



Structure-activity relationships for a class of selective inhibitors of the major cysteine protease from *Trypanosoma cruzi*

Rafael V. C. Guido, Gustavo H. G. Trossini, Marcelo S. Castilho, Glaucius Oliva, Elizabeth I. Ferreira & Adriano D. Andricopulo

To cite this article: Rafael V. C. Guido, Gustavo H. G. Trossini, Marcelo S. Castilho, Glaucius Oliva, Elizabeth I. Ferreira & Adriano D. Andricopulo (2008) Structure-activity relationships for a class of selective inhibitors of the major cysteine protease from *Trypanosoma cruzi*, Journal of Enzyme Inhibition and Medicinal Chemistry, 23:6, 964-973, DOI: [10.1080/14756360701810322](https://doi.org/10.1080/14756360701810322)

To link to this article: <https://doi.org/10.1080/14756360701810322>



Published online: 20 Oct 2008.



Submit your article to this journal [↗](#)



Article views: 704



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

Structure-activity relationships for a class of selective inhibitors of the major cysteine protease from *Trypanosoma cruzi*

RAFAEL V. C. GUIDO¹, GUSTAVO H. G. TROSSINI², MARCELO S. CASTILHO³,
GLAUCIUS OLIVA¹, ELIZABETH I. FERREIRA², & ADRIANO D. ANDRICOPULO¹

¹Laboratório 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, São Carlos-SP 13560-970, Brazil, ²Laboratório de Planejamento e Síntese de Quimioterápicos Potenciais Contra Endemias Tropicais, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, Av. Professor Lineu Prestes 580, São Paulo-SP 05508-900, Brazil, and ³Laboratório de Bioinformática e Modelagem Molecular, Faculdade de Farmácia, Universidade Federal da Bahia, Campus Universitário de Ondina, Salvador-BA 40170-290, Brazil

(Received 28 August 2007; accepted 25 October 2007)

Abstract

Chagas' disease is a parasitic infection widely distributed throughout Latin America, with devastating consequences in terms of human morbidity and mortality. Cruzain, the major cysteine protease from *Trypanosoma cruzi*, is an attractive target for antitrypanosomal chemotherapy. In the present work, classical two-dimensional quantitative structure-activity relationships (2D QSAR) and hologram QSAR (HQSAR) studies were performed on a training set of 45 thiosemicarbazone and semicarbazone derivatives as inhibitors of *T. cruzi* cruzain. Significant statistical models (HQSAR, $q^2 = 0.75$ and $r^2 = 0.96$; classical QSAR, $q^2 = 0.72$ and $r^2 = 0.83$) were obtained, indicating their consistency for untested compounds. The models were then used to evaluate an external test set containing 10 compounds which were not included in the training set, and the predicted values were in good agreement with the experimental results (HQSAR, $r^2_{\text{pred}} = 0.95$; classical QSAR, $r^2_{\text{pred}} = 0.91$), indicating the existence of complementarity between the two ligand-based drug design techniques.

Keywords: Chagas' disease, enzyme inhibitors, drug design, chemotherapy, QSAR

Introduction

Enzymes are extremely attractive targets for small-molecule drug intervention in human diseases [1–4]. The ability to efficiently design or discover novel, patentable new chemical entities which are potent and specific inhibitors of enzymes is, therefore, of great importance [5,6]. Chagas' disease or American trypanosomiasis is a major cause of morbidity and mortality in Latin America, where about 16–18 million people are infected with *Trypanosoma cruzi* [7]. The two nitroheterocyclic drugs (nifurtimox and benznidazole) available for the treatment of this parasitic disease have severe limitations, including

poor efficacy, high toxicity, and long courses of parenteral administration. Therefore, there is an urgent need for new drugs for chemotherapy of the disease [8–10]. This has led to a new paradigm for the discovery of drug candidates that act on specific enzymes or metabolic pathways. Cysteine proteases are essential components for the survival of several parasitic protozoa. Cruzain, the major cysteine protease of the protozoan parasite *T. cruzi*, has been identified as an attractive target for the development of a new generation of antitrypanosomal agents [11,12].

Structure- and ligand-based drug design techniques are vital components of modern drug discovery. Success in this endeavor has been driven by scientific

Correspondence: A. D. Andricopulo, Laboratório 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, São Carlos-SP 13560-970, Brazil. Tel: 55 16 3373 8095. Fax: 55 16 3373 9881. E-mail: aandrico@ifsc.usp.br

and technological innovation that enabled the discovery of new therapeutic agents [13,14]. Computational techniques have provided valuable information for the design of enzyme inhibitors with a wide variety of pharmacological properties, however, designing compounds that are effective against a single target is a difficult task [15]. The understanding of the chemical basis involved in the complex process of molecular recognition and biological activity, remains a major challenge. One of the key strategies in inhibitor drug design is the development of quantitative structure-activity relationships (QSAR) from a correct interpretation of the mutual binding-mode and mechanism of action within series of structurally related inhibitors. In the present study, hologram QSAR (HQSAR) and classical QSAR methods have been employed to explore both the molecular basis underlying the selective modulation of *T. cruzi* cruzain, and to investigate the QSAR of a series of chemically versatile thiosemicarbazone and semicarbazone derivatives as selective inhibitors of this trypanosomal cysteine protease.

Materials and methods

Data sets

The data set used for the QSAR studies contains 55 thiosemicarbazone and semicarbazone derivatives as inhibitors of the trypanosomal cysteine protease cruzain, which have been selected from the literature [16]. The structures and corresponding IC_{50} values for the whole set of inhibitors are included in Table I. The IC_{50} (concentration required for 50% inhibition of cruzain) values were converted to pIC_{50} ($-\log IC_{50}$) values and used as dependent variables in the QSAR analyses. The complete data set was divided into training (compounds 1–45, Table I) and test (compounds 46–55, Table I) sets. 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.

HQSAR analysis

The HQSAR modelling analyses, calculations and visualisations were performed using the SYBYL 7.3 package (Tripos Inc., St. Louis, USA) running on Red Hat Linux 7.3 workstations, as previously described [17–21]. Basically, HQSAR models can be affected by a number of parameters concerning hologram generation: hologram length, fragment size, and fragment distinction. The generation of the molecular holograms was carried out using several combinations of the following fragment distinction: Atoms (A), Bonds (B), Connections (C), Hydrogen atoms (H), Chirality (Ch), and Donor & Acceptor (DA).

The influence of fragment size, which controls the minimum and maximum length of fragments to be included in the hologram, was further investigated by using six distinct fragment sizes over the 12 default series of hologram lengths of 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins. The patterns of fragment counts from the training set inhibitors were then related to the experimental biological data using the partial least squares (PLS) analysis.

Classical QSAR analysis

Classical QSAR studies were carried out using the DRAGON 5.4 (Taletto SRL, Milan, Italy), BuildQSAR [22] and PIROUETTE 3.11 software (Infomatrix, Washington, USA), as previously described [19,20]. Briefly, 2D molecular descriptors, including topological descriptors, connectivity indices, 2D autocorrelation descriptors, and Burden eigenvalues indices, among others, were computed using the software DRAGON 5.4 and used as independent variables in the QSAR analyses. Next, about 950 molecular descriptors were subjected to the following selection criteria. Descriptors with zero variance as well as those with poor correlation to biological property ($r^2 < 0.10$) or those more than 0.97 pairwise correlated were discarded. This strategy afforded a total of 247 descriptors. Subsequently, the BuildQSAR software was employed to search for multiple linear regressions (MLR) models of up to 4 variables using genetic algorithm (GA). The selection protocol was applied to 50 individuals (models with up to 4 variables) that were evaluated through 200,000 generations. During this processes mutation (10%) and cross-over were allowed in order to increase model fitness, which was evaluated according to the correlation coefficient criteria. All descriptors present in the 10 best MLR models were pooled together, autoscaled and used for the PLS analyses performed with the PIROUETTE software.

QSAR models validation

The QSAR models were investigated using full cross-validated r^2 (q^2) PLS. Leave-one-out (LOO) cross-validation has been applied to determine the number of components that yield optimally predictive models. Leave-many-out (LMO) validation procedures with either 10 (LMO_{10}) or 5 (LMO_5) randomly selected groups were used as more rigorous tests 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 error of prediction (SDEP). External validation was conducted employing a test set of 10 compounds, which were not considered for QSAR model generation.

Table I. Chemical structures and corresponding pIC₅₀ values for a series of inhibitors of cruzain from *Trypanosoma cruzi*.

Training Set Inhibitors					
Cpd	Structure	pIC ₅₀	Cpd	Structure	pIC ₅₀
1		6.55	10		7.30
2		5.15	11		7.70
3		6.85	12		5.00
4		6.55	13		5.41
5		5.55	14		6.25
6		5.00	15		5.00
7		5.00	16		5.89
8		5.55	17		6.49
9		6.66	18		6.32
28		7.22	34		5.00
29		7.15	35		5.00
30		7.10	36		5.80
31		5.00	37		6.00
32		5.00	38		5.85
19		6.70	40		5.00
20		7.22	41		5.00
21		6.34	42		5.00
22		7.30	43		5.00
23		5.72	44		5.00
24		6.70			
25		7.10			
26		6.57			
27		6.64			

Table I – continued

Training Set Inhibitors								
Cpd	Structure	pIC ₅₀	Cpd	Structure	pIC ₅₀	Cpd	Structure	pIC ₅₀
33		5.00	39		5.00	45		5.00
Test Set Inhibitors								
46		7.00	51		5.42			
47		6.52	52		7.40			
48		6.25	53		5.00			
49		7.70	54		5.77			
50		6.77	55		5.00			

Results and discussion

Chemical and biological data

HQSAR and classical QSAR models were derived for a data set of 55 inhibitors of *T. cruzi* cruzain (Table I). The series of chemically attractive thiosemicarbazone and semicarbazone derivatives examined has substantial structural diversity. The bulk of the structural diversity lies in the nature of the substituent linked to the N1, N4 and C5 of the thiosemicarbazone and semicarbazone moieties (Figure 1).

The *in vitro* IC₅₀ values employed in this work were measured under the same experimental conditions [16], a fundamental requirement for QSAR studies [5,17,20]. The IC₅₀ values of the data set inhibitors vary from 20 to 10,000 nM, and are acceptably distributed across the range of property values (Figure 2).

The generation of consistent statistical models is dependent on the adequacy of the training and test sets used. Therefore, from the original data set, 45 inhibitors (compounds 1–45, Table I) were selected as members of the training set for model generation, whereas the other 10 inhibitors (compounds 46–55, Table I) were held as part of the test set for external validation.

A statistical cluster analysis revealed an appropriate composition of both training and test sets in terms of chemical diversity and distribution of the biological property across the range of IC₅₀ values. Thus, the data set is suitable for QSAR model development.

HQSAR analysis

HQSAR analyses require as input only 2D chemical structures and the corresponding biological activity data. Basically, the 2D QSAR analyses carried out

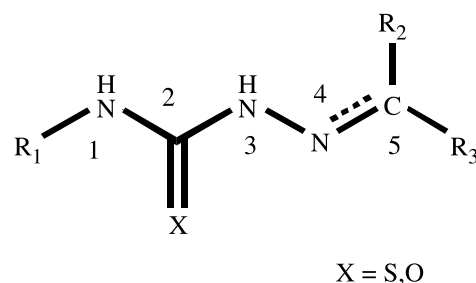


Figure 1. General scaffold of thiosemicarbazone and semicarbazone derivatives.

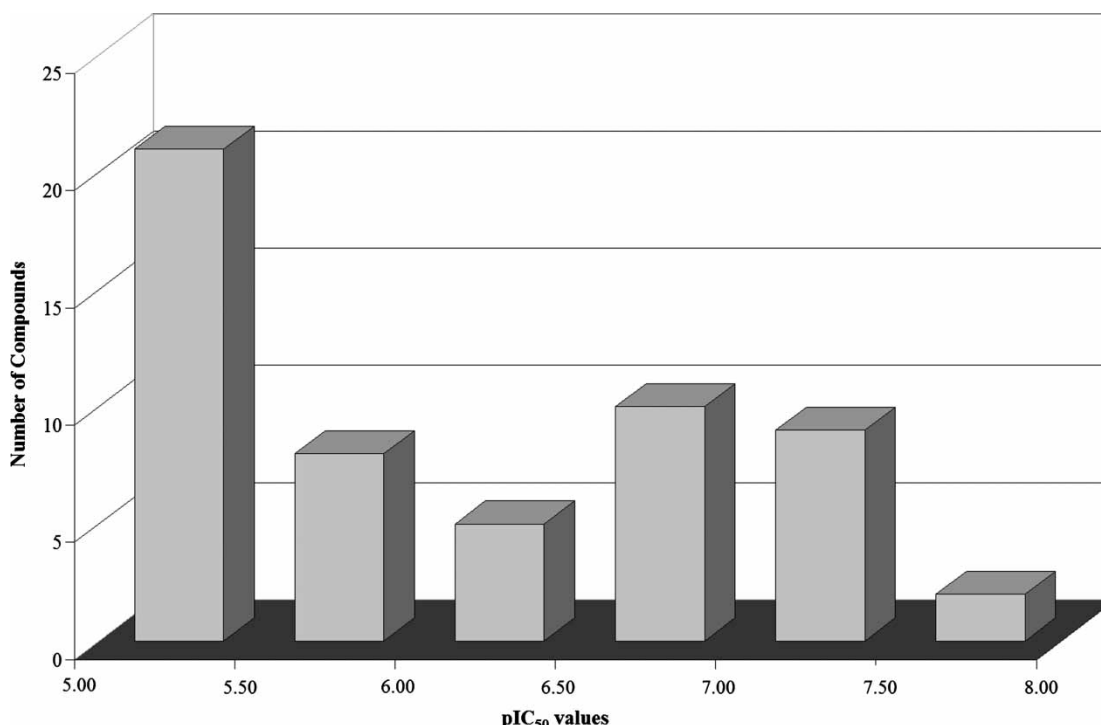


Figure 2. Distribution of pIC₅₀ values for the data set compounds.

in this work involved three main steps: (i) the generation of substructural fragments for each *T. cruzi* cruzain inhibitor member of the data set; (ii) the encoding of these fragments into a molecular hologram; (iii) the statistical generation of PLS QSAR models [20,23]. The molecular hologram (a specialised fragment fingerprint) generated consists of several substructural fragments of the thiosemicarbazone and semicarbazone derivatives that were counted into bins of a fixed length array. The bins represent all of the unique fragments included within a particular molecule and were assigned by a cyclic redundancy check (CRC) algorithm. The bin occupancies of the molecular hologram are structural descriptors (independent variables) encoding compositional and topological molecular information, that is, a fundamental component of molecular basis for the process of target recognition and biological activity. Thus, the molecular hologram includes information on the quantity and type of each fragment in each molecule of the data set, encoding all linear, branched, and overlapping molecular fragments without requiring the 3D alignment of the inhibitors.

HQSAR uses the molecular holograms and PLS to generate fragment-based structure-activity relationships. This investigation requires selecting values for parameters that specify the length of the hologram, as well as the size and type of the fragments that are to be encoded [24–26]. Therefore, several combinations of fragment distinction were considered during the QSAR modelling runs using the fragment size default

(4–7), as follows: A/B, A/B/C, A/B/C/H, A/B/C/H/Ch, A/C, A/C/Ch, A/C/DA, A/C/H/Ch, A/C/Ch/DA, A/B/H, A/B/C/Ch, A/B/DA, A/B/C/DA, A/B/H/DA, A/B/C/H/DA, and A/B/H/Ch/DA. The HQSAR analyses were performed over the 12 default series of hologram lengths ranging from 53 to 401 bins. The statistical results from the PLS analyses for the 45 training set inhibitors of *T. cruzi* cruzain using several fragment distinction combinations are summarized in Table II.

As can be seen, the best statistical results were obtained using A/C/Ch/DA as distinction information (model 9, $q^2 = 0.72$ and $r^2 = 0.94$, with SEE = 0.23). During the development of the several models, a considerable influence of the fragment distinction parameter on the predictive ability of the models was observed, as indicated by the values of q^2 in Table II (models 1–8, 10–15). This is probably due to an increase in poor information generated by the descriptors used for model development, which adds substantial noise to the cross-validation process [27]. Conversely, the inclusion of the fragment distinction Ch and DA has dramatically improved the predictive ability of the model 9 when compared to models 6 and 7, suggesting that steric and hydrogen-bonding properties play a major role towards the inhibitory potency in this series of cruzain inhibitors. These results highlight the importance of selecting the most appropriate combination of molecular descriptors for the generation of predictive models, which reflects the quantitative description of the structure-activity relationships within

Table II. HQSAR analysis for various fragment distinctions on the key statistical parameters using fragment size default (4-7).

Model	Fragment Distinction	Statistical Parameters				
		q^2	r^2	SEE	HL	PC
1	A/B	0.45	0.83	0.39	307	6
2	A/B/C	0.54	0.87	0.35	401	6
3	A/B/C/H	0.49	0.88	0.33	353	6
4	A/B/C/H/Ch	0.57	0.88	0.33	353	6
5	A/C	0.41	0.79	0.44	199	6
6	A/C/Ch	0.41	0.83	0.38	353	6
7	A/C/DA	0.24	0.49	0.65	97	3
8	A/C/H/Ch	0.38	0.73	0.49	151	4
9	A/C/Ch/DA	0.72	0.94	0.23	353	6
10	A/B/H	0.41	0.86	0.36	401	6
11	A/B/C/Ch	0.56	0.84	0.38	53	6
12	A/B/DA	0.27	0.52	0.64	83	4
13	A/B/C/DA	0.26	0.50	0.65	307	3
14	A/B/H/DA	0.52	0.80	0.42	353	5
15	A/B/C/H/DA	0.53	0.82	0.40	353	6
16	A/B/H/Ch/DA	0.67	0.92	0.27	151	6

q^2 , cross-validated correlation coefficient; r^2 , non-cross-validated correlation coefficient; SEE, non-cross-validated standard error; HL, hologram length; PC, optimal number of components. Fragment distinction: A, atoms; B, bonds; C, connections; H, hydrogen atoms; Ch, chirality; DA, donor and acceptor.

the data set. Subsequently, the influence of different fragment sizes on the key statistical parameters was further investigated for the best HQSAR model generated (Table III). Fragment size parameters control the minimum and maximum lengths of fragments to be included in the hologram fingerprint. The results for the different fragment sizes evaluated (2-5, 3-6, 4-7, 5-8, 6-9 and 7-10) show that the statistical quality of models improved with the increasing of the fragment size up to 6-9 (model 21, $q^2 = 0.75$ and $r^2 = 0.96$, with SEE = 0.20).

The internal statistical consistency of the model 21 was evaluated through the analysis of values of q^2 and SDEP obtained by more consistent cross-validation procedures using about 10% and 20% of training set compounds (LMO₁₀ and LMO₅, respectively). The cross-validation using 10 random groups yielded an average q_{10}^2 of 0.72 with SDEP₁₀ of 0.50, while an average q_5^2 of 0.70 with SDEP₅ of 0.52 was obtained for the cross-validation using 5 random groups. These results confirm the good stability and robustness of the model.

Although q^2 values may give a suitable representation of the model internal consistency and predictive power, the most valuable test of a QSAR model is its ability to predict the property value for new compounds, not included in the training set for model generation [24,28-30]. As the structure encoded in the 2D specialized fingerprints is directly related to the biological activity, HQSAR models should be useful to predict the activity of new structurally related molecules from its fingerprint [25,26]. Thus, the predictive power of the best QSAR model derived using the 45 training set inhibitors was assessed by predicting the biological property for 10 test set compounds (inhibitors 46-55, Table I). The results of the external evaluation are shown in Table IV, and the graphic results for the predicted versus experimental activities of both training set and test set are displayed in Figure 3.

The good agreement between experimental and predicted pIC₅₀ values for the test set cruzain inhibitors indicates the high quality of the HQSAR model ($r_{pred}^2 = 0.95$). The predicted values fall close to the experimental pIC₅₀ values, deviating by no more

Table III. HQSAR analysis for the influence of different fragment sizes on the statistical parameters.

Fragment Distinction	Model	Fragment Size	Statistical Parameters				
			q^2	r^2	SEE	HL	PC
A/C/Ch/DA	17	2-5	0.67	0.86	0.36	59	6
	18	3-6	0.70	0.92	0.26	307	6
	19	4-7	0.71	0.94	0.23	353	6
	20	5-8	0.73	0.95	0.21	353	6
	21	6-9	0.75	0.96	0.20	401	6
	22	7-10	0.73	0.95	0.22	401	6

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

Cpd	Experimental	HQSAR		Classical QSAR	
		Predicted	Residual ^a	Predicted	Residual ^a
46	7.00	6.75	0.25	6.59	0.41
47	6.52	6.51	0.01	6.60	-0.08
48	6.25	5.81	0.44	6.52	-0.27
49	7.70	7.97	-0.27	7.53	0.17
50	6.77	6.91	-0.14	6.75	0.02
51	5.42	5.62	-0.20	5.62	-0.20
52	7.40	7.25	0.15	6.98	0.42
53	5.00	4.85	0.15	4.93	0.07
54	5.77	5.78	-0.01	6.35	-0.58
55	5.00	5.15	-0.15	4.97	0.03

^a The difference between experimental and predicted values.

than 0.44 log units. From the low residual values, it can be seen that the HQSAR model obtained is substantially reliable and can be employed to predict the biological activity of novel cruzain inhibitors within this structural diversity.

Besides predicting the activities of untested molecules, the HQSAR model plays another important role in providing hints about the influence of different molecular subfragments to biological activity. It includes the identification of substructural patterns related to potency (IC_{50} values) in the data set of cruzain inhibitors. Therefore, the HQSAR model is graphically represented in the form of very intuitive contribution maps where the colours of the atoms reflect their relative contribution to the overall activity of the molecule. In this context, an accurate interpretation of the contribution maps revealed important molecular fragments related to biological

potency, as well as providing further insights into the underlying mechanism of action of this series of cruzain inhibitors. For the analysis of the contribution maps, colours at the red end of the spectrum (e.g., red, red-orange, and orange) reflect poor (or negative) contributions, while colours at the green end of the spectrum (e.g., yellow, green-blue, and green) reflect favorable (positive) contributions. Atoms with intermediate contributions are coloured white.

For example, the contribution of different subfragments for the most potent inhibitor of the series (compound **11**, Table I) is shown in Figure 4. As can be seen, the contribution map indicates that presence of the thiosemicarbazone moiety is favorable to the inhibitory activity. This is supported by the fact that the replacement of the thiosemicarbazone with a semicarbazone moiety (compounds **39–44**) did considerably affect the inhibitory potency, showing that the thiosemicarbazone moiety is essential for this series of enzyme inhibitors. In addition, the contribution map also highlights the importance of the sulfur atom (thiosemicarbazone) for intermolecular interactions with the catalytic cysteine of the *T. cruzi* cruzain binding pocket. Concomitantly, the model also shows the importance of the N4 and C5 alkyl substituents for activity. This is in agreement with the fact that compounds **9** and **10**, with ethyl substituent at C5, have enhanced inhibitory activity when compared to molecules without substituents on C5 (inhibitors **14–18**) [16].

Classical QSAR analysis

In order to obtain further insight into the structure-activity relationships for this series of cruzain inhibitors, we resorted to a classical QSAR approach. Recently, the integration between HQSAR and classical QSAR methods has been successfully applied [19,20], illustrating the complementary nature of these approaches as a powerful tool for medicinal chemistry studies and drug design.

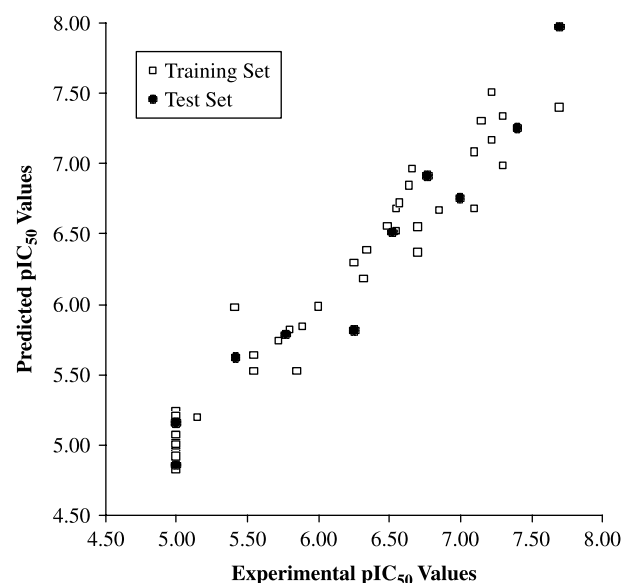


Figure 3. Plot of predicted values of pIC_{50} versus the corresponding experimental values for the training (open squares) and test (solid circles) set inhibitors for the final HQSAR model.

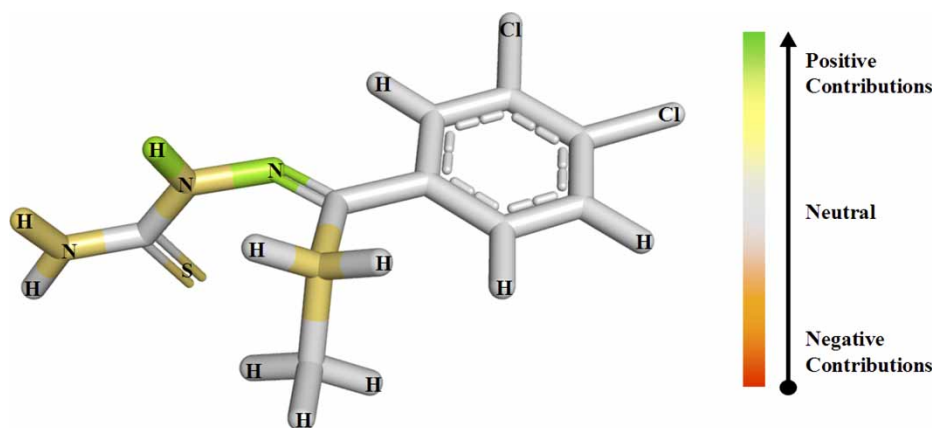


Figure 4. HQSAR contribution map for the most potent cruzain inhibitor 11.

Classical 2D QSAR studies require the calculation of molecular descriptors and the construction of mathematical models. In this work, 2D-topological, 2D-BCUTs, 2D-walk and path counts, 2D-autocorrelations, 2D-connectivity indices, 2D-information indices, 2D topological charge indices, and 2D-eigenvalue-based indices, were computed in the DRAGON 5.4 software and then used as independent variables in the QSAR modelling investigations. In order to reduce redundant and useless information, descriptors possessing zero variance, poor correlation to biological activity ($r^2 < 0.10$) as well as those with pairwise correlation superior to 0.97 were discarded. The descriptor space was further reduced through the selection of descriptors closely related to the biological activity. In order to achieve this task, BuildQSAR software was employed to search for the 10 best multiple linear regression (MLR) models of up to 4 variables using GA [31]. This strategy led to the selection of 12 descriptors (gathered from the MLR models) that were pooled together, autoscaled and used in the principal components analysis (PCA) and PLS analysis with the PIROUETTE software.

According to the PCA results, 4 principal components accounted for 91% of total variance, while additional components were considered irrelevant. A graphical analysis of the PCA results indicates that some molecular features of the data set inhibitors could be satisfactorily described by PC2. Less bulky compounds have negative values in PC2, while bulkier compounds have positive values. Weak inhibitors ($\text{pIC}_{50} = 5.0$) are broadly distributed along PC1 and PC2; however, none of them is located in the lower-right quadrant. In contrast, 75% of the most potent inhibitors ($\text{pIC}_{50} > 7.0$) are located in this region. These results suggest that the selected descriptors can be used for deriving QSAR models. Thus, PLS QSAR models were created using the LOO cross-validation procedure. As can be seen in Figure 5, the best statistical model shows good correlation within the training set ($q^2 = 0.72$ and $r^2 = 0.83$). The stability

and robustness of the model was further evaluated by the LMO cross-validation procedures. In both cases, no significant differences were obtained ($q_{10}^2 = 0.71$, with $\text{SDEP}_{10} = 0.52$; and $q_5^2 = 0.69$, with $\text{SDEP}_5 = 0.54$).

The predictive ability of the model was assessed through the prediction of the biological property for 10 test set inhibitors (compounds 46-55, Table I). The results, of similar quality compared to the HQSAR investigation, are listed in Table IV ($r_{\text{pred}}^2 = 0.91$). Besides demonstrating statistical significance, QSAR models should also provide useful chemical insights for drug design. This goal can be accomplished through a careful interpretation of the importance of the descriptors for the QSAR model. Figure 6 illustrates the influence of the descriptors over the regression vector, which can be regarded

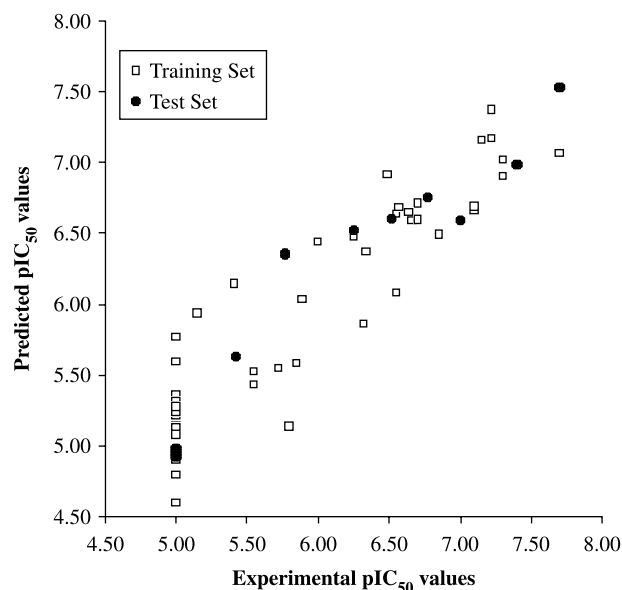


Figure 5. Plot of predicted values of pIC_{50} versus the corresponding experimental values for the training (open squares) and test (solid circles) set inhibitors for the final classical QSAR model.

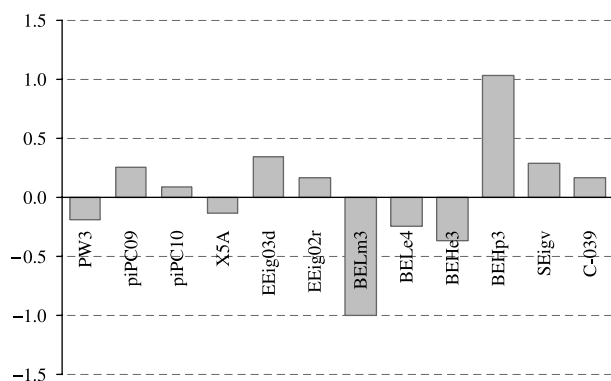


Figure 6. Descriptors score to regression vector.

as a weighted sum of the loadings included in the model. Accordingly, descriptors with small coefficients (PW3, PiPC09, PiPC10, X5A, EEig03d, EEig02r, BELe4, BELe3, SEIgv, C-039) did not significantly affect the quality of the QSAR model. Conversely, BELm3 and BEHp3 are very important topological descriptors that can be derived from the connectivity matrix, the matrix of eigenvectors, and a diagonal matrix of eigenvalues weighted by an atomic property (atomic masses and atomic polarizabilities, respectively). These descriptors are collectively known as BCUT metric and contain contributions from all atoms, thus reflecting the topology of the whole molecule. The opposite contribution of BELm3 and BEHp3 to the final model indicates that highly polarized undersized molecules have a better interaction profile to the cruzain binding site. For instance, the integrated analyses of classical QSAR and HQSAR models can be regarded as complementary in nature by suggesting that the inhibitory potency of compounds **9** and **10** might be improved in the presence of more polarizable groups of similar molecular mass in positions N4 and C5 adjacent to thiosemicarbazone moiety.

Conclusion

The use of fragment-based molecular holograms and topological descriptors, along with a standard and reproducible high quality biological data, allowed the generation of QSAR models exhibiting both good internal and external consistency, with substantial predictive power for this series of *T. cruzi* cruzain inhibitors. The information gathered from the HQSAR contribution maps, integrated with a careful analysis of the topological descriptors from the classical QSAR, shed some light on the effects of the substitution pattern related to biological activity within this series of thiosemicarbazone and semicarbazone derivatives. The consistent results confirmed that investigations can be carried out concomitantly to search for synergies between 2D QSAR technologies and drug design. In sum, as the

models reported herein have been internally and externally validated employing different 2D QSAR approaches, they appear to be valuable tools for ligand-based enzyme inhibitor design.

Acknowledgements

We gratefully acknowledge financial support from FAPESP (The State of São Paulo Research Foundation), FAPESB (The State of Bahia Research Foundation) and CNPq (The National Council for Scientific and Technological Development), Brazil.

References

- [1] Mestres J. Representativity of target families in the Protein Data Bank: Impact for family-directed structure-based drug discovery. *Drug Discov Today* 2005;10:1629–1637.
- [2] Robertson JG. Mechanistic basis of enzyme-targeted drugs. *Biochemistry* 2005;44:5561–5571.
- [3] Lespinet O, Labedan B. Orphan enzymes could be an unexplored reservoir of new drug targets. *Drug Discov Today* 2006;11:300–305.
- [4] Supuran CT, Scozzafava A. Carbonic anhydrases as targets for medicinal chemistry. *Bioorg Med Chem* 2007;15:4336–4350.
- [5] Farutin V, Masterson L, Andricopulo AD, Cheng J, Riley R, Hakimi R, Frazer J, Cordes EH. Structure-activity relationships for a class of inhibitors of purine nucleoside phosphorylase. *J Med Chem* 1999;42:2422–2431.
- [6] Andricopulo AD, Montanari CA. Structure-activity relationships for the design of small-molecule inhibitors. *Mini Rev Med Chem* 2005;5:585–593.
- [7] Beyrer C, Villar JC, Suwanvanichkij V, Singh S, Baral SD, Mills EJ. Neglected diseases, civil conflicts, and the right to health. *Lancet* 2007;370:619–627.
- [8] Barrett MP, Burchmore RJ, Stich A, Lazzari JO, Frasch AC, Cazzulo JJ, Krishna S. The trypanosomiasis. *Lancet* 2003; 362:1469–1480.
- [9] Nwaka S, Hudson A. Innovative lead discovery strategies for tropical diseases. *Nat Rev Drug Discov* 2006;5:941–955.
- [10] Nwaka S, Ridley RG. Virtual drug discovery and development for neglected diseases through public-private partnerships. *Nat Rev Drug Discov* 2003;2:919–928.
- [11] McKerrow JH, Engel JC, Caffrey CR, Related Articles, Cysteine protease inhibitors as chemotherapy for parasitic infections. *Bioorg Med Chem* 1999;7:639–644.
- [12] Doyle PS, Zhou YM, Engel JC, McKerrow JH. A Cysteine Protease Inhibitor Cures Chagas' Disease in an Immuno-deficient Murine Model of Infection. *Antimicrob Agents Chemother* 2007, doi: 10.1128/AAC.00436-07.
- [13] Eckert H, Bajorath J. Molecular similarity analysis in virtual screening: Foundations, limitations and novel approaches. *Drug Discov Today* 2007;12:225–233.
- [14] Klebe G. Virtual ligand screening: Strategies, perspectives and limitations. *Drug Discov Today* 2006;11:580–594.
- [15] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat Rev Drug Discov* 2004;3:935–949.
- [16] Du X, Guo C, Hansell E, Doyle PS, Caffrey CR, Holler TP, McKerrow JH, Cohen FE. Synthesis and structure-activity relationship study of potent trypanocidal thio semicarbazone inhibitors of the trypanosomal cysteine protease cruzain. *J Med Chem* 2002;45:2695–2707.
- [17] Honorio KM, Garratt RC, Andricopulo AD. Hologram quantitative structure-activity relationships for a series

- of farnesoid X receptor activators. *Bioorg Med Chem Lett* 2005;15:3119–3125.
- [18] Honorio KM, Garratt RC, Polikarpov I, Andricopulo AD. Hologram QSAR studies on farnesoid X receptor activators. *Lett Drug Des Discov* 2006;3:261–267.
- [19] Castilho MS, Guido RVC, Andricopulo AD. 2D Quantitative structure activity relationship studies on a series of cholesteryl ester transfer protein inhibitors. *Bioorg Med Chem* 2007;15: 6242–6252.
- [20] Castilho MS, Guido RVC, Andricopulo AD. Classical and hologram QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors. *Lett Drug Des Discov* 2007; 4:106–113.
- [21] Ferreira LG, Leita A, Montanari CA, Andricopulo AD. Hologram quantitative structure-activity relationships for a class of inhibitors of HIV-1 protease. *Lett Drug Des Discov* 2007;4:356–364.
- [22] Oliveira DB, Gaudio AC. BuildQSAR: A new computer program for QSAR analysis. *Quant Struct-Activ Relat* 2003; 19:599–601.
- [23] HQSARTM Manual. SYBYL 6.9.2, Tripos Inc., St. Louis, MO, 2003.
- [24] Honorio KM, Garratt RC, Polikarpov I, Andricopulo AD. 3D QSAR comparative molecular field analysis on Nonsteroidal farnesoid X receptor activators. *J Mol Graph Modell* 2007;25: 921–927.
- [25] Castilho MS, Postigo MP, de Paula CB, Montanari CA, Oliva G, Andricopulo AD. Two- and three-dimensional quantitative structure-activity relationships for a series of purine nucleoside phosphorylase inhibitors. *Bioorg Med Chem* 2006; 14:516–527.
- [26] Salum LD, Polikarpov I, Andricopulo AD. 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* 2007;25:434–442.
- [27] Wold S, Johansson E, Cocchi M. PLS — Partial Least Square projections to latent structures. In: Kubinyi H, editor. *3D QSAR in Drug Design*. Leiden: ESCOM; 1993. p 523–550.
- [28] Doweyko AM. 3D-QSAR illusions. *J Comput Aided Mol Des* 2004;18:587–596.
- [29] Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. *J Comput Aided Mol Des* 2003; 17:241–253.
- [30] Golbraikh A, Tropsha A. Beware of q^2 ! *J Mol Graph Model* 2002;20:269–276.
- [31] Liu H, Gramatica P. QSAR study of selective ligands for the thyroid hormone receptor beta. *Bioorg Med Chem* 2007; 15:5251–5261.