software Gaussian 03¹⁵ and new optimizations were carried out at Hartree-Fock level (HF/6-31G(d,p)).¹⁶ Using the optimized structures the calculation of molecular descriptors, geometric, electronic and topological were made using different computer programs, such as Gaussian 03,¹⁵ e-Dragon 1.0,¹⁷ ALOGPS 2.1,¹⁸ KRAKENX¹⁹ and Pharmaceutical Data Exploration Laboratory 2.21 (PaDEL),²⁰ totaling 3986 calculated descriptors for each molecule. In this study, the whole data set was autoscaled along all the variables, i.e., normalized and centered on the mean, so that they could be compared to each other on the same scale.

Next, using QSAR modeling²¹ software a matrix obtained with descriptors were reduced by eliminating those that presented the absolute Pearson correlation coefficient ($|r|$) lower than 0.3 with the logarithm of the

biological activity (pIC_{50}).²² After, the resulting matrix was subjected in the QSAR modeling to the ordered predictors selection (OPS).²¹ Finally, we performed a new selection of variables by means of genetic algorithm with the QSARINS program.²³ These steps attempt to remove variables that were little related to biological activity and that did not have any relevant information to the construction of the regression equation by the partial least squares (PLS) method.

Molecular docking

The three-dimensional (3D) structure of the cruzain enzyme was obtained from the Protein Data Bank (PDB) database under code 1ME4 with a resolution of 1.2 Å.²⁴ The enzyme and crystallographic ligand were prepared in the Chimera program.²⁵ Five computational packages were selected for evaluation of molecular docking procedures.

Initially, the water molecules present in the enzyme were removed, hydrogen atoms were added to the structures and the charges were calculated according to each program. For calculations performed in the DOCK 6 program,²⁶ AM1-BCC (Austin Method 1 with Bond Charge Correction) charges were used, while for calculations in the AutoDock 4.2 (ADT)²⁷ and AutoDockVina 1.1.1 (VINA)²⁸ programs, Gasteiger charges²⁹ were used for the ligand and Kollman charges³⁰ for the target (enzyme). In the FITTED 2.4 (FT)³¹ and Molegro Virtual Docker 5.5 (MVD)³² programs, the standard program charges were used, such that after the preparation of each structure in their respective programs, they were saved separately for the docking study.

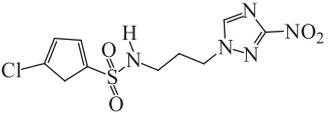
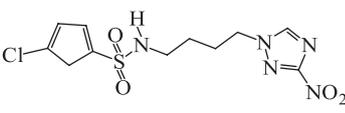
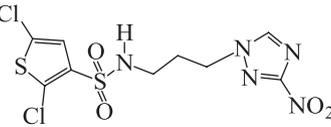
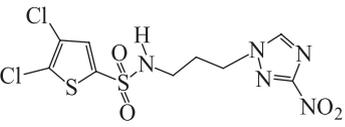
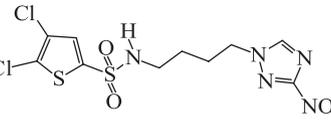
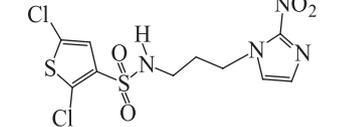
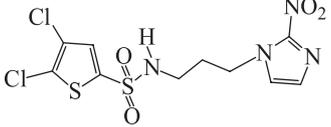
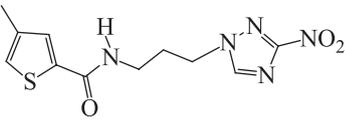
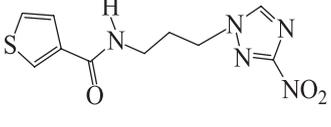
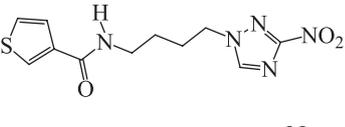
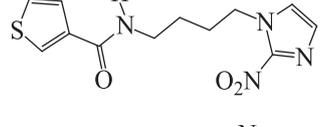
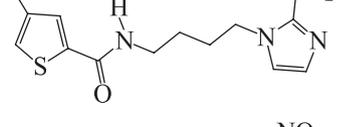
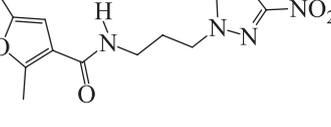
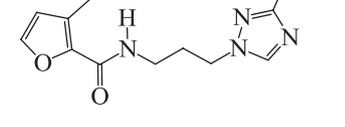
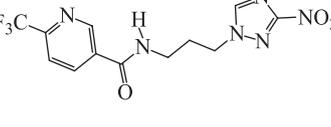
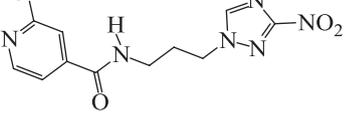
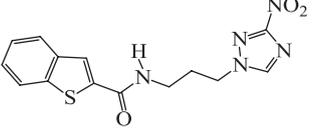
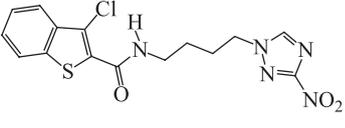
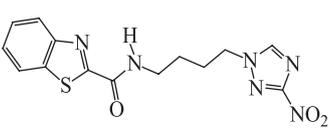
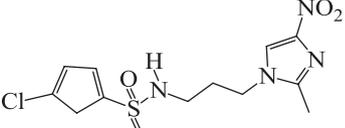
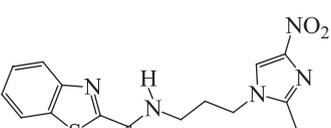
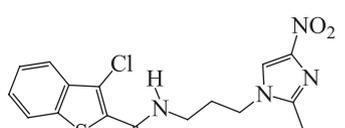
Redocking

To validate the packages used, the T10 crystallographic ligand was docked in the 1ME4 structure. To evaluate the convergence of the results, the procedure was performed in triplicate in five different programs (ADT, DOCK, VINA, MVD and FT), in which it was possible to determine the affinity energy, interactions between the receptor and the ligand, as well as the lowest root mean square deviation (RMSD) value using the fconv 1.24 program.³³ These parameters made it possible to select the program that best reproduced the interactions of the crystallographic ligand in the complex with the lowest RMSD value.

Molecular dynamics

The structures obtained on the molecular docking analysis were used as a starting point for classical molecular

Table 1. Selected dataset of 22 heteroarylamine/sulfonamide compounds (HA/S) and their respective pIC_{50} values

Number	Structure	pIC_{50}	Number	Structure	pIC_{50}
1		6.36	2		6.34
3		6.26	4		6.80
5		6.43	6		5.16
7		5.32	8		5.59
9		5.40	10		5.76
11		nd	12		nd
13		5.05	14		4.52
15		5.46	16		4.87
17		6.53	18		6.73
19		6.37	20		3.99
21		3.98	22		4.49

pIC_{50} : $-\log IC_{50}$, where IC_{50} is the concentration in mol L^{-1} for 50% growth inhibition; nd: not determined.

dynamics (MD) simulations by using two computational packages: Amber 16³⁴ and Q.³⁵ Initially, the protonation states of all amino acid residues of the enzyme were determined under pH 7.0 condition and analyzed by PDB2PQR/PROPKA server.³⁶ Then, molecular systems were built for each MD package as follows:

Amber system

On the Amber 16 package,³⁴ the all-atom force field Amber FF99SB³⁷ and general Amber force field (GAFF)³⁸ were selected as parameters set to describe protein and ligands in all MM (molecular mechanics) simulations, respectively. The restrained electrostatic potential (RESP) charges³⁹ of each ligand were calculated on the Gaussian 03¹⁵ package using the Hartree-Fock method with 6-31G* basis. The systems were solvated in a cubic water-box with the explicit solvation model TIP3P.⁴⁰ We used a distance of 12.0 Å between the cell wall and the solvated atoms of the system. Counter-ions were appropriately added to neutralize the system.

Initially, each solvated complex was energy-minimized by performing out a minimization (25,000 and 10,000 steps of each steepest descent and conjugate gradient approach, respectively) and then gradually heated up to 298 K (ensemble NVT) of MD simulations with a constraint of 25 kcal mol⁻¹ Å⁻² on the solute atoms of systems. After, each system was relaxed for 250 ps of MD simulations for density equilibration at constant temperature (298 K). All stages simulations employed a nonbonded cutoff of 9 Å, where particle mesh Ewald (PME) approach computed the long-range electrostatic forces. In all the simulations, the bonds involving hydrogen atoms were restricted by SHAKE algorithm⁴¹ and the motion equations were integrated at each 2 fs using Verlet algorithm.⁴² Finally, 50 ns of MM MD simulations were performed out for each system by using NPT ensemble. The SANDER module³⁴ of AmberTools 16 was used for all MM calculations. The structural analysis of each system was performed out by using RMSD and root mean square fluctuation (RMSF) plots obtained by CPPTRAJ module⁴³ of AmberTools 16.

Binding free energy calculations by using MM-Poisson-Boltzmann/generalized Born/surface area (PB(GB)/SA) approach

The last 10 ns of MM MD simulations of each system were used for binding free energy calculations. The snapshots were carefully extracted by using the CPPTRAJ module and submitted to the MMPBSA.py module⁴⁴ of AmberTools 16. The binding free energy values were computed by using MM-PB(GB)/SA approaches⁴⁵ as

implemented on the MMPBSA.py module. These methods have been used successfully on many biological systems^{46,47} as well as described elsewhere.⁴⁸ We used a single-trajectory MM-PB(GB)/SA protocol where protein-ligand complex, unbound protein, and free ligand conformation states are identical. In the equation 1, the free energy of binding (ΔG_{bind}) is approximated through:

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S \approx \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S \quad (1)$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdw}} \quad (2)$$

$$\Delta G_{\text{sol}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{(nonpolar)}} \quad (3)$$

From the equation 2 the term ΔE_{MM} refers to protein-ligand interactions on the gas phase, which are computed by internal energy (ΔE_{int}), electrostatic (ΔE_{ele}) and van der Waals (ΔE_{vdw}) contributions. ΔH is the enthalpy. The solvation free energy and entropic contributions are computed by ΔG_{sol} and $T\Delta S$ terms, respectively. The ΔG_{sol} (equation 3) is obtained computing polar ($\Delta G_{\text{PB/GB}}$) and nonpolar ($\Delta G_{\text{(nonpolar)}}$) electrostatic interactions by using implicit solvation models (PB or GB) and a linear relation to the solvent accessible surface area (SASA).⁴⁹

Q system

As described previously, molecular systems obtained from molecular docking calculations were used for classical MD simulations by using Q package,³⁵ where protein and ligands were described by all-atom force field OPLS (optimized potentials for liquid simulations).⁵⁰ Particularly, to compute the binding free energy, MD simulations were simulated in two different states: (i) ligand solvated only by water molecules and (ii) ligand bound to protein and solvated by water molecules. In each state, the systems were solvated in 18 Å radius water sphere with the explicit solvation model TIP3P⁴⁰ and subjected to polarization and radial constraints according to the surface-constrained all-atom solvent model^{35,51} to mimic the properties of bulk water at the sphere surface. A time step of 2.0 fs was used, and no positional restraints were applied. Solvent bonds and angles were constrained using the SHAKE algorithm.⁴¹

All classical MD simulations started with a heating and equilibration stage followed by subsequent data collection. All systems were thus gradually heated up to 298 K, and positional restraints on all solute heavy atoms were gradually released. Finally, each ligand was subjected to three MD simulations (for each state) during 3 ns. The trajectories and energy values were updated every 25 steps, and the same interval was used for the sampling of the ligand-surrounding interaction energies.

Linear interaction energy calculation

The binding free energy for each protein-ligand system was computed by using linear interaction energy (LIE) approach (ΔG_{LIE}).⁵² On this strategy, ΔG_{LIE} is calculated by the difference observed on (*i*) and (*ii*) states:

$$\Delta G_{LIE} = \alpha(\langle V_{vdw} \rangle_{complex} - \langle V_{vdw} \rangle_{free}) + \beta(\langle V_{ele} \rangle_{complex} - \langle V_{ele} \rangle_{free}) \quad (4)$$

where in the equation 4 the terms $\langle V_{vdw} \rangle$ and $\langle V_{ele} \rangle$ are the averages of the van der Waals and electrostatic interactions, respectively for each (*i*) and (*ii*) stated, scaled by the scaling factors α and β . The constant α has been empirically determined as 0.181, while β varies depending on the properties of the ligand.^{52,53} Here, α and β values used are 0.181 and 0.430, respectively.

Results and Discussion

QSAR 2D

Using the study by Papadopoulou *et al.*,¹¹ it was found that the biological activity values (pIC_{50}) varied from three logarithmic units, ranging from 3.98 (least potent compound) to 6.8 (most potent molecule). Figure 1 shows the histogram with the distribution of pIC_{50} values.

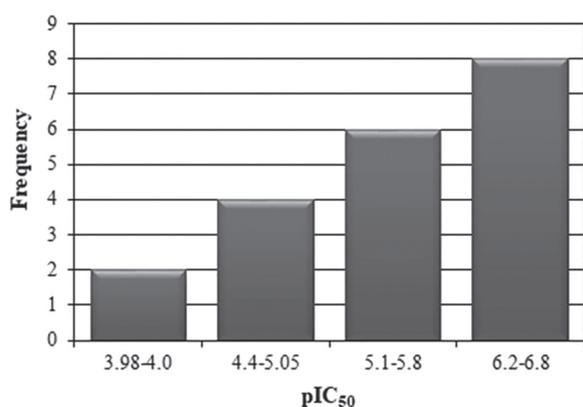


Figure 1. Histogram showing the distribution of the biological activity (pIC_{50}) in log units for HA/S compounds.

Subsequently, after determining the most stable conformations obtained by the HF/6-31G(d,p) method, several descriptors were calculated to form a data matrix. Thus, after calculating all the descriptors, the next step was to select those that were relevant (statistically significant) for the biological activity in question.

The compounds were divided into two sets called training and testing via the hierarchical cluster analysis (HCA) method. The analysis was carried out on autoscaled data by the Minitab 14 program,⁵⁴ using the Ward

linkage and Euclidean distance. For this, 15 compounds were randomly chosen for the training set (70% of the compounds), and the rest were used as a testing set (30% of the compounds) to verify the performance of the model. The following compounds were selected to comprise the testing set: 5, 9, 10, 14 and 21 and the other compounds were used for the training set.

After the cutoff by the correlation of ± 0.3 in relation to the biological activity, performed in the QSAR modeling program,²¹ of the 3986 descriptors initially calculated, a total of 385 remained. Using the OPS algorithm,⁵⁵ we arrived at a matrix of 25 descriptors and three latent variables (LVs). The use of the QSARINS program²³ allowed a reduction to the three final descriptors for the construction of the model.

The model was selected with two LVs containing 81.49% of accumulated information, in which 64.01% referred to the first LV and 17.48% referred to the second LV, and no anomalous sample was identified. Table 2 shows that the model was able to explain 93.2% and predict 88.1% of the total variance. The validation statistical parameters for the regression model, coefficient of determination (R^2) and the cross-validation coefficient of determination (Q^2_{LOO}), presented values above those recommended in the literature,⁵⁶ where $R^2 > 0.6$ and $Q^2_{LOO} > 0.5$.⁵⁶ When considering the values of R^2 and Q^2_{LOO} , the difference observed between them was 0.051, thus the model is within the limit suggested by Kiralj and Ferreira,⁵⁷

Table 2. Results of statistical parameters through cross-validation leave-N-out (LNO)

Statistical parameter	Value
R^2	0.932
Q^2_{LOO}	0.881
$R^2 - Q^2_{LOO}$	0.051
PRESSval	1.262
PRESScal	0.720
RMSECV	0.290
RMSEC	0.219
$F(n,n-k-1)$	82.24
$R^2_{adjusted}$	0.927
SS_y	10.57
n	15
LV	2

R^2 : coefficient of determination; Q^2_{LOO} : cross-validation coefficient of determination; PRESSval: predictive residual sum of squares of validation; PRESScal: predictive residual sum of squares of calibration; RMSECV: root mean square error of validation; RMSEC: root mean square error of calibration; $F(n,n-k-1)$: F test; $R^2_{adjusted}$: adjusted correlation coefficient; SS_y : sum of squares; n: number of compounds; LV: latent variables.

which is an indication that the model does not present data overfitting. Another important factor was the result of the F test value calculation (82.24), which was much higher than its tabulated reference value $F_{\text{tab}(2,12)} = 3.89$ with 95% significance ($\alpha = 0.05$), wherein it is more than five times larger than the tabulated value, where $F_{(2,12)}/F_{\text{tab}(2,12)} = 21.14$, which indicates that the information calibrated by equation 5 is really significant.

Table 2 presents the parameters for the results of the internal validation of the training set. In this step, the predictive residual sum of squares of validation (PRESSval) determines that the smaller the difference between the experimental value and the predicted pIC_{50} , the greater the predictive capacity of the model. The QSAR model proposed provides internal quality and statistical significance, because the PRESSval value is lower than the sum of squares of the response values (SSy), as indicated in Table 2.²² Similarly, the same can be observed for the sum of squares of the PRESScal calibration prediction residuals, which means the lower the value, the better the goodness of fit.^{57,58} The root mean square error of calibration (RMSEC) and root mean square error of validation (RMSECV), respectively also show better statistical quality of internal validation as the values approach zero, indicating a small deviation from the experimental value.⁵⁹

The adjusted correlation coefficient (R^2_{adjust}), which takes into account the number of compounds and latent variables in the QSAR model, shows that the regression equation has good explanatory capacity and since the difference between R^2 and R^2_{adjust} is less than 0.3, the number of descriptors present in the QSAR model is acceptable.⁶⁰

Among the three descriptors selected, two were obtained from the PaDEL program²⁰ (MATS7v and MDEN-23) and one from the e-Dragon program¹⁷ (R6e+).

Equation 5 presents the best regression model obtained for the training set according to the validation parameters:

$$\text{pIC}_{50} = -54.463 \times \text{R6e+} - 6.507 \times \text{MATS7v} + 0.438 \times \text{MDEN} + 5.806 \quad (5)$$

The non-scaled multivariate regression equation obtained from the self-scaled data in the QSAR modeling program is presented in equation 6.

$$\text{pIC}_{50} = -0.235 \times \text{R6e+} - 0.662 \times \text{MATS7v} + 0.259 \times \text{MDEN} \quad (6)$$

Equation 6 demonstrates that the MATS7v descriptor is the most important in the regression equation with the self-scaled regression vectors. In addition, it also shows

that the R6e+ and MATS7v descriptors have a negative contribution to biological activity; on the other hand, the MDEN-23 descriptor has a positive contribution.

One factor that can be assessed in the model is the coincidence between the algebraic signs of Pearson correlation coefficient values (r) for each descriptor with the pIC_{50} values (Table 3) and the signs of the regression coefficients in the model (equation 6). Table 3 shows that the model has descriptors whose sign of their coefficients coincides with the information provided by the correlation of the biological activity, which demonstrates the self-consistency of the model.⁶¹

Table 3. Standardized coefficients of the model and Pearson correlation coefficients (r)

Descriptor	Standardized coefficient	r
R6e+	-0.235	-0.729
MATS7v	-0.662	-0.904
MDEN-23	0.259	0.629

In the leave-N-out (LNO) validation test, it is recommended that about 20 to 30% of the number of samples be removed, and 5 samples (ca. 20%) were removed in this work. Figure 2 presents the LNO results, the mean and the standard deviation for the values of Q^2_{LNO} obtained for the three repetitions, removing from 1 to 5 samples (leave-5-out).

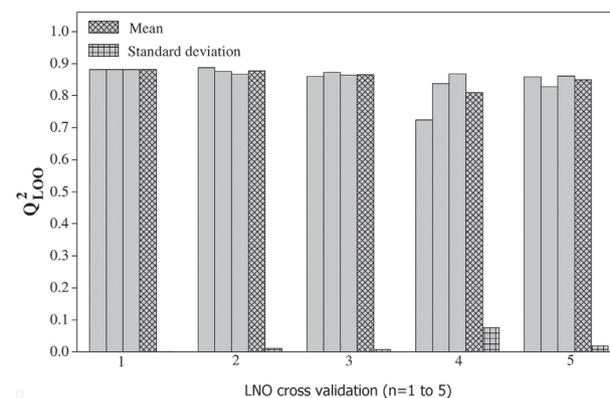


Figure 2. LNO cross-validation plot.

The LNO test shows that when removing up to five samples, the QSAR model built can be classified as robust, because it has high Q^2_{LNO} values, average Q^2_{LNO} values were close to the individual ones, and the difference of each Q^2_{LNO} value with respect to the Q^2_{LOO} value was not higher than 0.1, besides the standard deviations found were also below 0.1.⁵⁷

Figure 3 shows the result of the y randomization test, used to measure the possibility of chance correlation. An analysis of Figure 3 shows that the intercept of the regressions for R^2 and Q^2_{LOO} were below 0.3 and 0.05, respectively, indicating that the variance explained by the models is not due to chance correlation. Thus, the model can be considered robust.^{57,62}

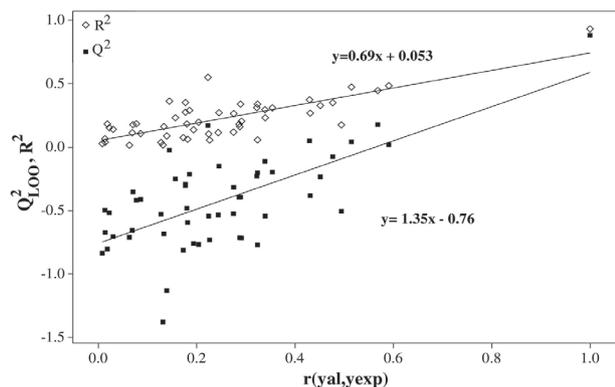


Figure 3. Results obtained for y -randomization test.

In Figure 4, from the 50 randomizations, it can be observed that all values obtained for R^2 and Q^2 were below 0.6 and 0.5, respectively. This fact shows that the variance explained by the model was not due to chance correlation.²²

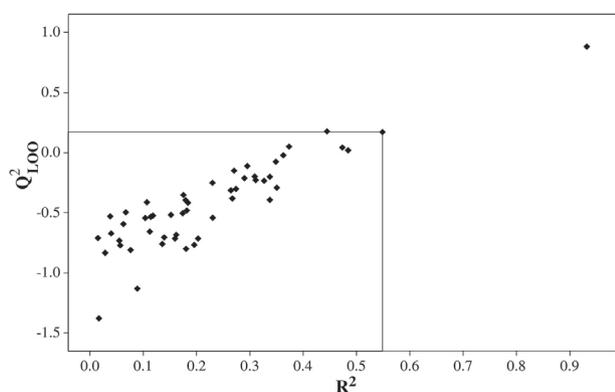


Figure 4. y -Randomization test.

An analysis of the Figure 5 chart of leverage by Student residuals shows the absence of compounds with anomalous behavior (outliers). Note that only compound 22 has a high leverage value, but a low Student residual value, and thus cannot be classified as an anomalous sample, such that removing it would cause a decrease in the quality parameters of the model.

The quality of the model is also evaluated according to the statistical parameters presented in Table 4 for the external prediction set, in which activity values are predicted for a set, called a “testing” set, made up of known

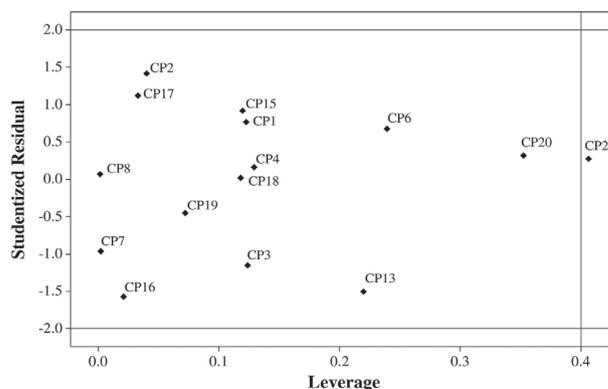


Figure 5. Student residuals *versus* leverage values plot.

Table 4. Statistical parameters to external validation data set

Parameter	Value
R^2_{pred}	0.963
SEP	0.191
$ARE_{pred} / \%$	3.36
k	1.004
k'	0.994
$ R^2_0 - R'^2_0 $	7.9×10^{-3}
$(R^2 - R^2_0) / (R^2)$	0.011
$(R^2 - R'^2_0) / (R^2)$	3.136×10^{-3}

R^2_{pred} : coefficient of determination of external validation; SEP: standard error of prediction; ARE_{pred} : average relative error of prediction; k , k' : Golbraikh-Tropsha slopes of the linear regression lines between the observed and the predicted activities in the external validation; $|R^2_0 - R'^2_0|$: Golbraikh-Tropsha absolute values of the difference of the determination coefficient of the linear relation between the observed and predicted values without an intercept (R^2_0), and the predicted and observed values without an intercept (R'^2_0); R^2 : coefficient of determination.

compounds that are not used for the construction of the regression model.

Table 4 shows that the model has good external prediction capacity, as it presents the value of R^2_{pred} (coefficient of determination of external validation) above 0.6, showing a good fit for the data.⁶³ The standard error of prediction (SEP) and average relative error of prediction (ARE_{pred}) showed low values, indicating a low prediction error.⁶⁰

Table 4 demonstrates that the Golbraikh and Tropsha parameters, based on straight-line inclinations, which were obtained by regression between observed and predicted values (k) and between predicted and observed (k') must be between 0.85 and 1.15 and the modulus of the difference between R^2_0 (determination coefficient of the linear relation between the observed and predicted values without an intercept) and (R'^2_0) (predicted and observed values without an intercept) less than 0.3.⁶⁴ Finally, the fact that the last two parameters are less than 0.1 shows that the proposed QSAR model has excellent predictive ability.⁶⁵

Thus, according to the internal validation parameters, it was found that the model can be classified as a good model, since according to the criteria used, the model did not present overfitting, had a small deviation from the experimental, is self-consistent and robust, does not suffer from chance correlation and shows good external prediction ability.

In a QSAR model, it is desirable to evaluate the selected molecular descriptors to assist in understanding the action mechanism for the development of new molecules that act as new therapeutic agents. As such, some conclusions on the variables can be drawn for the studied HA/S derivatives.

The R6e+ descriptor emphasizes the relevance of the electronic effects and takes into account Sanderson's electronegativity among a group of 6 atoms.⁵⁵ Equation 5 shows the inversely proportional relationship to biological activity, and the smaller the difference in electronegativity between atoms, the better the activity, explaining the activity of the CP1, CP2, CP4, CP5, CP7, CP10, CP13, CP18 and CP19 compounds.

The MATS7v two-dimensional autocorrelation descriptor expresses van der Waals volume values of a 7-atom group, thus determining the relationship between atomic volume and biological activity.⁶⁶ In the set of compounds, it is observed that the lower the value of the atomic volumes, the higher the biological activity when observed in equation 5. The inversely proportional relationship of the MATS7v descriptor and the biological activity can also be seen, proving that the most active CP4 and CP18 compounds present the lowest values of this property.

The MDEN-23 topological descriptor refers to the geometric mean of the molecular distances between all secondary and tertiary nitrogen, and this encodes the presence or absence of nitrogen containing substituents attached to a common skeleton.⁶⁷ The proportional relationship of the MDEN-23 descriptor and the biological

activity is observed, proving that the most active compounds present the highest values of this property.

Through the results obtained in the QSAR study, equations 5 and 6 and the properties that can contribute to the biological activity, it was possible to propose five compounds derived from the HA/S class, named LMM1 to LMM5, to integrate into the set of molecules.

The LMM1 compound was built with only one chlorine atom attached to the heterocyclic ring and maintaining the structure of compound 3 proposed by Papadopoulou *et al.*¹¹ In contrast, the LMM2 to LMM4 compounds were constructed by substituting the nitro group with the amino to avoid possible toxic activities and increasing the number of carbon atoms between the heterocyclic rings with the intention of decreasing the value of the most influential descriptor obtained in the regression equation with the self-scaled regression vectors (MATS7v).

The LMM5 compound was designed to be of the lowest biological activity by replacing the chlorine atoms of the heterocyclic ring with the bromine atom in order to increase the values of the MATS7v and R6e+ descriptors. Next, it was possible to calculate their properties and predict their respective biological activity values using equation 5, as shown in Table 5 and Figure 6.

Table 5. Descriptors selected and the predicted values by the PLS model for proposed compounds derived from the HA/S class

Compound	R6e+	MATS7v	MDEN-23	pIC ₅₀ / (μmol L ⁻¹)
LMM1	0.045	-0.100	2.435	5.06
LMM2	0.047	-0.135	1.209	4.65
LMM3	0.036	-0.214	1.155	5.74
LMM4	0.023	-0.277	1.310	6.93
LMM5	0.049	0.388	1.462	1.24

pIC₅₀: -log IC₅₀, where IC₅₀ is the concentration in mol L⁻¹ for 50% growth inhibition.

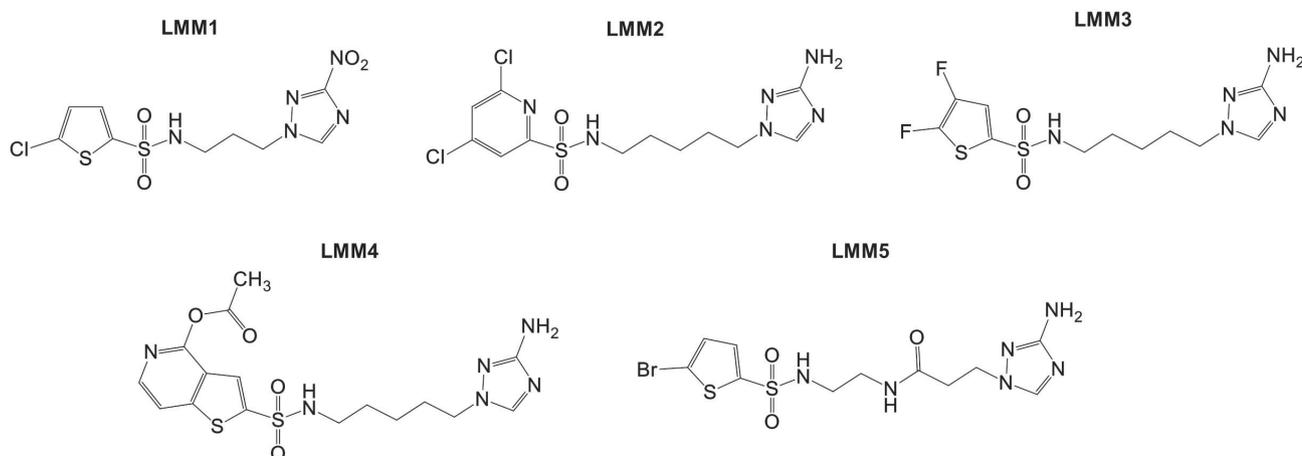


Figure 6. Proposed compounds derived from the HA/S class.

Thus, Table 5 demonstrates that the predicted biological activity values for the LMM1 and LMM3 compounds showed intermediate values to the pIC_{50} values, whereas LMM4 shows a higher activity against *T. cruzi* than compound **4** elaborated by Papadopoulou *et al.*¹¹ (Table 1). This is mainly due to the contribution of the MAT7v descriptor to the construction of the model. In addition, it was possible to predict a derivative with a low pIC_{50} value, since the carbon chain was increased, the distances between nitrogens were reduced and a bromine atom was added.

Molecular docking

The study performed by Papadopoulou *et al.*¹¹ verified the HA/S activity against the amastigote form of *T. cruzi*, and that cruzain, expressed in all evolutionary forms of the parasite, acts in the invasion, replication and differentiation stages, and is therefore a key enzyme in the search for inhibitors against Chagas disease.^{6,7} Thus, due to their importance, the action of these compounds on the parasite's cruzain enzyme was investigated.

For the determination of the most suitable program and procedures for molecular docking calculations of the 27 HA/S derivatives (CP1-CP22 and LMM1-LMM5), the enzyme structure was recovered from the PDB under

code 1ME4.²⁴ Molecular redocking of the crystallographic ligand (T10) for program selection was performed in five different programs, such as DOCK 6,²⁶ ADT,²⁷ VINA,²⁸ FT³¹ and MVD.³² RMSD values were evaluated using the fconv 1.24 program.³³

After the redocking of the co-crystallized ligand (T10) in triplicate in the five programs, it was observed that the FT program presented the lowest RMSD value (equal to 2.91), and thus was chosen to dock the 27 compounds in the 1ME4 enzyme.

The interactions of the 27 compounds were analyzed in the PoseView online server.⁶⁸ In order to select the compounds against the cruzain enzyme of *T. cruzi*, the structures were classified according to the interaction energy values obtained by the FT, as well as a visual inspection taking into account the hydrogen bonds, hydrophobic interactions, and the biological activity (pIC_{50}) values obtained. Table 6 and Figure 7 show that CP2 and LMM4 stood out for their high interaction energy and pIC_{50} values, as well as their hydrogen bond interactions with the main active site residuals that are part of the catalytic triad, such as Cys25, Gly66, Asn69 and His159, as well as hydrophobic interactions with the Gly65, Leu67, Ala133, and Glu205 residuals.

In the conformation obtained by the molecular docking for CP2, this compound showed hydrogen bonds with the

Table 6. Hydrogen bond (HB), hydrophobic interactions (HI), pIC_{50} values and docking energy (DE)

Compound	HB ^a	HI ^b	pIC_{50} ^c	DE ^d / (kcal mol ⁻¹)
CP2	Cys25 and Gly66	Cys25	6.34	-20.96
LMM4	Gly66, Asn69 and His159	Cys25, Gly65, Leu67, Ala133 and Glu205	6.93	-20.42

^aHydrogen bonds; ^bhydrophobic interactions; ^cbiological activity, $-\log IC_{50}$, where IC_{50} is the concentration in mol L⁻¹ for 50% growth inhibition; ^ddocking energy.

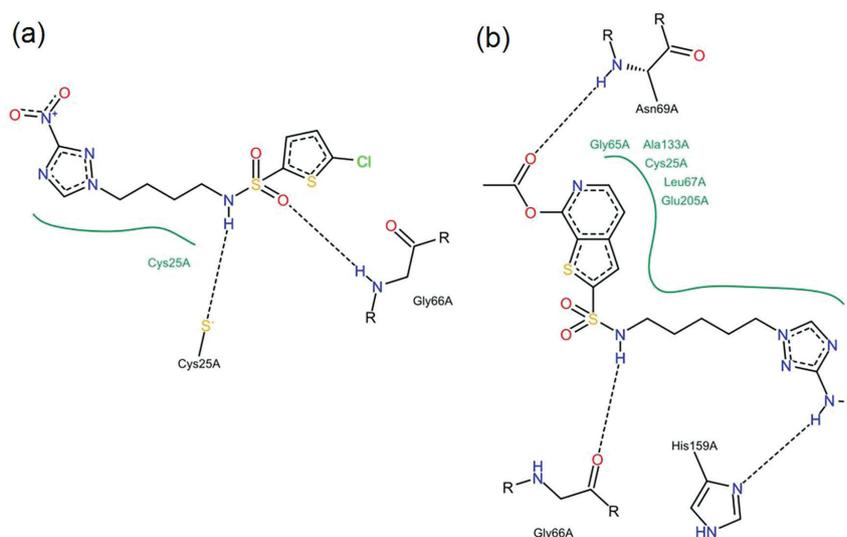


Figure 7. Interactions predicted from docking simulations between each compound and cruzain enzyme: (a) CP2 and (b) LMM4. Hydrogen bonds are shown in dashed black lines and hydrophobic interactions are in green. Images obtained from PoseView Web 1.97.0.⁶⁸

residues of amino acid Cys25 and Gly66 (Figure 7). The Cys25 belongs to the active cruzain enzyme site, and this residue is located in the S1 subsite of the enzyme, while the Gly66 is located in the S3 subsite, highlighting that the Cys25 residual is part of the enzyme catalytic triad.^{24,69}

In the conformation obtained for the LMM4 ligand, it is possible to observe hydrogen interactions with the Gly66, Asn69 and His159 residues. The His159 is part of the catalytic triad and the amino acid residues Leu67, Asn69, Ala133 and Glu205 are located in the S2 subsite, and according to a study by Durrant *et al.*,⁷⁰ it appears that these flexible residuals can be explored in the drug design process. In addition, a study by Gillmor *et al.*⁷¹ showed that the Glu205 is important because it adjusts itself to restructure part of the active site, conferring a strong interaction with basic or hydrophobic inhibitors in the S2 subsite.

After docking the 27 compounds in the cruzain, the HA/S derivatives were docked to the homologous human cathepsin L enzyme recovered from the PDB under code 5MAE.⁷² Thus, docking calculations showed that these compounds did not interact with the 5MAE enzyme when subjected to interaction analysis on the PoseView online server.⁶⁸

Molecular dynamics

Experimental and computational studies^{73,74} have been performed to verify the specificity and characterize the cruzain enzyme site in order to facilitate directed drug design. Free energy calculations have been increasingly used to provide quantitative predictions in order to prioritize compounds based on their estimated activity.

After molecular docking simulation, CP2 and LMM4 were selected for MD calculations because these ligands present hydrogen bonds with key residues, in addition to lower energy values obtained by docking with the enzyme of *T. cruzi* and pIC₅₀ values predicted by the PLS model above 6 μ M.

The structural stability of the systems was analyzed using RMSD values as a function of simulation time (Figure 8). For the systems evaluated in the Amber program,³⁴ it is observed that in all systems, the RMSD values are less than 2.2 Å, showing few conformational changes in the complex's structure. In addition, an analysis of Figure 8 shows that the most stable complex is that referring to the LMM4 ligand, which has a lower RMSD than CP2. Similar results were also found for the same systems in the Q program (Supplementary Information).

The flexibility of protein regions with respect to each ligand was observed by RMSF. Figure 9 shows that the regions corresponding to the amino acid ranges Val54-Leu67, Glu95-His106 and Thr148-Gln159, which

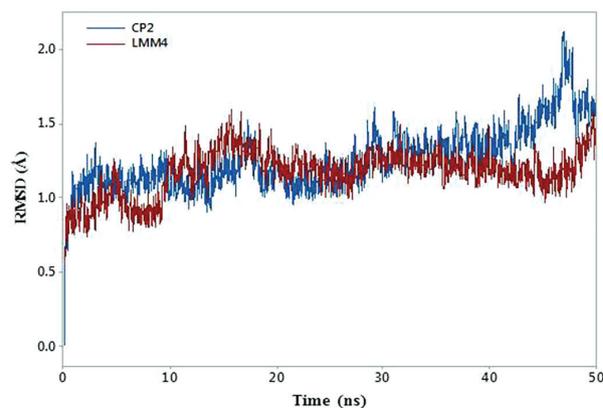


Figure 8. Graphical representation of the RMSD values versus simulation time for the complex CP2 (blue line) and LMM4 (red line) studied.

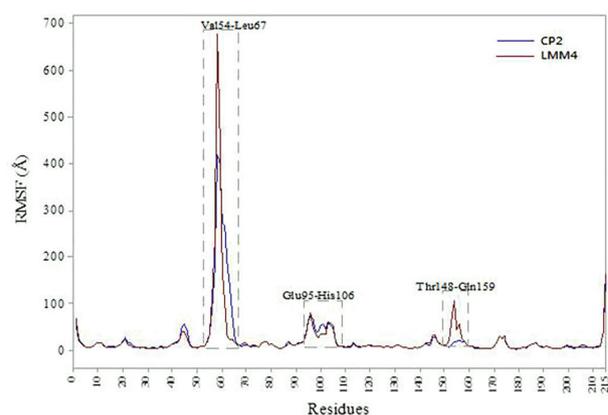


Figure 9. Graphical representation of the structural fluctuations taking into account residues of complex CP2 (blue line) and LMM4 (red line) from the MD simulations.

correspond to the loop regions, showed greater fluctuations. It is also worth highlighting the region between the residuals Val54-Leu67, which showed significant deviations from the two ligands. Studies conducted by Wiggers *et al.*⁶⁹ show the importance of these residuals located in the S3 subsite of the cruzain enzyme, which demonstrates their importance in the molecular recognition process and, consequently, in inhibiting the enzyme.

MD simulations are an important tool for including receiver flexibility and calculating the binding free energy value, at which this value can be used to quantify the affinity between an enzyme-inhibitor type system. Thus, the calculation of binding free energy values is an important tool in the drug design process, and is considered a key factor in proposing powerful enzyme inhibitors.⁴⁷ In this context, to evaluate the affinity of the CP2 and LMM4 compounds in the cruzain enzyme of *T. cruzi*, the binding free energy values were determined by the MM-GB(PB)SA and LIE methods in the Amber 16³⁴ and Q programs,³⁵ respectively.

The calculation of the binding free energy for the complexes formed between CP2 and LMM4 was performed

Table 7. Binding free energies with standard deviation computed by the MM-GB(PB)SA and LIE methods

Complex	$\Delta G_{\text{MM-GBSA}}^a$ / (kcal mol ⁻¹)	$\Delta G_{\text{MM-PBSA}}^b$ / (kcal mol ⁻¹)	ΔG_{LIE}^c / (kcal mol ⁻¹)
Cruzain-CP2	-19.14 ± 3.23	-21.93 ± 3.70	-11.95 ± 0.86
Cruzain-LMM4	-29.61 ± 3.85	-26.55 ± 3.94	-14.48 ± 0.78

^aBinding free energy values obtained from MM-GBSA (molecular mechanics with generalized Born surface area) method; ^bbinding free energy values obtained from MM-PBSA (molecular mechanics with Poisson-Boltzmann surface area) method; ^cbinding free energy values obtained from LIE (linear interaction energy) method.

using the last 10 ns of the MD trajectories for the CP2 and LMM4 complexes in the Amber 16 program³⁴ (Table 7). Table 7 shows that the formation of the two complexes formed by the enzyme and the ligands is favorable. Furthermore, comparing the binding free energy values (MM-GB(PB)SA) shows that the LMM4 ligand presented the lowest energy values when compared to the CP2 proposed by Papadopoulou *et al.*¹¹

The values found in Table 7 show that the formation of the two complexes was also favorable by the LIE method, and once again the complex formed by the LMM4 ligand, proposed in the QSAR model, presented a lower energy value, that is, a higher affinity than the CP2 compound. Therefore, for the three free energy calculation methods

(MM-GB(PB)SA and LIE), the results corroborate each other.

The energy contribution of each residue for the complexes formed by the CP2 and LMM4 compounds obtained by the MM-GBSA method are shown in Figure 10. This energy decomposition analysis shows the residues that contribute most significantly to the total interaction energy and, therefore, to the stabilization of the complex formed by ligand CP2 are Cys25, Trp26, Gly65, His162 and Gly163. Other favorable interactions were with residuals Ser24, Gly66, Leu67, Ala138 and Leu160. For the complex formed by LMM4, there is a greater number of residuals that contribute to the stabilization of the complex. In addition, to those previously mentioned,

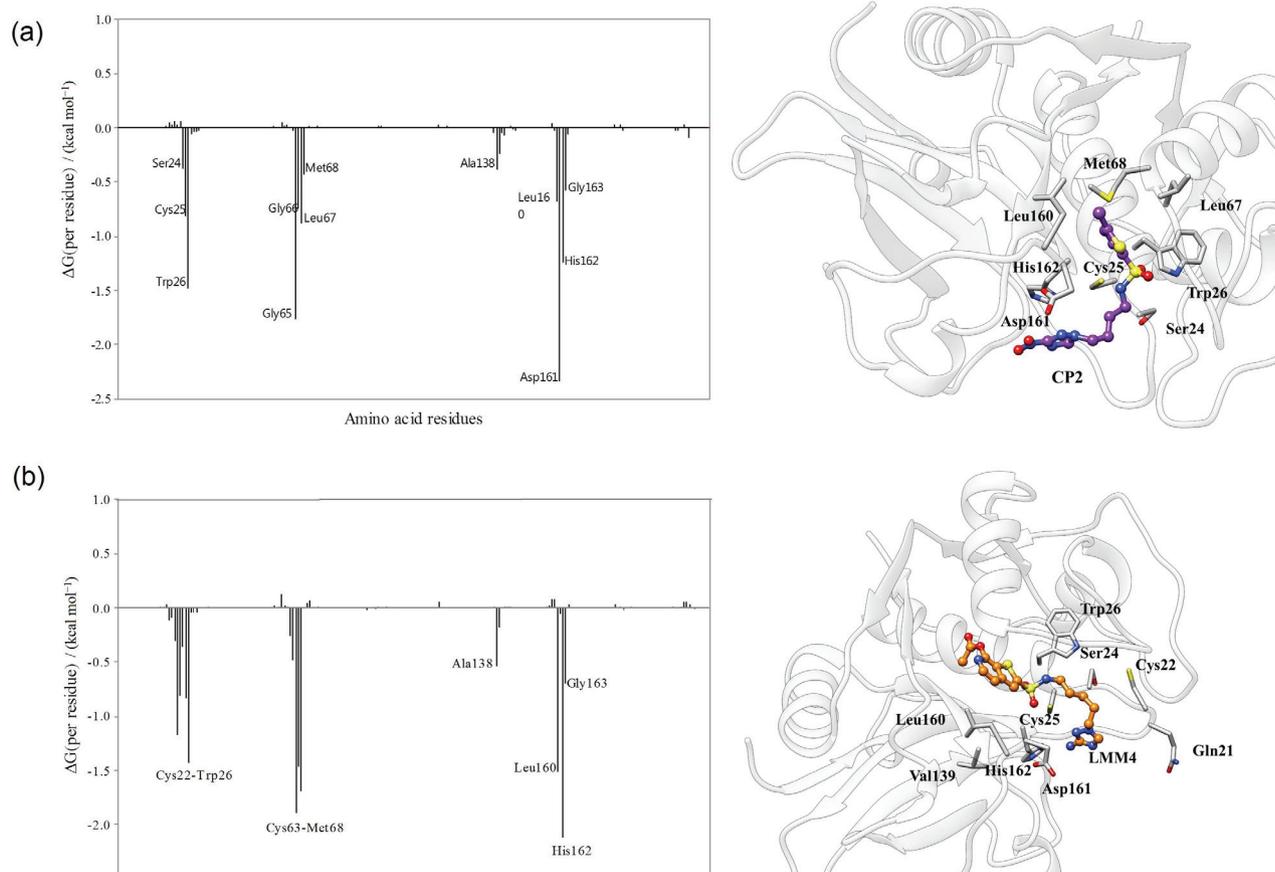


Figure 10. Graphical representation of the interaction energy *per* residue (left) for the complex formed (a) CP2 and (b) LMM4, obtained by Amber 16 program.³⁴

Gln21, Cys22, Gly23, Cys63, Ser64, Met68 and Val139 residues were important for LMM4. It is important to highlight that the interactions with residuals Cys25, Gly65, Gly66 and Leu67 had already been observed in the molecular docking results. It is noteworthy that the catalytic triad formed by the Gly23, Cys25 and Gly65 residuals are responsible for the trypanosomide activity observed.⁷ Furthermore, a greater interaction of the LMM4 ligand with the amino acid residues of the cruzain enzyme can be observed in the S1 (Gly23, Cys25 and Trp26), S3 (Ser64, Gly65, Gly66 and Gly67) and S1' (His162) subsites, which were important for fixation and stability of the inhibitor at the active site, results that corroborate studies by Bryant *et al.*⁷⁵

Figure 11 shows the energy contribution of each residue calculated by Q program,³⁵ and from these results is verified the stabilization of the system formed by the CP2 and LMM4 ligands with the enzyme (Figures 11a and 11b). Figures 11a-11b show that residuals Gly23, Ser24, Cys25, Trp26, Gly65, Gly66, Leu67, Met68, Leu160, Asp161, His162 and Gly163 contributed favorably. Thus, the results found by the LIE method corroborate those found by the

MM-GBSA for compounds CP2 and LMM4. Interestingly, the LMM4 compound made an important interaction with the Glu208 residue (Figure 11b), located at the end of the S2 subsite, which contributes to the flexibility of this subsite to accommodate volumous inhibitors in this region.^{71,76} Thus, despite the different approaches used to evaluate the proposed systems, the results indicated that the LMM4 obtained by the QSAR model does in fact present more effective interactions in relation to CP2, and thus can be indicated as a potential inhibitor of the cruzain enzyme of *T. cruzi*.

The QSAR model obtained made the proposition of the LMM4 ligand possible for the purpose of improving trypanocidal activity. Besides, the results of docking and MD showed that the LMM4 compound had low RMSD values, lower free energy value when compared to the structure obtained in the study by Papadopoulou *et al.*¹¹ Also, the LMM4 showed a greater number of residues, using different methods of MD, that contribute to the stabilization of the enzyme-inhibitor complex. Thus, it was possible to propose a promising molecule that may lead to an increase in the inhibitory capacity of the derivatives against *T. cruzi*.

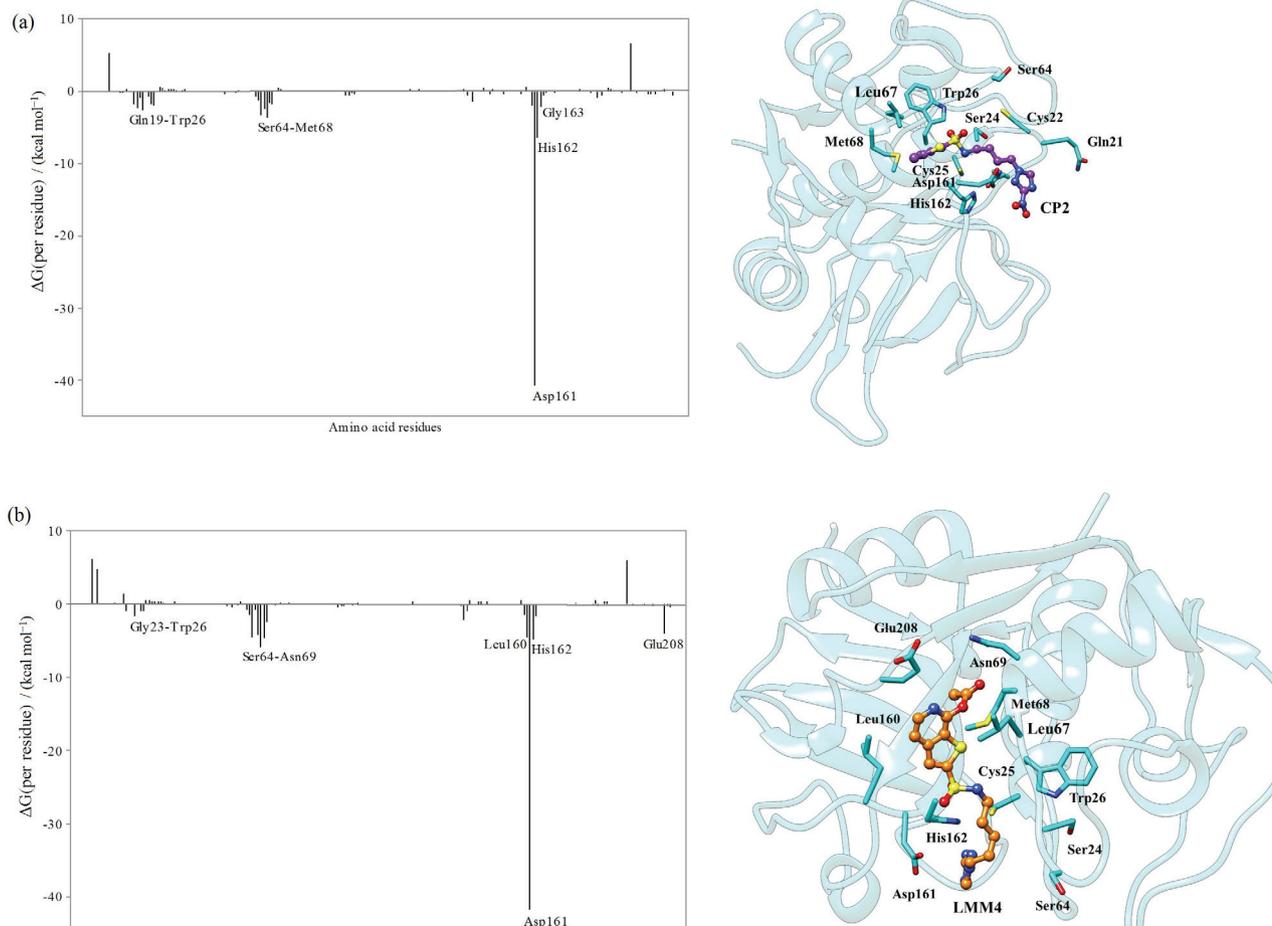


Figure 11. Graphical representation of the interaction energy *per* residue (left) for the complex formed (a) CP2 and (b) LMM4, obtained by Q program.³⁵

Conclusions

In this study, we performed out a molecular modeling study on derivatives of HA/S molecules as cruzain inhibitors using QSAR, molecular docking, MD simulations and binding free energy calculations. The QSAR method was employed by using PLS method, and the proposed molecular models were validated by using internal and external statistics parameters set, which show high quality and suitable prediction capacity. Besides, the final models were satisfactory according to the statistical validation results, where the PLS method leads to propose five new HA/S derivatives.

Molecular docking method was then applied to explain the results of QSAR and explore the interactions between new HA/S derivatives and catalytic site of cruzain enzyme, a potential target against Chagas disease. The molecular docking results suggest that amino acid residues such as Gln19, Cys25, Gly66, Leu157, Asp158, Glu205 and Ser208 were determined as the key residues for inhibitory activity.

The MD simulations and binding free energy calculations were computed on different packages and approaches, showed that the MD results were consistent with the findings obtained from QSAR and molecular docking results, and they suggested that LMM4 compound is more stable than inhibitor (CP2) when complexed to cruzain enzyme. In addition, the binding free energy decomposition pointed out the key residues involved in the binding of LMM4. Therefore, we could have a better understanding of the QSAR and the binding features of new HA/S derivatives using the molecular docking calculations and five new compounds with potential cruzain activity were obtained.

Finally, molecular modeling techniques such as QSAR, molecular docking, and MD simulations were successfully applied to design and to predict the biological activity of the new LMM4 compound against the cruzain enzyme. These results can be used on drug design field of new HA/S derivatives as potent inhibitor of cruzain from *T. cruzi*.

Supplementary Information

RMSD plots of MD simulations performed out by Amber and Q programs for CP2 and LMM4 systems are available free of charge at <http://jbcs.sbq.org.br> as PDF file.

Acknowledgments

The authors acknowledge the PROPESP/UFPA (Pró-Reitoria de Pesquisa e Pós-Graduação da UFPA) for the

financial support to improve the quality of this article. J. R. A. S. thanks to National Council for Scientific and Technological Development (CNPQ, grant 407096/2016-7) for their financial support and South African Centre for High-Performance Computing (<https://www.chpc.ac.za/>) and University of Florida Research Computing (<http://researchcomputing.ufl.edu>) for providing computational resources.

Author Contributions

André L. P. F. Souza and Fábio J. B. Cardoso conducted QSAR and molecular docking calculations; Lucas S. Martins performed out MD simulations and binding free energy calculations; Cláudio N. Alves, José R. A. Silva and Fábio A. Molfetta were responsible for the project administration, data analysis and wrote the manuscript. All authors read and approved the final manuscript.

References

1. Pauli, I.; Ferreira, L. G.; de Souza, M. L.; Oliva, G.; Ferreira, R. S.; Dessoay, M. A.; Slafer, B. W.; Dias, L. C.; Andricopulo, A. D.; *Future Med. Chem.* **2017**, *9*, 641.
2. Martins-Melo, F. R.; Lima, M. D.; Ramos, A. N.; Alencar, C. H.; Heukelbach, J.; *Trop. Med. Int. Health* **2014**, *19*, 943.
3. Ferreira, R. S.; Dessoay, M. A.; Pauli, I.; Souza, M. L.; Krogh, R.; Sales, A. I. L.; Oliva, G.; Dias, L. C.; Andricopulo, A. D.; *J. Med. Chem.* **2014**, *57*, 2380.
4. Cianni, L.; Sartori, G.; Rosini, F.; de Vita, D.; Pires, G.; Lopes, B. R.; Leitão, A.; Burtoloso, A. C. B.; Montanari, C. A.; *Bioorg. Chem.* **2018**, *79*, 285.
5. Urbina, J. A.; *Curr. Pharm. Des.* **2002**, *8*, 287.
6. Saraiva, A. P. B.; Miranda, R. M.; Valente, R. P. P.; Araujo, J. O.; Souza, R. N. B.; Costa, C. H. S.; Oliveira, A. R. S.; Almeida, M. O.; Figueiredo, A. F.; Ferreira, J. E. V.; Alves, C. N.; Honorio, K. M.; *Chem. Biol. Drug Des.* **2018**, *92*, 1475.
7. da Silva-Junior, E. F.; Franca, P. H. B.; Ribeiro, F. F.; Mendonca-Junior, F. J. B.; Scotti, L.; Scotti, M. T.; de Aquino, T. M. D.; de Araujo-Junior, J. X.; *Curr. Comput.-Aided Drug Des.* **2018**, *14*, 68.
8. Ferreira, L. G.; Andricopulo, A. D.; *Pharmacol. Ther.* **2017**, *180*, 49.
9. dos Santos, A. M.; Cianni, L.; de Vita, D.; Rosini, F.; Leitão, A.; Laughton, C. A.; Lameira, J.; Montanari, C. A.; *Phys. Chem. Chem. Phys.* **2018**, *20*, 24317.
10. Papadopoulou, M. V.; Bloomer, W. D.; Rosenzweig, H. S.; Chatelain, E.; Kaiser, M.; Wilkinson, S. R.; McKenzie, C.; Ioset, J. R.; *J. Med. Chem.* **2012**, *55*, 5554.
11. Papadopoulou, M. V.; Bloomer, W. D.; Rosenzweig, H. S.; Wilkinson, S. R.; Kaiser, M.; *Eur. J. Med. Chem.* **2014**, *87*, 79.

12. <https://chemaxon.com/presentation/marvin-the-next-generation-of-chemical-drawing>, accessed in July 2020.
13. Stewart, J. J. P.; *MOPAC 2016*; Stewart Computational Chemistry, Colorado Springs, USA, 2016.
14. Stewart, J. J. P.; *J. Mol. Model.* **2013**, *19*, 1.
15. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; 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.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; *Gaussian 03 User's Reference*; Gaussian Inc., Pittsburgh, USA, 2003.
16. Roothaan, C. C. J.; *Rev. Mod. Phys.* **1960**, *32*, 179.
17. Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V.; Radchenko, E.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V.; *J. Comput.-Aided Mol. Des.* **2005**, *19*, 453.
18. <http://www.vcclab.org/lab/alogps/>, accessed in July 2020.
19. Venkatraman, V.; Alsberg, B. K.; *J. Mol. Model.* **2016**, *22*, 93.
20. Yap, C. W.; *J. Comput. Chem.* **2011**, *32*, 1466.
21. Martins, J. P. A.; Ferreira, M. M. C.; *Quim. Nova* **2013**, *36*, 554.
22. de Melo, E. B.; Martins, J. P. A.; Jorge, T. C. M.; Friozi, M. C.; Ferreira, M. M. C.; *Eur. J. Med. Chem.* **2010**, *45*, 4562.
23. Gramatica, P.; Cassani, S.; Chirico, N.; *J. Comput. Chem.* **2014**, *35*, 1036.
24. Huang, L.; Brinen, L. S.; Ellman, J. A.; *Bioorg. Med. Chem.* **2003**, *11*, 21.
25. 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.
26. Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D.; *RNA* **2009**, *15*, 1219.
27. 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.
28. Trott, O.; Olson, A. J.; *J. Comput. Chem.* **2010**, *31*, 455.
29. Gasteiger, J.; Marsili, M.; *Tetrahedron* **1980**, *36*, 3219.
30. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P.; *J. Am. Chem. Soc.* **1984**, *106*, 765.
31. Corbeil, C. R.; Englebienne, P.; Moitessier, N.; *J. Chem. Inf. Model.* **2007**, *47*, 435.
32. Thomsen, R.; Christensen, M. H.; *J. Med. Chem.* **2006**, *49*, 3315.
33. Neudert, G.; Klebe, G.; *Bioinformatics* **2011**, *27*, 1021.
34. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J.; *J. Comput. Chem.* **2005**, *26*, 1668.
35. Marelus, J.; Kolmodin, K.; Feierberg, I.; Aqvist, J.; *J. Mol. Graphics Modell.* **1998**, *16*, 213.
36. Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A.; *Nucleic Acids Res.* **2007**, *35*, W522.
37. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C.; *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712.
38. Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A.; *J. Comput. Chem.* **2004**, *25*, 1157.
39. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A.; *J. Phys. Chem.* **1993**, *97*, 10269.
40. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.; *J. Chem. Phys.* **1983**, *79*, 926.
41. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C.; *J. Comput. Phys.* **1977**, *23*, 327.
42. Verlet, L.; *Phys. Rev.* **1968**, *165*, 201.
43. Roe, D. R.; Cheatham, T. E.; *J. Chem. Theory Comput.* **2013**, *9*, 3084.
44. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E.; *J. Chem. Theory Comput.* **2012**, *8*, 3314.
45. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E.; *Acc. Chem. Res.* **2000**, *33*, 889.
46. Homeyer, N.; Gohlke, H.; *Mol. Inf.* **2012**, *31*, 114.
47. Ganesan, A.; Coote, M. L.; Barakat, K.; *Drug Discovery Today* **2017**, *22*, 249.
48. Genheden, S.; Ryde, U.; *Expert Opin. Drug Discovery* **2015**, *10*, 449.
49. Hou, T. J.; Wang, J. M.; Li, Y. Y.; Wang, W.; *J. Chem. Inf. Model.* **2011**, *51*, 69.
50. Jorgensen, W. L.; Tiradorives, J.; *J. Am. Chem. Soc.* **1988**, *110*, 1657.
51. King, G.; Warshel, A.; *J. Chem. Phys.* **1989**, *91*, 3647.
52. Hansson, T.; Marelus, J.; Aqvist, J.; *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27.
53. Almlof, M.; Carlsson, J.; Aqvist, J.; *J. Chem. Theory Comput.* **2007**, *3*, 2162.
54. *Minitab 14*; Minitab Inc., State College, PA, USA, 2014.
55. de Campos, L. J.; de Melo, E. B.; *J. Mol. Graphics Modell.* **2014**, *54*, 19.

56. Birck, M. G.; Campos, L. J.; de Melo, E. B.; *Quim. Nova* **2016**, *39*, 567.
57. Kiralj, R.; Ferreira, M. M. C.; *J. Braz. Chem. Soc.* **2009**, *20*, 770.
58. Gaudio, A. C.; Zandonade, E.; *Quim. Nova* **2001**, *24*, 658.
59. Ribeiro, T. P. P.; Manarin, F. G.; de Melo, E. B.; *Ecotoxicol. Environ. Saf.* **2018**, *153*, 175.
60. Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C. P.; Agrawal, R. K.; *Int. J. Drug Des. Discovery* **2001**, *2*, 511.
61. Kiralj, R.; Ferreira, M. M. C.; *J. Chemom.* **2010**, *24*, 681.
62. Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P.; *Environ. Health Perspect.* **2003**, *111*, 1361.
63. Melagraki, G.; Afantitis, A.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Igglessi-Markopoulou, O.; *J. Comput.-Aided Mol. Des.* **2007**, *21*, 251.
64. Golbraikh, A.; Tropsha, A.; *J. Mol. Graphics Modell.* **2002**, *20*, 269.
65. Xu, Q.; Xu, J.; *Monatsh. Chem.* **2017**, *148*, 645.
66. Todeschini, R. C. V.; *Molecular Descriptors for Cheminformatics*, vol. 41; Wiley-VCH: Weinheim, 2010.
67. Polishchuk, P.; *J. Chem. Inf. Model.* **2017**, *57*, 2618.
68. Stierand, K.; Rarey, M.; *ACS Med. Chem. Lett.* **2010**, *1*, 540.
69. Wiggers, H. J.; Rocha, J. R.; Fernandes, W. B.; Sesti-Costa, R.; Carneiro, Z. A.; Cheleski, J.; da Silva, A. B. F.; Juliano, L.; Cezari, M. H. S.; Silva, J. S.; McKerrow, J. H.; Montanari, C. A.; *PLoS Neglected Trop. Dis.* **2013**, *7*, e2370.
70. Durrant, J. D.; Keränen, H.; Wilson, B. A.; McCammon, J. A.; *PLoS Neglected Trop. Dis.* **2010**, *4*, e676.
71. Gillmor, S. A.; Craik, C. S.; Fletterick, R. J.; *Protein Sci.* **1997**, *6*, 1603.
72. Giroud, M.; Ivkovic, J.; Martignoni, M.; Fleuti, M.; Trapp, N.; Haap, W.; Kuglstatter, A.; Benz, J.; Kuhn, B.; Schirmeister, T.; Diederich, F.; *ChemMedChem* **2017**, *12*, 257.
73. Alvarez, L. H.; Gomes, D. E. B.; Gonzalez, J. E. H.; Pascutti, P. G.; *PLoS One* **2019**, *14*, e0211227.
74. Santos, L. H.; Waldner, B. J.; Fuchs, J. E.; Pereira, G. A. N.; Liedl, K. R.; Caffarena, E. R.; Ferreira, R. S.; *J. Chem. Inf. Model.* **2019**, *59*, 137.
75. Bryant, C.; Kerr, I. D.; Debnath, M.; Ang, K. K. H.; Ratnam, J.; Ferreira, R. S.; Jaishankar, P.; Zhao, D. M.; Arkin, M. R.; McKerrow, J. H.; Brinen, L. S.; Renslo, A. R.; *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6218.
76. Chen, F.; Sun, H. Y.; Wang, J. M.; Zhu, F.; Liu, H.; Wang, Z.; Lei, T. L.; Li, Y. Y.; Hou, T. J.; *RNA* **2018**, *24*, 1183.

Submitted: December 20, 2019

Published online: August 6, 2020

