

Quantitative structure–activity relationships for a series of inhibitors of cruzain from *Trypanosoma cruzi*: Molecular modeling, CoMFA and CoMSIA studies

Gustavo H.G. Trossini^a, Rafael V.C. Guido^b, Glaucius Oliva^b, Elizabeth I. Ferreira^a, Adriano D. Andricopulo^{b,*}

^aLaboratório de Planejamento e Síntese de Quimioterápicos Potenciais Contra Endemias Tropicais, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, Av. Professor Lineu Prestes 580, 05508-900, São Paulo, SP, Brazil

^bLaboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos, SP, Brazil

ARTICLE INFO

Article history:

Received 4 December 2008

Received in revised form 27 February 2009

Accepted 2 March 2009

Available online 14 March 2009

Keywords:

Cruzain

Drug design

Chagas' disease

CoMFA

CoMSIA

ABSTRACT

Human parasitic diseases are the foremost threat to human health and welfare around the world. Trypanosomiasis is a very serious infectious disease against which the currently available drugs are limited and not effective. Therefore, there is an urgent need for new chemotherapeutic agents. One attractive drug target is the major cysteine protease from *Trypanosoma cruzi*, cruzain. In the present work, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) studies were conducted on a series of thiosemicarbazone and semicarbazone derivatives as inhibitors of cruzain. Molecular modeling studies were performed in order to identify the preferred binding mode of the inhibitors into the enzyme active site, and to generate structural alignments for the three-dimensional quantitative structure–activity relationship (3D QSAR) investigations. Statistically significant models were obtained (CoMFA, $r^2 = 0.96$ and $q^2 = 0.78$; CoMSIA, $r^2 = 0.91$ and $q^2 = 0.73$), indicating their predictive ability for untested compounds. The models were externally validated employing a test set, and the predicted values were in good agreement with the experimental results. The final QSAR models and the information gathered from the 3D CoMFA and CoMSIA contour maps provided important insights into the chemical and structural basis involved in the molecular recognition process of this family of cruzain inhibitors, and should be useful for the design of new structurally related analogs with improved potency.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Parasitic diseases are a major global cause of illness, long-term disability, and death, with severe socio-economic consequences for millions of people worldwide. According to the World Health Organization, Chagas' disease is a serious and life threatening disease, affecting 10–14 million people [1]. This problem is aggravated by the rate at which populations are growing, especially in the developing world, where there is a lack of adequate sanitation facilities and poor hygiene practices which present breeding grounds for disease, illness and suffering [2–4]. In spite of the alarming health, economic and social consequences of these parasitic infections, the limited existing drug therapy (the nitroheterocyclic compounds nifurtimox and benznidazole) suffers from a combination of drawbacks including poor efficacy, resistance and serious side effects. Therefore, there is an urgent

need for new drugs that can overcome resistance and are safe and effective for use in human [5–8].

Cysteine proteases are widespread in nature [9]. Their implication in numerous vital processes of several parasitic protozoa makes them highly attractive targets for drug design [10–12]. Cruzain (EC 3.4.22), the major cysteine protease of the parasite *Trypanosoma cruzi*, plays a pivotal role during the infection of host cells, replication, and metabolism [13], and has been considered an important target for the development of new antitrypanosomal agents [14–17].

Structure- and ligand-based drug design approaches have become fundamental components of modern drug discovery [18–20]. Quantitative structure–activity relationship (QSAR) methods have been successfully employed to assist the design of new small molecule drug candidates, ranging from enzyme inhibitors to receptor ligands [21–25]. In the present work, molecular docking, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) have been performed to investigate a series of chemically diverse thiosemicarbazone and semicarbazone derivatives as inhibitors of cruzain

* Corresponding author. Tel.: +55 16 3373 8095; fax: +55 16 3373 9881.
E-mail address: aandrico@ifsc.usp.br (A.D. Andricopulo).

from *T. cruzi*. The predictive 3D QSAR models along with the information gathered from 3D contour maps provided important insights into the structural and chemical basis for potent cruzain inhibition within this series of (thio)semicarbazone derivatives. The identification of key intermolecular features associated potency and selectivity should be a valuable tool for the design of new inhibitors within this structural diversity with enhanced properties.

2. Materials and methods

2.1. Data set

The data set used for the QSAR studies contains 55 thiosemicarbazone and semicarbazone derivatives as inhibitors of the *T. cruzi* cysteine protease cruzain [26]. Representative examples of the chemical structures and corresponding pIC_{50} values ($-\log \text{IC}_{50}$, where IC_{50} is the concentration of compound required for 50% inhibition of cruzain activity) employed in this work are shown in Table 1. The complete version of Table 1, including the chemical structures and corresponding biological data, is provided as supplementary information, as well as the chemical structures in SMILES format for the complete data set.

2.2. Computational approach

The QSAR modeling analyses, calculations, and visualizations for CoMFA and CoMSIA were performed using the SYBYL 8.0 package (Tripos Inc., St. Louis, MO, USA) running on Red Hat Enterprise Linux workstations. The 3D structures of the cruzain inhibitors were constructed using standard geometric parameters of the molecular modeling software package SYBYL 8.0. Each single optimized conformation of each molecule in the data set was energetically minimized employing the Tripos force field [27] and the Powell conjugate gradient algorithm [28] with a convergence criterion of 0.05 kcal/mol Å and Gasteiger–Hückel charges [29]. A statistical cluster analysis was carried out with Tsar 3D version 3.3 (Accelrys, San Diego, USA) using the complete linkage clustering method (Euclidean distances) with no data standardization.

2.3. Molecular modeling and structural alignment

Molecular docking and scoring protocols as implemented in FlexX (BioSolveIT GmbH, Sankt Augustin, Germany) were used to investigate the possible binding conformations of the ligands within the cruzain binding pocket. The X-ray crystallographic data for cruzain in complex with the peptide-like inhibitor 1-(1-benzyl-3-hydroxy-2-oxo-propylcarbamoyl)-2-phenyl-ethyl-carbamic acid benzyl ester determined at 1.2 Å (PDB ID 1ME4) used in the docking simulations were retrieved from the Protein Data Bank (PDB) [30]. The ligand and water molecules were removed from the binding pocket and hydrogen atoms were added in standard geometry using the Biopolymer module implemented in SYBYL 8.0. Histidines, glutamines, and asparagines residues within the binding site were manually checked for possible flipped orientation, protonation, and tautomeric states with Pymol 0.99 (DeLano Scientific, San Carlos, USA) side-chain wizard script. The binding site was defined as all the amino acid residues encompassed within an 8.0-Å radius sphere centered on the bound ligand.

The structural alignment protocol was applied on the basis of the inhibitors docked conformation within the cruzain binding site generated by FlexX with default parameters [31]. The docking procedures were repeated 30 times for each data set compound. The implemented FlexX scoring function was employed to select the representative conformation for each data set compound. Subsequently, each of the 30 conformations of each compound was

submitted to the DrugScore^{ONLINE} website to rescore the proposed binding modes [32,33]. Finally, the individual ranks obtained from the scoring functions were added to give a rank order list, and then only the top-ranked poses were employed to produce the structural alignment for the 3D QSAR studies. The aligned data set is depicted in Fig. 2.

2.4. 3D QSAR: CoMFA and CoMSIA models

CoMFA [34,35] and CoMSIA [36,37] analyses were carried out on the basis of the interaction energies between a suitable probe atom and the aligned ligand atoms calculated within a cubic lattice of 2 Å spacing. The grid box embedded all ligands with a margin of at least 4 Å in each direction. A positively charged sp^3 carbon atom was used as the probe atom for calculating steric and electrostatic CoMFA fields applying SYBYL standard parameters (TRIPOS standard field, dielectric constant 1/ r , cutoff 30 kcal/mol). CoMSIA fields were computed for steric, electrostatic and hydrophobic properties, using a probe of charge +1, radius of 1, hydrophobicity of +1, and an attenuation factor of 0.3 for the Gaussian distance-dependent function.

Regression analysis was carried out by SAMPLS [38] using the partial least-squares (PLS) to derive a linear correlation between the CoMFA or CoMSIA fields and the biological property of the cruzain inhibitors. Leave-one-out (LOO) cross-validation analysis was applied to determine the number of components that yield optimally predictive models. Minimum-sigma (column filtering) was set at 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points where energy variation is below the threshold. Subsequently, a PLS analysis was performed without cross-validation, applying no column filtering. Leave-many-out (10 groups, LMO₁₀; and five groups, LMO₅) procedures were used as a more rigorous test to assess model stability and statistical significance. Each random cross-validation run was repeated 25 times to obtain mean values for q^2 and the corresponding standard deviation of error of prediction (SDEP) values. The region focusing method was applied to increase the resolution of the CoMFA and CoMSIA models. Progressive scrambling method was applied to determine the sensitivity of the QSAR models to chance correlations.

The derived 3D QSAR models were externally validated with a test set of compounds, which were not considered for QSAR model generation. After generation of the PLS training set models, the dependent variables (pIC_{50}) were predicted for the test set compounds, allowing predictive r^2 values to be determined for individual 3D QSAR models [39,40]. The r^2_{pred} is a procedure analogous to the cross-validated q^2 , and is calculated by using the formula:

$$r^2_{\text{pred}} = \frac{\text{SD} - \text{PRESS}}{\text{SD}} \quad (1)$$

where SD is the sum of squared deviation between the biological activities of the test set molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the observed and the predicted activities of the test molecules. Finally, CoMFA and CoMSIA contour maps were generated by interpolation of the pairwise products between the PLS coefficients and the standard deviations of the corresponding descriptors values.

3. Results and discussion

3.1. Chemical and biological data

The quality of the biological data under investigation as well as the structural diversity of the data set is important foundations for

Table 1Representative examples of the chemical structures and biological property (pIC₅₀) of the series of (thio)semicarbazone cruzain inhibitors employed in the 3D QSAR studies.

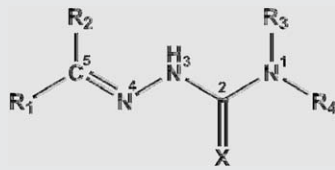
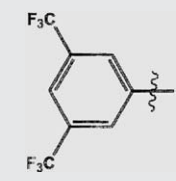
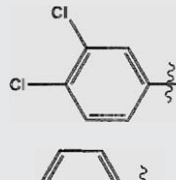
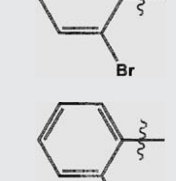
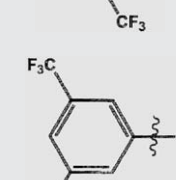
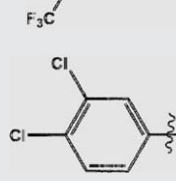
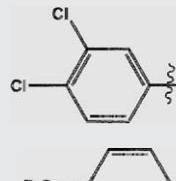
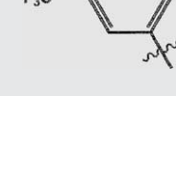

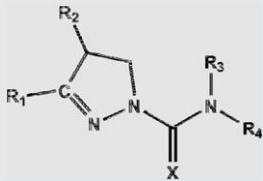
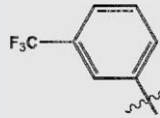
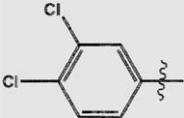
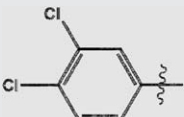
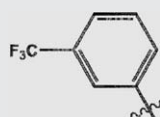
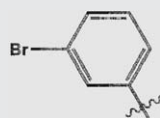
Cpd	X	R ₁	R ₂	R ₃	R ₄	pIC ₅₀
Training set:						
						
10	S		CH ₂ CH ₃	H	H	7.70
11	S		CH ₂ CH ₃	H	H	7.70
13	S		CH ₃	H	H	5.41
15	S		H	H	H	5.00
20	S		CH ₃	H	H	7.22
21	S		H	H	H	6.34
22	S		CH ₃	H	H	7.30
49	S		CH ₂ CH ₃	H	H	7.30

Table 1 (Continued)

Cpd	X	R ₁	R ₂	R ₃	R ₄	pIC ₅₀
Training set:						
						
36	S		CH ₃	H	H	7.10
39	S		H	H	H	7.22
40	S		CH ₃	H	H	7.15
41	S		H	H	H	7.10
55	S		CH ₃	H	H	7.10

successful QSAR studies [39–41]. In this work, a series of 55 thiosemicarbazone and semicarbazone derivatives as selective inhibitors of cruzain was collected from one publication reported by one laboratory [26]. The *in vitro* IC₅₀ values employed in this work were measured under the same experimental conditions, a fundamental requirement for QSAR studies [21–25]. The IC₅₀ values were converted into the corresponding pIC₅₀ values (Table 1, supplementary information) and used as dependent variables in the 3D QSAR investigations. The values of pIC₅₀ span approximately three orders of magnitude and are acceptably distributed across the pIC₅₀ range values (Fig. 1). The series of chemically attractive (thio)semicarbazone derivatives examined

has substantial structural diversity. The bulk of the structural diversity lies in the nature of the substituents linked to the N1, N4 and C5-position of the (thio)semicarbazone moieties (numerated according Table 1).

The generation of consistent statistical models depends on the proper selection of both training and test sets in terms of structural diversity and property values distribution. From the original data set of 55 inhibitors, 45 compounds (1–45, Table 1) were selected as members of the training set for model construction, and the other 10 compounds (46–55, Table 1) as members of the test set for external model validation, in the ratio of about 4:1 (approximately 20%). The compounds of the training and test sets were carefully selected in order to ensure appropriate property coverage on the entire range of pIC₅₀ values. A statistical cluster analysis confirmed that the composition of both training and test sets is representative of the whole data set (Table 1, Fig. 1). Thus, the data set is appropriate for the purpose of QSAR model development. CoMFA and CoMSIA molecular fields were used as independent variables, whereas the pIC₅₀ values were employed as dependent variable in the PLS regression analyses to derive QSAR models. The same training set was employed for all 3D QSAR analysis, and the internal and external predictive ability of the models was assessed by their *q*² and predictive *r*² values (external test set predictions), respectively.

3.2. Molecular alignment

The determination of the 3D spatial arrangement and inter-molecular interaction patterns required for ligand binding and

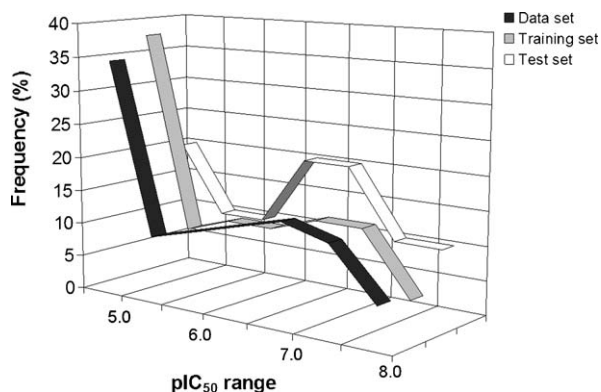


Fig. 1. Distribution of pIC₅₀ values for the training set, test set and complete data set of *T. cruzi* cruzain inhibitors.

selectivity can critically affect the outcome of 3D QSAR studies, since the analyses are highly dependent on the quality of the alignments. The knowledge about the possible structural alignment of the set of inhibitors is a valuable tool to advance our understanding of the molecular basis underlying the mechanism of enzyme inhibition [18,39]. Based on this assumption, the high-resolution crystal structure (PDB ID 1ME4, 1.2 Å) of cruzain in complex with the hydroxymethyl ketone inhibitor T10 was retrieved from the PDB and used in the docking simulations. High-resolution data (<1.5 Å) usually indicate that the model is probably more than 95% a consequence of the observed data. Consequently, errors and uncertainties in the assignment of the atomic positions are less likely to occur, which favorably contribute to the accuracy of the molecular modeling studies [42].

Structure-based molecular docking simulations were performed in this study in order to investigate the most relevant conformations of the data set (thio)semicarbazone derivatives within the cruzain binding site. The molecular alignment was accomplished using the top-ranked poses obtained by the combination of FlexX and DrugScore scoring functions (Fig. 2). Considering that the molecular alignment generated represents the ligand-bound conformations within the enzyme binding pocket, a detailed analysis of the proposed binding modes was carried out to examine the docking results. According to the FlexX results, the (thio)carbonyl moiety of the (thio)semicarbazone

Table 2

CoMFA and CoMSIA statistical results.

Model	q^2	r^2	N	SEE	F	Fraction (%)		
						S	E	H
CoMFA	0.78	0.95	6	0.186	170.302	54	46	–
CoMSIA	0.73	0.91	6	0.292	65.111	26	25	48

q^2 , Leave-one-out (LOO) cross-validated correlation coefficient; N , optimum number of components; r^2 , non-cross-validated correlation coefficient; SEE, standard error of estimate; F , F -test; S , steric field; E , electrostatic field; H , hydrophobic field.

derivatives were oriented in the direction of the cysteine and histidine catalytic residues (Cys25 and His159, respectively) favorably positioned for a nucleophilic attack (Fig. 2A). These findings are in good agreement with the proposed mechanism of inhibition, which suggests that a reversible covalent binding of the S^- anion of the active site residue Cys25 occurs with the C=S or C=O bond of the small molecule inhibitors [26]. As can be seen in Fig. 2B, the docked poses are reasonably well predicted and superposed in the *T. cruzi* cruzain catalytic site, confirming the robustness of the docking strategy and the quality of the generated structural alignment. Subsequently, the aligned molecules were quantitatively evaluated using the CoMFA and CoMSIA methods.

3.3. 3D QSAR analysis

3.3.1. Statistical results

The CoMFA and CoMSIA methods are based on the assumption that changes in binding affinities of the ligands are related to changes in molecular properties represented by molecular fields [34–37]. The alignment rule is a crucial component in 3D QSAR studies, affecting the calculations of the fields and the results of the statistical analyses. Therefore, in order to investigate the structural and chemical features related to the biological activity of this series of cruzain inhibitors, the 3D alignment developed in this work (Fig. 2) was analyzed using the CoMFA and CoMSIA fields, and the respective QSAR models were generated using the training set of 45 compounds (Table 1). The PLS method was used to construct QSAR models relating molecular descriptors with inhibitory activities. The results of the CoMFA and CoMSIA analyses are summarized in Table 2.

As can be seen, significant correlation coefficients were obtained, with cross-validated correlation coefficients q^2_{LOO} of 0.78 and 0.73, and conventional non-cross-validated correlation coefficients r^2 of 0.95 and 0.91, for CoMFA and CoMSIA, respectively. In addition, the relative contributions of the CoMFA steric and electrostatic fields to the biological activity were 54% and 46%, respectively, while the CoMSIA steric, electrostatic, and hydrophobic descriptors accounted for 26%, 25%, and 48%, respectively (Table 2). LMO procedures were performed in both cases as a more rigorous test for the stability and statistical significance of the models. Accordingly, the data set was divided into 10 (LMO_{10}) and 5 (LMO_5) randomly selected groups, and subsequently, each group was left out during the cross-validation process. Each model was evaluated 25 times by measuring its accuracy in predicting the activity of the remaining 10% and 20% data set compounds. The mean $q^2_{LMO_5}$ and $q^2_{LMO_{10}}$ values were, respectively, 0.77 and 0.75 for CoMFA; and 0.72 and 0.70 for CoMSIA. The results confirmed the stability and reliability of the 3D QSAR models, considering that the statistical values obtained for the LMO analyses were comparable to those of the LOO analyses. Progressive scrambling of the data set was also carried out to determine the sensitivity of the QSAR models to chance correlations and to test the stability of the models. The results further confirmed the consistency of the models as defined by the critical

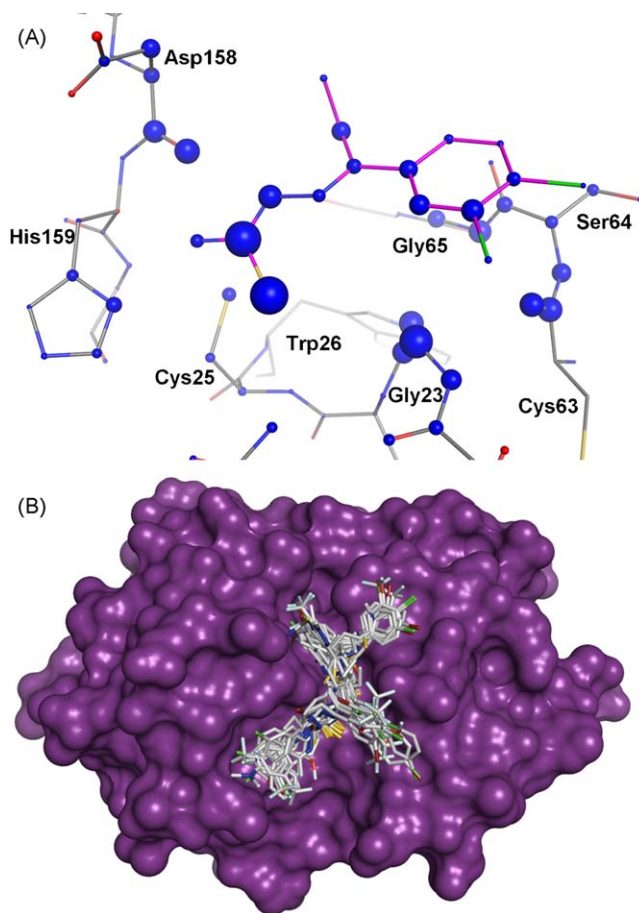


Fig. 2. (A) FlexX docking pose of compound **11** within the cruzain binding site (PDB ID 1ME4). Favorable interacting atoms are surrounded by blue spheres, whereas unfavorable interactions are shown in red spheres. The sizes of the spheres correspond to the values of the contributing per-atom according to DrugScore (i.e., the larger the blue spheres, the larger the attractive interactions between ligand and the amino acids residues). (B) Three-dimensional data set alignment for conformations generated by FlexX within the *T. cruzi* cruzain binding site.

Table 3

Experimental and predicted activities (pIC_{50}) with residual values for the test set compounds.

Compounds	Experimental	CoMFA		CoMSIA	
		Predicted	Residual	Predicted	Residual
46	6.52	6.96	−0.44	6.59	−0.07
47	7.00	6.89	0.11	6.97	0.03
48	6.25	5.39	0.86	5.51	0.74
49	7.30	6.93	0.37	6.50	0.80
50	6.77	6.92	−0.15	7.03	−0.26
51	5.42	5.73	−0.31	5.53	−0.11
52	5.00	4.80	0.20	5.61	−0.61
53	5.00	5.34	−0.34	5.38	−0.38
54	5.77	5.95	−0.18	6.29	−0.52
55	7.10	7.08	0.02	7.00	0.10

slope, and optimum statistics for cSDEP and Q^{*2} obtained at the end of different runs. In sum, the models generated in this study are very stable and possess substantial statistical significance and high internal predictive ability.

Although the q^2_{LOO} and q^2_{LMO} procedures may give a suitable representation of the predictive power of the models for untested cruzain inhibitors, the external validation process can be considered the most valuable validation method [43–47]. Therefore, the predictive ability of the CoMFA and CoMSIA models derived using the 45 training set molecules was assessed by predicting pIC_{50} values for 10 test set molecules (compounds 45–55, Table 1), which were completely excluded from model generation. Prior to prediction, the test set compounds were processed identically to the training set compounds. On the basis of the appropriate representation of chemical diversity and distribution of property values (Table 1 and Fig. 1), the test set meets the requirements for the purpose of external model validation. The detailed list of experimental and predicted pIC_{50} values are shown in Table 3 and the graphic results simultaneously displayed in Fig. 3.

The good agreement between experimental and predicted values for the test set molecules indicates the reliability of the constructed CoMFA and CoMSIA models (Table 3 and Fig. 3). The predicted pIC_{50} values fall close to the experimental values, not deviating by more than 0.86 log units. Significant predictive r^2 values of 0.81 and 0.79 were obtained for the CoMFA and CoMSIA models, respectively, which indicate a reasonable predictive power for untested compounds. No outliers were detected in this series of (thio)semicarbazone inhibitors.

3.3.2. 3D QSAR contour maps

In addition to prediction of property values for the test set molecules, the derived CoMFA and CoMSIA model should also provide important insights into molecular properties intrinsically related to receptor-binding affinity and biological activity. In this context, a detailed analysis of the interaction fields is essential to successful SAR studies. One prominent feature of the 3D QSAR methods employed in this study is the easy visualization of regions in space responsible for increases or decreases in the values of a particular type of dependent variable (e.g., pIC_{50}).

3.3.2.1. CoMFA contour maps. The molecular specialized fields can be analyzed considering two aspects in the CoMFA contour maps. Unfavorable steric regions are represented in yellow and favorable steric regions in green, while red contours represent regions where electronegative substituents may increase the biological activity, and blue contours indicate regions where electropositive groups would contribute to enhance the inhibitory potency. The CoMFA steric contour maps are shown in Fig. 4A. The favorable steric maps have good complementarity with the cruzain binding site, highlighting regions of the S1 (Gly65, Gly66 and Leu67) and S2 (Asp158)

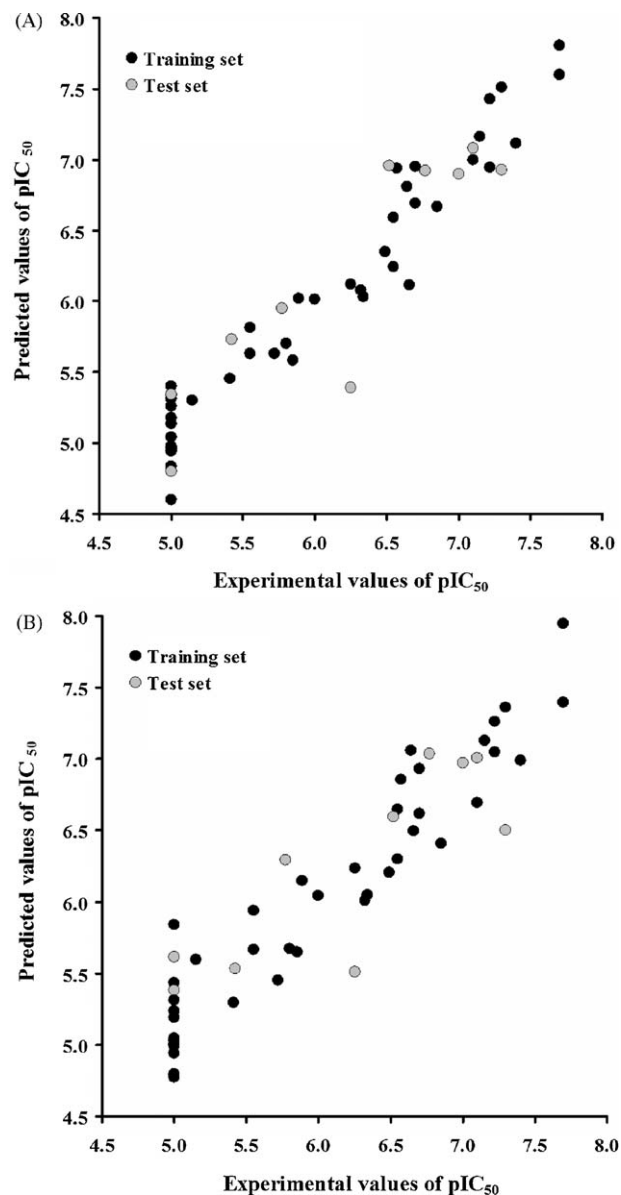


Fig. 3. Plot of predicted vs. experimental values of pIC_{50} for the training and test set cruzain inhibitors obtained with CoMFA (A) and CoMSIA (B) model.

pockets. The green contours surrounding the ethyl substituent of the most potent compound **11** (pIC_{50} = 7.70) indicate that bulkier groups in these regions would improve the inhibitory potency. These observations are in agreement with previous SAR and QSAR studies [19,26], which suggested the importance of bulky substituents on the C5-position (compounds **21**, **22**, and **11**; with pIC_{50} values of 6.34, 7.30, 7.70, respectively). The unfavorable steric map (Fig. 4A) is near to the main chain of Gly65, where some of the less potent inhibitors orient their aryl substituents.

The CoMFA electrostatic fields show a red contour located in a solvent-exposed region, as well as a blue field surrounding the NE2 atom of the side chain of Gln19. This indicates a favorable interaction with the electronegative sulfur atom of the thiosemicarbazone moiety and shows a good agreement with the chemical environment of the enzyme. Considering that the thiosemicarbazone derivatives are described as covalent reversible inhibitors of cruzain, the proposed mechanism of inhibition suggests a nucleophilic attack of the thiolate anion of the Cys25 residue on the C atom of the thiosemicarbazone moiety assisted by the transfer of the His159 proton to the thiosemicarbazone sulphur

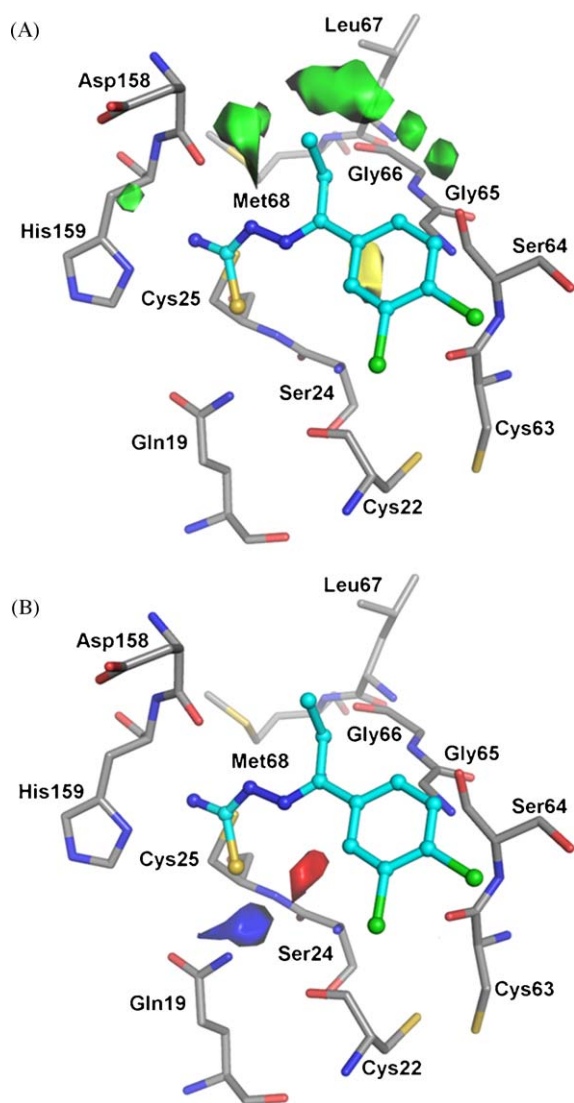


Fig. 4. CoMFA contour maps (standard deviation \times coefficient). The inhibitor **11** is represented by ball-and-stick model in the active site of cruzain. The hydrogen atoms were omitted for clarity. (A) Steric contour maps are shown in yellow and green (-0.01 and 0.009 kcal/mol, respectively). (B) Electrostatic contour maps are shown in red and blue (-0.05 and 0.01 kcal/mol, respectively).

[26]. In addition, in order to gain further structural insights into the molecular mechanism of enzyme inhibition, a detailed analysis of the CoMFA electrostatic maps was conducted with respect to the binding mode of the cruzain inhibitors. The blue field surrounding the side chain of Gln19 suggests that this residue plays an important role in the covalent reversible mechanism of inhibition. According to this model, the side chain amino group of Gln19 is close to the S atom of the thiosemicarbazone moiety to ensure a reasonable orientation of the hydrogen bond, and a possible role in the backward reaction. The H-bond interaction might orient the S atom of the tetrahedral adduct in a favorable position to facilitate the proton abstraction by the His159 residue, followed by the transfer of electrons from the adduct to the thiolate anion of the Cys25 residue (Fig. 5).

3.3.2.2. CoMSIA contour maps. The CoMSIA steric, electrostatic and hydrophobic fields are represented within cruzain binding site (Fig. 6). Fig. 6A shows the sterically favored (green) and disfavored (yellow) regions. The green contours mapped around the ethyl substituent suggest that bulkier groups are tolerated at this position. A similar sterically favorable region is observed in the CoMFA map as shown previously (Fig. 4A). This could be one of the reasons why the derivatives containing ethyl or methyl-pyrazoline substituents are among the most potent inhibitors of the series (compounds **10**, **11**, **36**, **40**, **49**, **55**). Conversely, the yellow contour near to the C6-position of the aryl group indicates that the introduction of bulky substituents at this position is unfavorable for potency (compounds **13** and **15**). The CoMSIA electrostatic contour fields are illustrated in Fig. 6B. The red contour surrounding the main chain of the Asp158 denotes the region where positively charge groups might exist in the enzyme binding site, and thus an electronegative substituent on the inhibitor scaffold could be favored. The other two red contour maps indicate that the presence of electronegative groups would increase the inhibitory potency. The blue contours in Fig. 6B suggest that the presence of electropositive substituents at N4 of the thiosemicarbazone moiety would be favored to enhance the biological potency. As can be seen, the CoMSIA electrostatic maps agree with the chemical environment of the cruzain binding site, suggesting that the main chain carbonyl group and the side chain of the Cys22 and Ser24 residues, respectively, are important sites for inter-molecular electrostatic interactions.

The hydrophobic (yellow) and hydrophilic (white) contour fields in Fig. 6C indicate regions especially correlated with the inhibitory potency of this series of (thio)semicarbazones. Taking into account the cruzain binding site environment, it is possible to

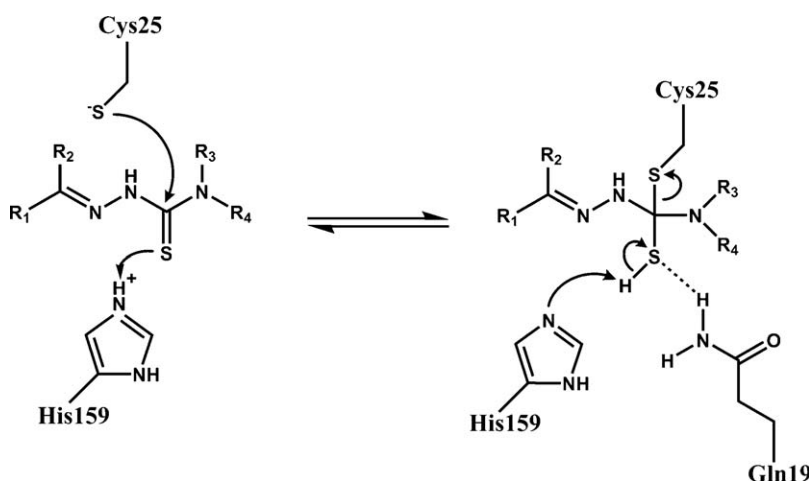


Fig. 5. Proposed mechanism of reversible covalent interaction between (thio)semicarbazones and cruzain based on molecular modeling studies.

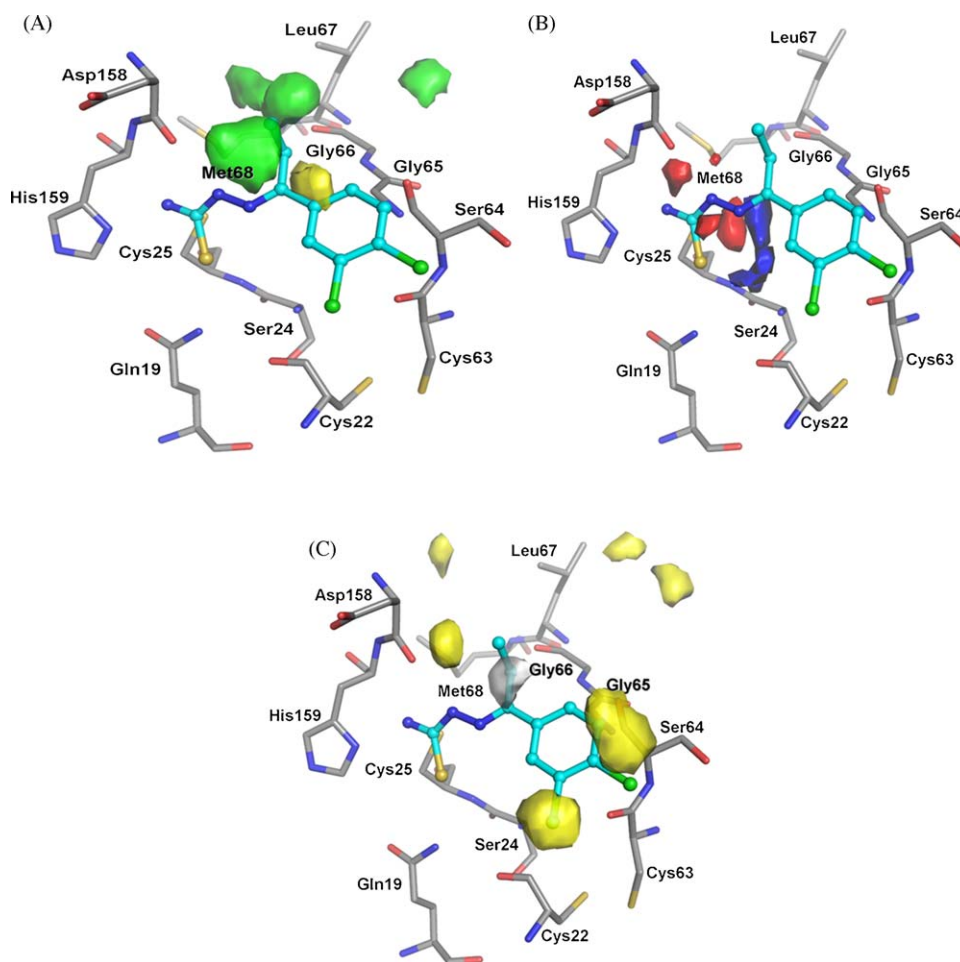


Fig. 6. CoMSIA contour maps (standard deviation \times coefficient). The inhibitor **11** is represented by ball-and-stick model in the active site of cruzain. The hydrogen atoms were omitted for clarity. (A) Steric contour maps are shown in yellow and green (-0.003 and 0.0007 kcal/mol, respectively). (B) Electrostatic contour maps are shown in red and blue (-0.01 and 0.001 kcal/mol, respectively). (C) Hydrophobic contour maps are shown in gray and yellow (-0.03 and 0.03 kcal/mol, respectively).

note that the hydrophilic region extends into the solvent exposed area. In contrast, the hydrophobic favorable regions are located into the S1 (Gly65, Gly66 and Leu67) and S2 (Asp158) pockets.

These contour maps partially overlaps with the CoMSIA steric favorable map (Fig. 6A), suggesting that ligands with bulky and hydrophobic substituents at these sites would be beneficial for enhanced potency. These findings are in agreement with previous experimental results, which indicate that bulky and hydrophobic substituents at this position are beneficial for inhibitory activity [26]. In addition, there are other two yellow contours that can be seen in Fig. 6C surrounding the 3-, 4-, and 5-position of the aryl group, indicating that hydrophobic groups at these positions are favorable for inhibitory potency (e.g., compounds **10**, **11**, **20**, **22**, **36**, and **39–41**). The identification of these important structural features should be useful for the design of new structurally related inhibitors of *T. cruzi* with improved potency.

4. Conclusions

A comprehensive study of enzyme–ligand interactions using a combination of experimental and computational approaches is essential for the design of compounds with improved affinity, selectivity and potency [18,48]. Robust 3D QSAR models were developed on the basis of the docked conformations of a class of selective inhibitors of cruzain, the major cysteine protease from *T. cruzi*, and the models possess high internal and external consistency, showing substantial predictive power. The CoMFA

and CoMSIA contour maps emphasized important regions in 3D space where modifications of steric, electrostatic and hydrophobic properties would be strongly associated with concomitant changes in the observed inhibitory potency. In addition, these observations were compatible with the 3D protein environment in the cruzain binding site and, therefore, could be used to guide further structural modifications as well as the structure-based design of new structurally related analogs with improved potency. The integration of molecular modeling and 3D QSAR studies was a useful tool for the identification of chemical and structural features responsible for inhibitory potency of this family of thiosemicarbazone and semicarbazone derivatives.

Acknowledgements

We gratefully acknowledge financial support from the State of São Paulo Research Foundation (FAPESP, Fundação de Amparo à Pesquisa do Estado de São Paulo), and the National Council for Scientific and Technological Development (CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgs.2009.03.001.

References

- [1] WHO—World Health Organization, Technical Report Series, 2002.
- [2] A.M. Strosberg, K. Barrio, V.H. Stinger, J. Tashker, J.C. Wilbur, L. Wilson, K. Woo, Chagas disease: A Latin American Nemesis, Institute for OneWorld Health—IOWH, Technical Report, 2007, 1–110.
- [3] J.W. Sanders, G.S. Fuhrer, M.D. Johnson, M.S. Riddle, The epidemiological transition: the current status of infectious diseases in the developed world versus the developing world, *Sci. Prog.* 91 (2008) 1–37.
- [4] D.H. Molyneux, Control of human parasitic diseases: context and overview, *Adv. Parasitol.* 61 (2006) 1–45.
- [5] R. Pink, A. Hudson, M.A. Mouriès, M. Bendig, Opportunities and challenges in antiparasitic drug discovery, *Nat. Rev. Drug Discov.* 9 (2005) 727–740.
- [6] S. Nwaka, R.G. Ridley, Virtual drug discovery and development for neglected diseases through public–private partnerships, *Nat. Rev. Drug Discov.* 11 (2003) 919–928.
- [7] S. Nwaka, A. Hudson, Innovative lead discovery strategies for tropical diseases, *Nat. Rev. Drug Discov.* 11 (2006) 941–955.
- [8] P.J. Hotez, D.H. Molyneux, A. Fenwick, J. Kumaresan, S.E. Sachs, J.D. Sachs, L. Savioli, Control of neglected tropical diseases, *N. Engl. J. Med.* 357 (2007) 1018–1027.
- [9] B. Turk, Targeting proteases: successes, failures and future prospects, *Nat. Rev. Drug Discov.* 5 (2006) 785–799.
- [10] P.J. Rosenthal, Cysteine proteases of malaria parasites, *Int. J. Parasitol.* 34 (2004) 1489–1499.
- [11] T.C. O'Brien, Z.B. Mackey, R.D. Fetter, Y. Choe, A.J. O'Donoghue, M. Zhou, C.S. Craik, C.R. Caffrey, J.H. McKerrow, A parasite cysteine protease is key to host protein degradation and iron acquisition, *J. Biol. Chem.* 283 (2008) 28934–28943.
- [12] Z.B. Mackey, T.C. O'Brien, D.C. Greenbaum, R.B. Blank, J.H. McKerrow, A cathepsin B-like protease is required for host protein degradation in *Trypanosoma brucei*, *J. Biol. Chem.* 279 (2004) 48426–48433.
- [13] J.J. Cazzulo, V. Stoka, V. Turk, Cruzipain, the major cysteine proteinase from the protozoan parasite *Trypanosoma cruzi*, *Biol. Chem.* 378 (1997) 1–10.
- [14] M. Sajid, J.H. McKerrow, Cysteine proteases of parasitic organisms, *Mol. Biochem. Parasitol.* 120 (2002) 1–21.
- [15] Z. Jia, S. Hasnain, T. Hiram, X. Lee, J.S. Mort, R. To, C.P. Huber, Crystal structures of recombinant rat cathepsin B and a cathepsin B-inhibitor complex. Implications for structure-based inhibitor design, *J. Biol. Chem.* 270 (1995) 5527–5533.
- [16] J.H. McKerrow, C. Caffrey, B. Kelly, P. Loke, M. Sajid, Proteases in parasitic diseases, *Annu. Rev. Pathol.* 1 (2006) 497–536.
- [17] K. Brak, P.S. Doyle, J.H. McKerrow, J.A. Ellman, Identification of a new class of nonpeptidic inhibitors of cruzain, *J. Am. Chem. Soc.* 130 (2008) 6404–6410.
- [18] R.V.C. Guido, G. Oliva, A.D. Andricopulo, Virtual screening and its integration with modern drug design technologies, *Curr. Med. Chem.* 15 (2008) 37–46.
- [19] A.D. Andricopulo, M.B. Akoachere, R. Krogh, C. Nickel, M.J. McLeish, G.L. Kenyon, L.D. Arscott, C.H. Williams, E. Davioud-Charvet, K. Becker, Specific inhibitors of *Plasmodium falciparum* thioredoxin reductase as potential antimalarial agents, *Bioorg. Med. Chem. Lett.* 16 (2006) 2283–2292.
- [20] H. Eckert, J. Bajorath, Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches, *Drug Discov. Today* 12 (2007) 225–233.
- [21] L.B. Salum, I. Polikarpov, A.D. Andricopulo, Structural and chemical basis for enhanced affinity and potency for a large series of estrogen receptor ligands: 2D and 3D QSAR studies, *J. Mol. Graph. Modell.* 26 (2007) 434–442.
- [22] R.V.C. Guido, G.H.G. Trossini, M.S. Castilho, G. Oliva, E.I. Ferreira, A.D. Andricopulo, Structure–activity relationships for a class of selective inhibitors of the major cysteine protease from *Trypanosoma cruzi*, *J. Enz. Inhib. Med. Chem.* 23 (2008) 964–973.
- [23] M.S. Castilho, R.V.C. Guido, A.D. Andricopulo, 2D quantitative structure–activity relationship studies on a series of cholesteryl ester transfer protein inhibitors, *Bioorg. Med. Chem.* 15 (2007) 6242–6252.
- [24] M.S. Castilho, R.V.C. Guido, A.D. Andricopulo, Classical and hologram QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors, *Lett. Drug Des. Discov.* 4 (2007) 106–113.
- [25] R.V.C. Guido, M.S. Castilho, S.G.R. Mota, G. Oliva, A.D. Andricopulo, Classical and hologram QSAR studies on a series of inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase, *QSAR Comb. Sci.* 27 (2008) 768–781.
- [26] X. Du, C. Guo, E. Hansell, P.S. Doyle, C.R. Caffrey, T.P. Holler, J.H. McKerrow, F.E. Cohen, Synthesis and structure–activity relationship study of potent trypanocidal thio semicarbazone inhibitors of the trypanosomal cysteine protease cruzain, *J. Med. Chem.* 45 (2002) 2695–2707.
- [27] M. Clark, R.D. Cramer, N. Van Opdenbosch, Validation of the general purpose Tripos 5.2 force field, *J. Comput. Chem.* 10 (1989) 982–1012.
- [28] M.J.D. Powell, Restart procedures for the conjugate gradient method, *Math. Program.* 12 (1977) 241–254.
- [29] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges, *Tetrahedron* 36 (1980) 3219–3228.
- [30] L. Huang, L.S. Brinen, J.A. Ellman, Crystal structures of reversible ketone-based inhibitors of the cysteine protease cruzain, *Bioorg. Med. Chem.* 11 (2003) 21–29.
- [31] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, *J. Mol. Biol.* 261 (1996) 470–489.
- [32] H. Gohlke, M. Hendlich, G. Klebe, Knowledge-based scoring-function to predict protein–ligand interactions, *J. Mol. Biol.* 295 (2000) 337–356.
- [33] DrugScore^{ONLINE}, version 0.9, 2008, available at: <http://www.agklebe.de/>.
- [34] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, *J. Am. Chem. Soc.* 110 (1988) 5959–5967.
- [35] R.D. Cramer, D.E. Patterson, J.D. Bunce, Recent advances in comparative molecular field analysis (CoMFA), *Prog. Clin. Biol. Res.* 291 (1989) 161–165.
- [36] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, *J. Med. Chem.* 37 (1994) 4130–4146.
- [37] G. Klebe, U. Abraham, Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries, *J. Comput. -Aided Mol. Des.* 13 (1999) 1–10.
- [38] B.L. Bush, R.B. Nachbar Jr., Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA, *J. Comput. -Aided Mol. Des.* 7 (1993) 587–619.
- [39] R.V.C. Guido, G. Oliva, C.A. Montanari, A.D. Andricopulo, Structural basis for selective inhibition of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase: molecular docking and 3D QSAR studies, *J. Chem. Inf. Model.* 48 (2008) 918–929.
- [40] K.M. Honório, R.C. Garratt, I. Polikarpov, A.D. Andricopulo, 3D QSAR comparative molecular field analysis on nonsteroidal farnesoid X receptor activators, *J. Mol. Graph. Modell.* 25 (2007) 921–927.
- [41] T.L. Moda, L.G. Torres, A.E. Carrara, A.D. Andricopulo, PK/DB: database for pharmacokinetic properties and predictive in silico ADME models, *Bioinformatics* 24 (2008) 2270–2271.
- [42] A.M. Davis, S.J. Teague, G.J. Kleywegt, Application and limitations of X-ray crystallographic data in structure-based ligand and drug design, *Angew. Chem. Int. Ed. Engl.* 42 (2003) 2718–2736.
- [43] L.B. Salum, I. Polikarpov, A.D. Andricopulo, Structure-based approach for the study of estrogen receptor binding affinity and subtype selectivity, *J. Chem. Inf. Model.* 48 (2008) 2243–2253.
- [44] A. Golbraikh, A. Tropsha, Beware of q²! *J. Mol. Graphics Modell.* 20 (2002) 269–276.
- [45] A. Golbraikh, M. Shen, Z. Xiao, Y.D. Xiao, K.H. Lee, A. Tropsha, Rational selection of training and test sets for the development of validated QSAR models, *J. Comput. -Aided Mol. Des.* 17 (2003) 241–253.
- [46] A.M. Doweyko, 3D-QSAR illusions, *J. Comput. -Aided Mol. Des.* 18 (2004) 587–596.
- [47] M.D. AbdulHameed, A. Hamza, J. Liu, C.G. Zhan, Combined 3D-QSAR modeling and molecular docking study on indolinone derivatives as inhibitors of 3-phosphoinositide-dependent protein kinase-1, *J. Chem. Inf. Model.* 48 (2008) 1760–1772.
- [48] A.D. Andricopulo, C.A. Montanari, Structure–activity relationships for the design of small-molecule inhibitors, *Mini Rev. Med. Chem.* 5 (2005) 585–593.