J. Braz. Chem. Soc., Vol. 22, No. 9, 1718-1726, 2011.
Printed in Brazil - ©2011 Sociedade Brasileira de Química
0103 - 5053 $6.00+0.00

Article

Descriptor- and Fragment-based QSAR Models for a Series of *Schistosoma mansoni* Purine Nucleoside Inhibitors

Humberto F. Freitas,[a] Matheus P. Postigo,[b] Adriano D. Andricopulo*,[b] and Marcelo S. Castilho*,[a]

[a]*Departamento do Medicamento, Faculdade de Farmácia, Universidade Federal da Bahia, 40170-115 Salvador-BA, Brazil*

[b]*Laboratório de Química Medicinal e Computacional, Instituto de Física de São Carlos, Universidade de São Paulo, 13566-970 São Carlos-SP, Brazil*

A enzima purina nucleosídeo fosforilase de *Schistosoma mansoni* (*Sm*PNP) é um alvo molecular atrativo para o tratamento de importantes doenças infecciosas parasitárias, com especial ênfase para o seu papel na descoberta de novos fármacos contra a esquistossomose, uma doença tropical que afeta cerca de 200 milhões de pessoas em 74 áreas endêmicas no mundo todo. No presente trabalho, a potência inibitória foi determinada e estudos das relações quantitativas entre a estrutura e atividade (QSAR), baseados em descritores e fragmentos, foram desenvolvidos para uma série de 9-deazaguaninas que atuam como inibidores da *Sm*PNP. Parâmetros estatísticos significantes (modelo baseado em descritor: $r^2 = 0,79$; $q^2 = 0,62$, $r^2_{pred} = 0,52$; e modelo baseado em fragmento: $r^2 = 0,95$; $q^2 = 0,81$; $r^2_{pred} = 0,80$) foram obtidos, indicando o potencial dos modelos para compostos ainda não testados. O modelo baseado em fragmento foi então usado para predizer a potência inibitória de um conjunto teste de compostos, e os valores preditos estão em boa concordância com os resultados experimentais.

The enzyme purine nucleoside phosphorylase from *Schistosoma mansoni* (*Sm*PNP) is an attractive molecular target for the treatment of major parasitic infectious diseases, with special emphasis on its role in the discovery of new drugs against schistosomiasis, a tropical disease that affects millions of people worldwide. In the present work, we have determined the inhibitory potency and developed descriptor- and fragment-based quantitative structure-activity relationships (QSAR) for a series of 9-deazaguanine analogs as inhibitors of *Sm*PNP. Significant statistical parameters (descriptor-based model: $r^2 = 0.79$, $q^2 = 0.62$, $r^2_{pred} = 0.52$; and fragment-based model: $r^2 = 0.95$, $q^2 = 0.81$, $r^2_{pred} = 0.80$) were obtained, indicating the potential of the models for untested compounds. The fragment-based model was then used to predict the inhibitory potency of a test set of compounds, and the predicted values are in good agreement with the experimental results.

Keywords: purine nucleoside phosphorylase, schistosomiasis, fragment-based, descriptors, QSAR

Introduction

Purine nucleoside phosphorylase (PNP, EC 2.4.2.1) plays an important role in the purine salvage pathway and has long been explored in drug design for the therapy of cancer and auto-immune diseases.[1] More recently, the PNP enzyme has also been investigated as a potential target for the treatment of parasitic infectious diseases, such as malaria and schistosomiasis.[2-4] In particular, the parasite *Schistosoma mansoni*, one of the etiologic agents of human schistosomiasis, lacks the *de novo* pathway for purine biosynthesis and depends entirely on the salvage pathway for its purine requirements for synthesis of RNA and DNA.[5-9] In this context, the use of selective PNP inhibitors from *S. mansoni* (*Sm*PNP) can cause purine starvation, leading to death of the parasite. Schistosomiasis is a major infectious disease that affects 200 million people in 74 endemic areas worldwide.[4] Praziquantel, the only effective drug for the treatment of the disease, has been in use for more than two decades and significant resistance has emerged in different geographic regions.[10-12]

*e-mail: aandrico@ifsc.usp.br, castilho@ufba.br

This scenario prompted us to investigate several 9-deazaguanine analogs, which have been described as promising *Sm*PNP inhibitors.[10] In the present study, we have collected values of IC_{50} for a series of ground-state inhibitors of *Sm*PNP and used the data to create descriptor- and fragment-based quantitative structure-activity relationship (QSAR) models which show substantial predictive promise. Our strategy took advantage of previous structure-based drug design (SBDD) studies that revealed essential requirements for *Sm*PNP binding affinity and selectivity (*e.g.*, binding to the hydrophobic pocket near Phe161, H-bonding to Tyr201).[10] The results reported herein revealed important molecular requirements for the design of new PNP inhibitors with improved potency.

Experimental

Biochemical assays and data set composition

The data set of twenty six SmPNP inhibitors (**1**-**26**, Table 1) employed in this work consists of one guanine (**18**), one 9-substituted-guanine (**21**), one 9-substituted-

Table 1. Chemical structure and biological activity of deazaguanine analogs employed in QSAR model development

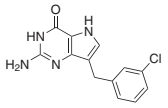
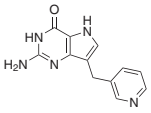
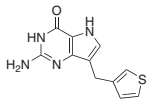
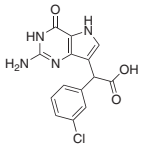
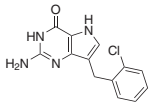
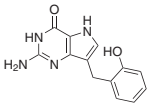
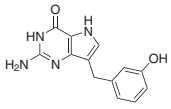
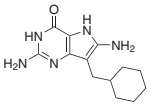
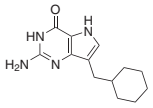
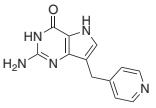
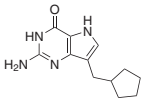
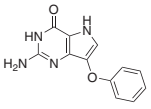
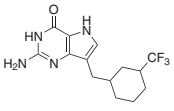
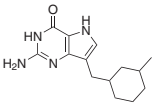
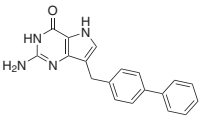
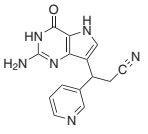
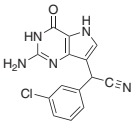
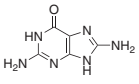
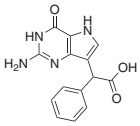
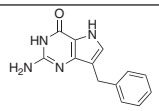
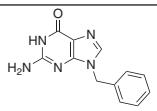
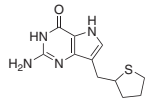
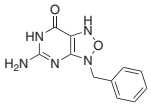
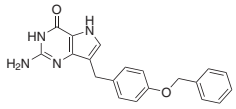
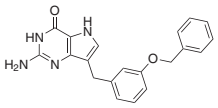
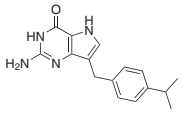
Training set					
Compound	Structure	IC_{50} / μM	Compound	Structure	IC_{50} / μM
1[b]		0.51 ± 0.03	**2**[a]		0.82 ± 0.05
3[b]		0.77 ± 0.08	**4**		200 ± 16
5[b]		0.87 ± 0.06	**6**[b]		2.08 ± 0.3
7[b]		0.79 ± 0.06	**8**[b]		6.40 ± 0.7
9		22 ± 1.8	**10**[a]		3 ± 0.4
11[b]		27.96 ± 2.1	**12**		4.50 ± 0.4
13		18.35 ± 1.5	**14**		18.28 ± 1.5
15[b]		0.10 ± 0.01	**16**		24 ± 2.2
17		39.15 ± 3.2	**18**[b]		23.32 ± 2
19		200 ± 18			

Table 1. continuation

Test set					
Compound	Structure	IC_{50} / μM	Compound	Structure	IC_{50} / μM
20[a]		1.35 ± 0.09	**21**[b]		39 ± 2.6
22[b]		12.38 ± 0.08	**23**		39.12 ± 7.3
24[b]		0.50 ± 0.08	**25**[a]		0.49 ± 0.08
26[a]		0.145 ± 0.01			

[a]Biological data available in reference 10; [b]biological data available in reference 17.

oxadiazolo-guanine (**23**) and several 9-substituted-9-deazaguanine derivatives, kindly supplied by BioCryst Pharmaceuticals Inc. Kinetic measurements were carried out spectrophotometrically with the aid of a Cary100 UV-Vis spectrophotometer, using a standard coupled assay as previously described.[10,13-16] The reaction mixture contained 5 nmol L^{-1} *Sm*PNP (as the monomer), 50 mmol L^{-1} phosphate buffer (K_3PO_4, pH 7.4), 10 μmol L^{-1} inosine, and xanthine oxidase 40 milliunits mL^{-1}. Uric acid formation was monitored at 293 nm, in triplicate at 25 °C (extinction coefficient for uric acid, ε_{293} = 12.9 L $mmol^{-1}$ cm^{-1}).[15] The percentage of inhibition was calculated according to the following equation:

$$\% \text{ of Inhibition} = 100 \times (1 - V_i / V_0)$$

where, V_i and V_0 are the initial velocities (enzyme activities) determined in the presence and in the absence of inhibitor, respectively. Compound **3**, a known *Sm*PNP inhibitor, was used as a positive control for enzyme inhibition.[10] Values of IC_{50} (concentration of compound required for 50% inhibition of *Sm*PNP) for the whole series of inhibitors were independently determined by making rate measurements for at least six inhibitor concentrations. The type of inhibition was determined for a subset of potent inhibitors as described previously. All kinetic parameters were determined from the collected data by nonlinear regression employing the SigmaPlot enzyme kinetics module. The values represent means of at least three individual experiments. Values of IC_{50} for inhibitors **1**-**3**, **5**-**9**, **11**, **12**, **14** and **19**-**26**, measured at 10 μmol L^{-1} inosine, are in good agreement with those previously described,[17] whereas comparable values are not available for the other inhibitors of the data set. The chemical structures of all *Sm*PNP inhibitors used in the modeling studies were constructed in the SYBYL 8.0 package (Tripos Inc., St. Louis, USA) and the energy was computed in a single point calculation using the AM1 semi-empirical method (keywords: 1SCF XYZ ESP NOINTER SCALE=1.4 NSURF=2 SCINCR=0.4 NOMM) as implemented in the MOPAC module. A hierarchical cluster analysis (HCA), carried out with Pirouette 4.0 software (Infometrix, Washington, USA), using the complete linkage clustering method and Euclidean distances, guided the division of the complete dataset into training (compounds **1**-**19**, Table 1) and test (compounds **20**-**26**, Table 1) sets so that both datasets present structural diversity and cover the whole dataset potency range.

Descriptor-based QSAR approach

About 2,500 2D molecular descriptors, including topological descriptors, connectivity indices, 2D autocorrelation and physicochemical descriptors and so forth, were computed using the DRAGON 5.5 software (Talette SRL, Milan, Italy) and then pre-selected as follows: descriptors with high inter-correlation (≥ 97%) or those poorly related to the biological property ($r^2 < 0.10$) were discarded. This strategy yielded 218 physicochemical descriptors that were employed to build multiple linear regression models (MLR) with up to 3 descriptors per model, as available in MOBYDIGS 1.0 software (Talette SRL, Milan, Italy). The MLR models were generated by genetic algorithm using the following fitting criteria: QUIK rule (0.005), asymptotic Q2 rule (–0.005), redundancy

RP rule (0.1) and overfitting RN rule (0.01).[18] Due to the stochastic nature of the genetic algorithm, the search was carried out using ten independent populations of 100 models each that evolved for more than 1000 generations or at least one million steps. The descriptors found in the 10 best models of each population were polled together, autoscaled and employed to develop partial least squares (PLS) models, as implemented in the PIROUETTE 4.0 software (Infometrix, Washington, USA).

Fragment-based QSAR strategy

Statistical HQSAR modeling was carried out as previously described.[19-21] Briefly, each molecule in the dataset is broken down into several unique structural fragments (linear, branched, and overlapping), which are arranged within the bins of a fixed length array (53 to 401 bins) to form a molecular hologram. The bin occupancies can be considered as structural descriptors encoding compositional and topological molecular information. Parameters that affect hologram generation such as hologram length, fragment size and fragment distinction (atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor/acceptor (DA)) were evaluated during model development, using default fragment size 4-7 over the 12 default series of hologram lengths. Next, the influence of fragment size was further investigated for the best models. All models generated in this study were investigated using the full cross-validated r^2 (q^2) partial least squares (PLS) leave-one-out (LOO) method.

QSAR model validation

External validation was carried out using a test set of seven compounds, which were not considered for the purpose of QSAR model development. The predictive ability of the models was estimated as described previously.[22]

Results and Discussion

In the present work, a series of twenty six structurally diverse compounds (Table 1, and Supplementary Information) was evaluated to determine the *in vitro* potency (IC_{50}) through kinetic studies. As expected based on previous studies,[10,17] these are competitive inhibitors of *Sm*PNP. For instance, double reciprocal plots of velocity as a function of substrate for compounds **15** and **16** show that V_{max} (intercept value of $1/v_0$) is constant at all inhibitor concentrations, whereas the apparent value of K_M (x-intercept, $-1/K_M$) changes with increasing inhibitor concentration (Figure 1). This experimental behavior is observed for all *Sm*PNP inhibitors, whose IC_{50} values range from 0.1 to 200 μM, a factor of potency of 2000.

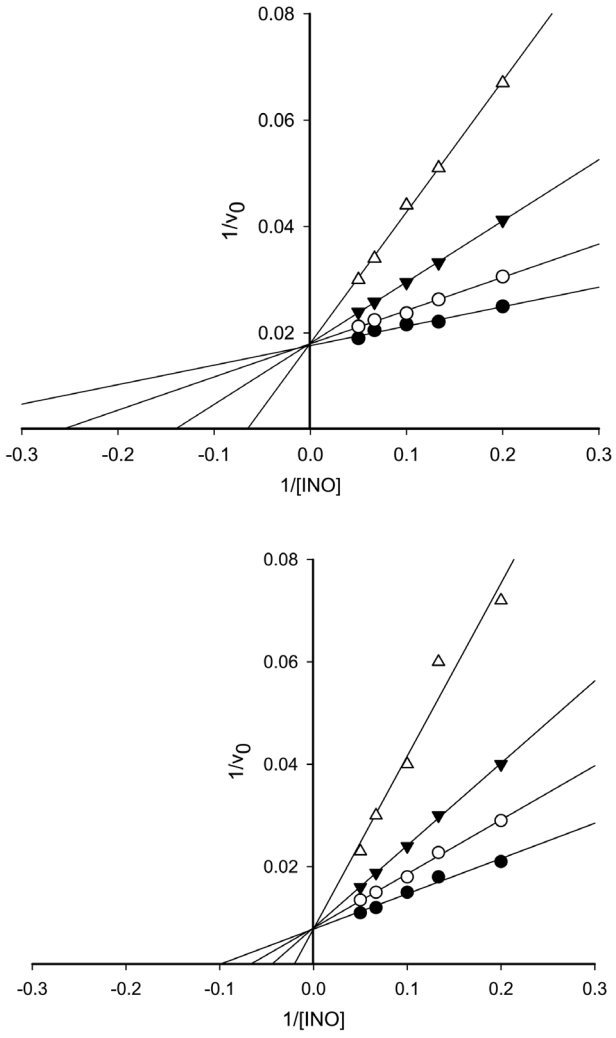


Figure 1. Competitive inhibition profile of compounds **15** and **16** against *Sm*PNP. Kinetic experiments were conducted in the presence of increasing concentrations of the inhibitors (upper graphic: **15**: 0.2 μmol L^{-1} (○), 0.4 μmol L^{-1} (▼) and 0.8 μmol L^{-1} (△); lower graphic: **16**: 2.5 μmol L^{-1} (○), 5.0 μmol L^{-1} (▼) and 10.0 μmol L^{-1} (△)). The absence of inhibitor is depicted by ●.

Although structure-activity relationships (SAR) have been widely described in the last decades for ground-state mammalian PNP inhibitors, the opposite situation is true for *Sm*PNP inhibitors. It was only more recently that the first SAR studies were provided in the literature, describing key structural requirements for *Sm*PNP affinity and selectivity.[10,17] These studies suggest that hydrophobic interactions in the active site of *Sm*PNP play an important role in the binding affinity of the inhibitors. In spite of their significance and usefulness, the SAR information, of qualitative nature, would gain strategic advantages in drug

design through the incorporation of statistical predictive modeling capabilities.[23] In this context, QSAR models are useful tools for the quantitative analysis of the internal consistency and predictive ability of different data sets of compounds, with the advantage of revealing important molecular features associated with biological activities.[24,25]

The synergy between descriptor-based and fragment-based QSAR models has been a valuable approach to boost SAR studies, due to the complementary nature of these ligand-based drug design (LBDD) strategies.[26,27] Thus, our initial efforts focused on the development of QSAR models by means of topological descriptors that account for molecular size, shape and branching through graph theoretical invariants (using the DRAGON 5.5 software). Additional information regarding molecular charge and polarizability was also considered through the weighting of the descriptors.[28] A total of 2489 descriptors were calculated, and the highly correlated and those that convey no information towards the biological activity (constant and $r^2 < 0.10$) were excluded from further consideration. This protocol afforded 218 descriptors that were employed to build a number of preliminary QSAR models by multiple linear regression (MLR), containing up to 3 descriptors. While the best MLR model obtained showed good internal statistical parameters (n = 19, $r^2 = 0.82$, $q^2 = 0.78$), the predictive ability was poor ($r^2_{pred} = 0.17$). This suggests that the chemical and structural features captured in the model do not extend beyond the chemical space of training-set compounds, limiting its usefulness in drug design. Therefore, we resorted to more powerful statistical tools, such as PLS. For this purpose, the descriptors found in the 10 best models from each population were gathered, autoscaled and used for further independent QSAR modeling.

Although our initial QSAR models showed inferior statistical parameters ($r^2 = 0.64$ and $q^2 = 0.51$, and 3 components), the iterative exclusion of the descriptors that presented a lower contribution to the regression vector led to improved models. The final QSAR model ($r^2 = 0.79$, $q^2 = 0.62$, and 2 principal components) (Table 2) showed an increased predictive ability ($r^2_{pred} = 0.52$) when compared to the MLR models (Figure 2 and Table 3), though insufficient for guiding the design of more potent *Sm*PNP inhibitors.

Thus, the analysis of the descriptors that have major contributions to the QSAR regression vector would depict misleading structure-activity relationships that hold true only for the training set compounds. In fact, the low predictive ability of descriptor-based QSAR models may suggest that compounds **22** and **24** are outliers, however, as can be seen below, a careful investigation indicates that their high residual values are a consequence of topological descriptors shortcomings, such as ineffective sampling of the deazapurine-analogs chemical space.

Table 2. Descriptors considered in the final QSAR model

Symbol	Definition / Description
MSD	mean square distance index (Balaban)
Jhetv	Balaban-type index from van der Waals weighted distance matrix
PW4	path/walk 4 - Randic shape index
X3A	Average connectivity index chi-3
MATS6e	Moran autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities
MATS8e	Moran autocorrelation - lag 8 / weighted by atomic Sanderson electronegativities
EEig02x	eigenvalue 02 from edge adj. matrix weighted by edge degrees
EEig08x	eigenvalue 08 from edge adj. matrix weighted by edge degrees
EEig01r	eigenvalue 01 from edge adj. matrix weighted by resonance integrals
BEHv1	highest eigenvalue n. 1 of Burden matrix / weighted by atomic van der Waals volumes
BEHv6	highest eigenvalue n. 6 of Burden matrix / weighted by atomic van der Waals volumes
BELv6	lowest eigenvalue n. 6 of Burden matrix / weighted by atomic van der Waals volumes
GGI10	topological charge index of order 10
SEigv	eigenvalue sum from van der Waals weighted distance matrix
SEige	eigenvalue sum from electronegativity weighted distance matrix

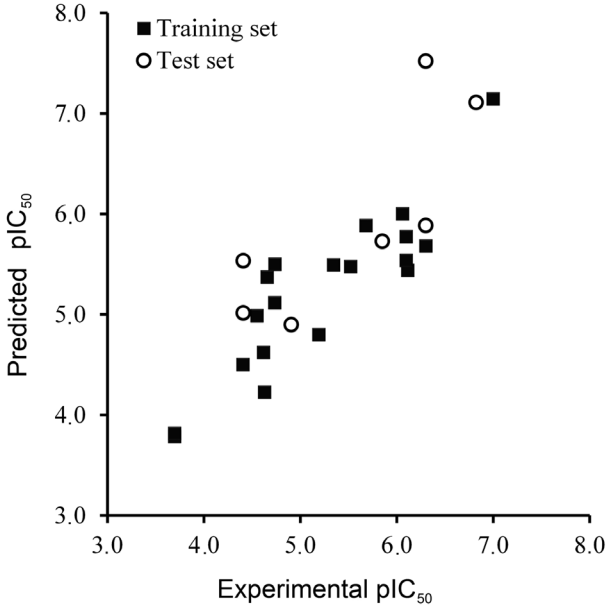


Figure 2. Plot of predicted *vs.* experimental values of pIC_{50} for the 26 *Sm*PNP inhibitors (training and test sets) according to the 2D descriptor-based QSAR model.

As part of our strategies in medicinal chemistry, we employed the fragment-based hologram QSAR (HQSAR) approach to investigate the crucial structural features

Table 3. Predicted pIC_{50} values according to the descriptor-based and fragment-based QSAR models

Compound	Experimental pIC_{50}	Descriptor-based model		Fragment-based model[a]	
		Predicted pIC_{50}	Residual	Predicted pIC50	Residual
20	5.85	5.73	0.12	6.13	–0.28
21	4.91	4.90	0.01	5.02	–0.11
22	6.30	7.52	–1.22	6.27	0.03
23	6.82	7.11	–0.29	6.34	0.48
24	4.41	5.53	–1.13	5.09	–0.68
25	4.41	5.89	0.42	5.01	–0.60
26	6.30	5.01	–0.61	6.31	–0.01

[a]Model derived using the fragment distinction A/B/H/Ch, and fragment size 4-7.

related to *Sm*PNP inhibition. HQSAR is an interesting method for this particular study, as no 3D structural information is required (*e.g.*, macromolecular target, putative binding information).[20,21] HQSAR investigations require the evaluation of parameters that specify the length of the hologram, as well as the size and type of fragment that are to be encoded. Several combinations of fragment distinction were considered during the QSAR modeling runs. The generation of molecular fragments was carried out using the following fragment distinctions: atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA). In order to assess the process of hologram generation, several combinations of these parameters were considered using the fragment size default (4-7) as follows: A/B/C, A/B/C/H, A/B/C/H/Ch, A/B/C/H/Ch/DA, A/B/H, A/B/Ch, A/B/DA, A/B/H/Ch, A/B/Ch/DA, A/B/H/DA and A/B/H/Ch/DA (Table 4). The patterns of fragment counts from the training set inhibitors were then related to the experimental biological data using PLS, as summarized in Table 4.

The influence of fragment distinction parameters has considerable effects on the quality of the models. As it can be seen in Table 4, the best statistical results among all models were obtained for models 5 (q^2 = 0.79, r^2 = 0.96, and 4 components) and 8 (q^2 = 0.81, r^2 = 0.95, and 4 components). These models were derived using A/B/H and A/B/H/Ch as fragment distinction, respectively. The use of other fragment distinctions into the molecular holograms did not improve the statistical quality of the models as shown in Table 4. It is worth noting that due to the intrinsic nature of different and highly diverse data sets, several different combinations of fragments must be considered in order to generate the best final HQSAR model.[29]

Previously, it has been shown that an extensive H-bonding network is responsible for the binding affinity of the 9-deazaguanine derivatives into the active site of *Sm*PNP.[10] This is in good agreement with our present studies, in which the presence of the fragment distinction H is highlighted in the best models 5 and 8. The influence of different fragment size in the statistical

Table 4. Influence of fragment distinction over the statistical parameters of HQSAR models, using default fragment size (4-7)

Model	Fragment distinction	q^2	r^2	SEE	HL	N
1	A/B/C	0.54	0.91	0.68	151	3
2	A/B/C/H	0.65	0.97	0.61	401	4
3	A/B/C/H/Ch	0.74	0.95	0.53	353	4
4	A/B/C/H/Ch/DA	0.67	0.97	0.60	353	4
5	A/B/H	0.79	0.96	0.47	199	4
6	A/B/Ch	0.73	0.95	0.54	353	4
7	A/B/DA	0.72	0.96	0.55	83	4
8	A/B/H/Ch	0.81	0.95	0.46	59	4
9	A/B/Ch/DA	0.73	0.95	0.54	257	4
10	A/B/H/DA	0.73	0.97	0.54	199	4
11	A/B/H/Ch/DA	0.65	0.94	0.61	199	4

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

Table 5. Influence of different fragment sizes on the statistical parameters of the two best fragment-based models

Fragment distinction	Fragment size	q^2	r^2	SEE	HL	*N*
A/B/H	2-5	0.70	0.96	0.57	199	4
	3-6	0.77	0.97	0.50	199	4
	4-7	0.79	0.96	0.47	199	4
	5-8	0.57	0.94	0.68	199	4
A/B/H/Ch	2-5	0.70	0.95	0.56	59	4
	3-6	0.49	0.95	0.73	59	4
	4-7	0.81	0.95	0.46	59	4
	5-8	0.68	0.94	0.58	59	4

parameters was further investigated for the two best HQSAR models (models 5 and 8, Table 4). Fragment size parameters control the minimum and maximum length of fragments to be included in the hologram fingerprint. Table 5 summarizes the statistical results for the distinct fragment sizes used to generate the QSAR models. As it can be seen, the variation of fragment size did not lead to the generation of better HQSAR models, and, therefore, the best statistical results were obtained with default fragment size (4-7) in both cases (A/B/H, model 5; and A/B/H/Ch, model 8).

It is important to note that the high q^2 values obtained for the best HQSAR models do not imply automatically that these models would possess high predictive ability for external compounds.[30] The most important test of a QSAR model is its ability to predict the property value for new structurally related compounds. The predictive power of the best HQSAR model derived using the training set molecules (model 8; fragment distinction A/B/H/Ch, and fragment size 4-7) was assessed by predicting pIC50 values for 7 test set molecules (compounds **20**-**26**, Table 1) that were completely excluded during the training of the model. The results are listed in Table 3, and the graphic results for the experimental *vs.* predicted activities of both training set and test set are displayed in Figure 2. The good agreement between experimental and predicted values for test set compounds indicates the reliability of the constructed HQSAR model (r^2_{pred} = 0.80). The graphic results further show the consistency between experimental and predicted pIC_{50} values of both training and test sets. The low residual values shown in Table 3 suggests that the HQSAR model obtained can be used to predict the biological activity of novel compounds within this structural class. The predicted values fall close to the experimental pIC_{50} values, deviating by less than 0.7 log units. The results show that the test set compounds are well predicted without any outliers (Figure 3). On the other hand, the quality of the results obtained for the external prediction of model 5 (r^2_{pred} = 0.71), under similar conditions, was not comparable with that of the model 8 (results not shown).

Useful fragment-based QSAR models should not only have statistical quality and predictive power, but also provide hints about which molecular fragments may be important to activity. Usually, the interpretation of the descriptors found in QSAR equations gives some clues about key electronic and steric components, which are essential for the biological property. Besides that, HQSAR has the advantage of offering an alternative and easier way to analyze the individual atomic contributions through a visual assessment of the different molecules of the data set. During the HQSAR analysis, the molecules can be colored to reflect their contribution (*e.g.*, positive, neutral or detrimental) to the biological activity of interest. The colors reflecting poor contributions are at the red end of the spectrum (red, red orange, and orange), while the

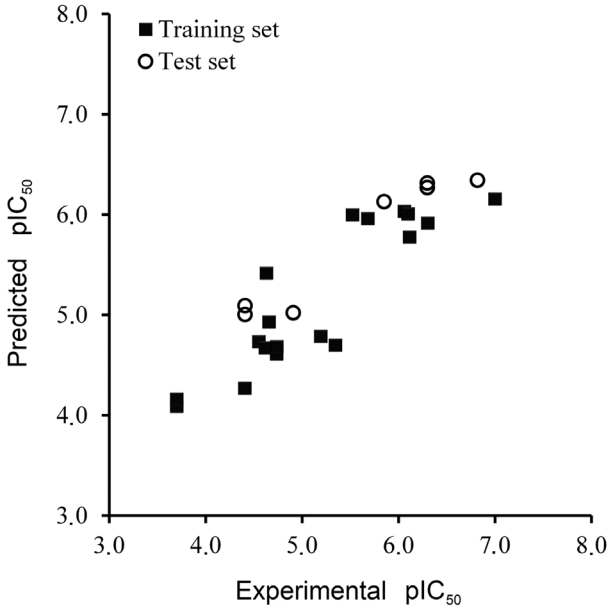


Figure 3. Plot of predicted *vs.* experimental values of pIC_{50} for the 26 *Sm*PNP inhibitors (training and test sets) for the best HQSAR model (A/B/H/Ch).

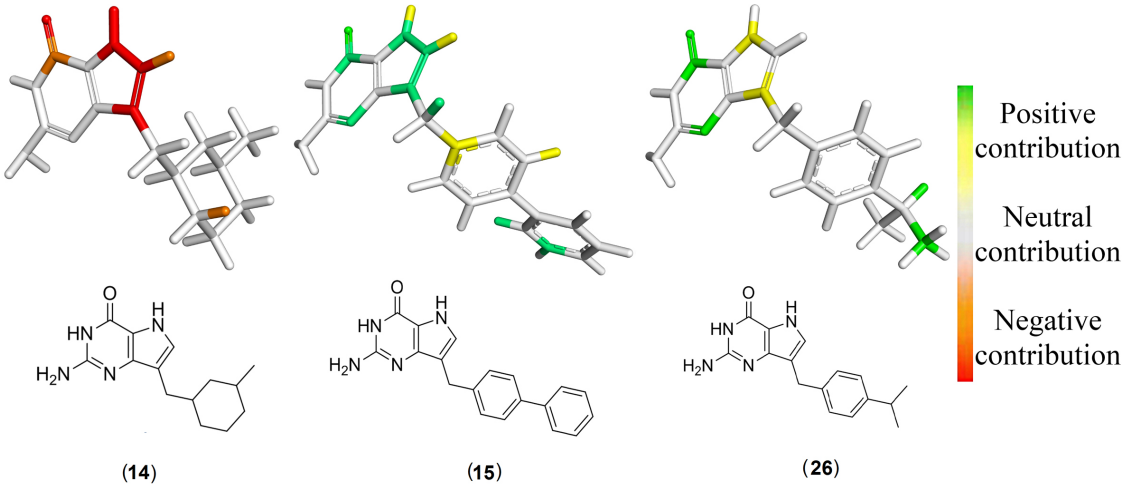


Figure 4. HQSAR contribution map: 2D and 3D chemical structure of the *Sm*PNP inhibitors **14** (IC_{50} = 18.28 μM), **15** (IC_{50} = 0.1 μM) and **26** (IC_{50} = 0.15 μM).

colors reflecting favorable contributions are at the green end (yellow, green blue, and blue). Atoms colored white reflect neutral contributions.[31] Surprisingly, comparison of the contribution maps of compounds **14**, **15** and **26** reveal that the purine ring might have opposing effects toward potency (Figure 4). This result can be explained by the H-bonding requirements in the *Sm*PNP active site. On one hand, it has been proposed that compounds possessing aryl groups in the 9 position of the purine ring (such as **15** and **26**) can reach the hydrophobic pocket in the vicinity of Phe161.[17] On the other hand, 9-substituted compounds with shorter and non-planar chains can bind loosely, being easily displaced by water molecules. Taken together, these evidences clarify the opposite role of the fragments of compound **14** in the H-bonding to Asn245 and Glu203 (reddish colored, poor H-bonding capability) in comparison with the corresponding fragments in compounds **15** and **26** (colored in green, stronger H-bonding network).

Conclusions

In spite of the urgent need for novel drugs for tropical infectious diseases, the investments in research and development (R&D) have been inadequate, as a consequence of the lack of interest shown by the major pharmaceutical and biotechnological companies. In order to circumvent this problem, most of the efforts devoted to the area of neglected diseases is observed in academia and non-governmental organizations, through public-private partnerships.[32] However, the main focus is on the early efforts to identify good targets or identify new leads for individual diseases, leaving a crucial gap in the current research and development pipeline. In this work, we have generated important descriptor- and fragment-based QSAR models for a series of 9-deazaguanines as potent inhibitors of *Sm*PNP, showing high internal and external consistency. In addition, the fragment-based model exhibited high predictive power for new compounds within this structural diversity. The molecular information gathered in this study should be useful for future efforts in the design of new inhibitors having increased affinity and selectivity.

Supplementary Information

Supplementary data are available free of charge at http://jbcs.sbq.org.br as pdf file.

Acknowledgments

We gratefully acknowledge financial support from the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB), São Paulo Research Foundation (FAPESP) and the National Council for Scientific and Technological Development (CNPq), Brazil. We are also grateful to BioCryst Pharmaceuticals, Inc. for the gift of the inhibitors employed in this work.

References

1. Castilho, M. S.; Postigo, M. P.; de Paula, C. B.; Montanari, C. A.; Oliva, G.; Andricopulo, A. D.; *Bioorg. Med. Chem.* **2006**, *14*, 516.
2. Bzowska, A.; Kulikowska, E.; Shugar, D.; *Pharmacol. Ther.* **2000**, *88*, 349.
3. Pereira, H. M.; Cleasby, A.; Pena, S. D. J.; Franco, G. R.; Garratt, R. C.; *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 1096.
4. http://www.who.int/schistosomiasis/strategy/en/, accessed in December 2010.

5. Senft, A. W.; Crabtree, G. W.; *Biochem. Pharmacol.* **1977**, *26*, 1847.
6. Shi, W.; Ting, L.; Kicska, G. A.; Lewandowicz, A.; Tyler, P. C.; Evans, G. B.; Furneaux, R. H.; Kim, K.; Almo, S. C.; Schramm, V. L.; *J. Biol. Chem.* **2004**, *279*, 18103.
7. Pereira, H. D. M.; Franco, G. R.; Cleasby, A.; Garratt, R. C.; *J. Mol. Biol.* **2005**, *353*, 584.
8. Guido, R. V. C.; Oliva, G.; Andricopulo, A. D.; *Curr. Med. Chem.* **2008**, *15*, 37.
9. Azevedo Jr., W. F.; Soares, M. B.; *Curr. Drug Targets* **2009**, *10*, 193.
10. Castilho, M. S.; Postigo, M. P.; Pereira, H. M.; Oliva, G.; Andricopulo, A. D.; *Bioorg. Med. Chem.* **2010**, *18*, 1421.
11. Webster, M.; Fallon, P. G.; Fulford, A. J.; Butterworth, A. E.; Ouma, J. H.; Kimani, G.; Dunne, D. W.; *Parasite Immunol*. **1997**, *19*, 333.
12. Sabra, A. N.; Botros, S. S.; *J. Parasitol.* **2008**, *94*, 537.
13. Postigo, M. P.; Guido, R. V. C.; Oliva, G.; Castilho, M. S.; Pitta, I. R.; Albuquerque, J. F. C.; Andricopulo, A .D.; *J. Chem. Inf. Model.* **2010**, *50*, 1693.
14. Farutin, V.; Masterson, L.; Andricopulo, A. D.; Cheng, J.; Riley, B.; Hakimi, R.; Frazer, J. W.; Cordes, E. H.; *J. Med. Chem.* **1999**, *42*, 2422.
15. Andricopulo, A. D.; Yunes, R. A.; *Chem. Pharm. Bull.* **2001**, *49*, 10.
16. Kim, B. K.; Cha, S.; Parks Jr., R. E.; *J. Biol. Chem.* **1968**, *243*, 1771.
17. Postigo, M. P.; Krogh, R.; Terni, M. F.; Pereira, H. M.; Oliva, G.; Castilho, M. S.; Andricopulo, A. D.; *J. Braz. Chem. Soc.* **2011**, *3*, 583.
18. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M.; *Anal. Chim. Acta* **2004**, *515*, 199.
19. Guido, R. V. C.; Castilho, M. S.; Mota, S. G. R.; Oliva, G.; Andricopulo, A. D.; *QSAR Comb. Sci.* **2008**, *27*, 768.
20. Borchhardt, D.; Castilho, M. S.; Andricopulo, A. D.; *Lett. Drug Des. Discovery* **2008**, *5*, 57.
21. Moda, T. L.; Montanari, C. A.; Andricopulo, A. D.; *Bioorg. Med. Chem.* **2007**, *15*, 7738.
22. Schuurmann, G.; Ebert, R.; Chen, J.; Wang, B.; Kuhen, R.; *J. Chem. Inf. Model.* **2008**, *48*, 2140.
23. Andricopulo, A. D.; Salum, L. B.; Abraham, D. J.; *Curr. Top. Med. Chem. (Sharjah, United Arab Emirates)* **2009**, *9*, 771.
24. Salum, L. B.; Polikarpov, I.; Andricopulo, A. D.; *J. Chem. Inf. Model.* **2008**, *48*, 2243.
25. Honorio, K. M.; Polikarpov, I.; Garratt, R. C.; Andricopulo, A. D.; *J. Mol. Graph. Model.* **2007**, *25*, 921.
26. Salum, L. B.; Andricopulo, A. D.; *Molec. Divers.* **2009**, *13*, 277.
27. Mota, S. G. R.; Barros, T. F.; Castilho, M. S.; *J. Braz. Chem. Soc.* **2009**, *20*, 451.
28. Caballero, J.; Garriga, M.; Fernandez, M.; *J. Comput. Aided Mol. Des.* **2005**, *19*, 771.
29. Castilho, M. S.; Guido, R. V. C.; Andricopulo, A. D.; *Lett. Drug Des. Discovery* **2007**, *4*, 106.
30. Golbraikh, A.; Tropsha, A.; *J. Mol. Graph. Model.* **2002**, *20*, 269.
31. ftp://dddc.ac.cn/manual/sybyl/pdf/hqsar.pdf accessed in December 2010
32. Nwaka, S.; Ridley, R. G.; *Nat. Rev. Drug Discovery* **2003**, *2*, 919; Nwaka, S.; Hudson, A.; *Nat. Rev. Drug Discovery* **2006**, *5*, 941.

Submitted: January 13, 2011
Published online: June 16, 2011

FAPESP has sponsored the publication of this article.

J. Braz. Chem. Soc., Vol. 22, No. 9, S1, 2011.
Printed in Brazil - ©2011 Sociedade Brasileira de Química
0103 - 5053 $6.00+0.00

Supplementary Information

Descriptor- and Fragment-based QSAR Models for a Series of *Schistosoma mansoni* Purine Nucleoside Inhibitors

Humberto F. Freitas,[a] Matheus P. Postigo,[b] Adriano D. Andricopulo*[,b] and Marcelo S. Castilho*[,a]

[a]Departamento do Medicamento, Faculdade de Farmácia, Universidade Federal da Bahia, 40170-115 Salvador-BA, Brazil

[b]Laboratório de Química Medicinal e Computacional, Instituto de Física de São Carlos, Universidade de São Paulo, 13566-970 São Carlos-SP, Brazil

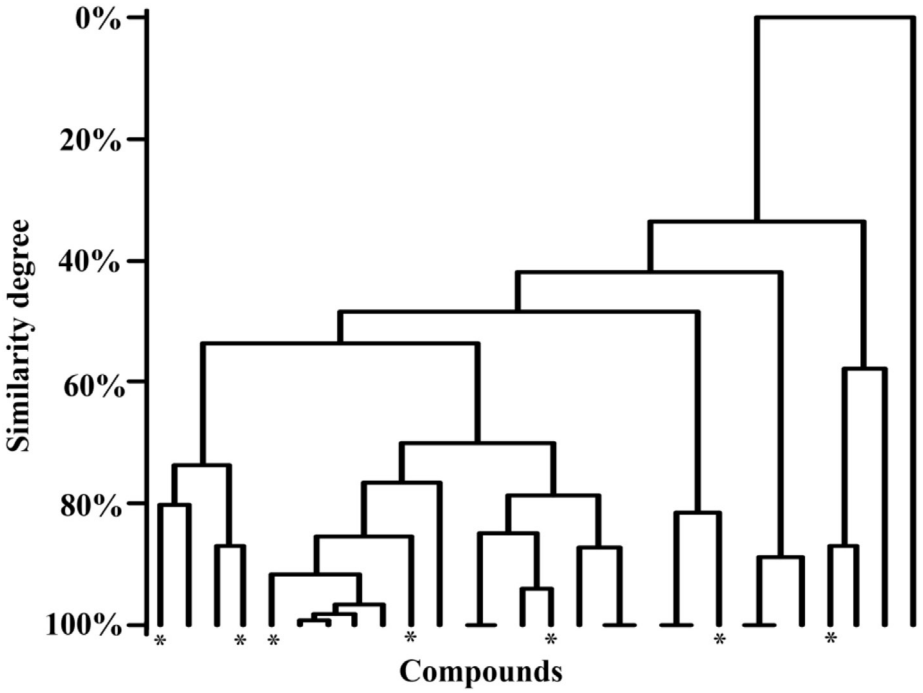


Figure S1. Hierarquical cluster analysis of all deazapurine analogs employed in QSAR model development. Test set compounds are marked with *.

*e-mail: aandrico@ifsc.usp.br, castilho@ufba.br