Design of Novel *N*-Myristoyltransferase Inhibitors of *Leishmania donovani* Using Four-Dimensional Quantitative Structure-Activity Relationship Analysis

Letícia Santos-Garcia,[a] Daniela R. Silva,[a] Letícia C. Assis,[a] Tamiris M. de Assis,[a] Giovanna C. Gajo,[a] Ítalo Antônio Fernandes,[a] Teodorico C. Ramalho[a] and Elaine F. F. da Cunha,[a]*

[a]Departamento de Química, Universidade Federal de Lavras, PO Box 3037, 37200-000 Lavras-MG, Brazil

N-Myristoylation protein is catalyzed by *N*-myristoyltransferase (NMT), an essential target in *Leishmania donovani*, the causative agent of kala-azar. Four-dimensional quantitative structure-activity relationship (4D-QSAR) analysis was applied to a series of 77 *Leishmania donovani* NMT inhibitors. Then, three new compounds were proposed using QSAR models. In addition, molecular docking was performed to predict the binding affinities and interaction modes among the proposed compounds and the NMT active site. *In silico* absorption, distribution, metabolism and excretion (ADME) evaluation was performed and potential inhibitors demonstrated satisfactory pharmacokinetic properties.

Keywords: leishmaniasis, inhibitors, 4D-QSAR, docking, ADME

Introduction

Visceral leishmaniasis (VL), also known as kala-azar, is a parasitic infection caused by *Leishmania donovani* and if untreated, it is fatal in more than 95% of cases within 2 years after disease onset. In 2016, the World Health Organization (WHO) and Gilead Sciences reported that there are around 200000 to 400000 new cases worldwide, and the disease is endemic in more than 80 countries. VL is characterized by irregular bouts of fever, weight loss, enlargement of the spleen and liver and anaemia.[1]

Chemotherapy is the main way to deal with this disease. The first-line drug, pentavalent antimony, has several adverse effects, such as vomiting, nausea, anorexia, myalgia, abdominal pain, headache, arthralgia, and lethargy and can rarely cause the severe reaction of fatal cardiac arrhythmia.[2] Other drugs, such as pentamidine, amphotericin B, paromomycin and miltefosine have several disadvantages, such as lengthy treatment, need for hospitalization and drug resistance. Only one of the commonly used drugs, miltefosine, can be administered orally, however, it has a high teratogenic potential.[3] Furthermore, the high treatment cost is also a barrier to effective disease control in poor countries.

N-Myristoyltransferase (NMT) catalyzes the covalent co-translational attachment of the fatty acid, myristate (C14:0) to the amino-terminal glycine residue after the removal of the initiating methionine.[4] *N*-Myristoylation plays an important role in protein-protein interactions of membrane proteins which, in turn, facilitate a variety of signal transduction pathways.[4] NMT has been found to be essential in *Leishmania donovani* life cycle,[5] becoming highly promising as a drug target for the treatment of VL.[6]

The design of new bioactive molecules can be accomplished by several different methods, one of which is the quantitative structure activity relationship (QSAR), which relates biological data to the chemical structure. Of the various types of QSAR, 4D-QSAR is a technique based on the 3D-QSAR approach and uses the conformational sampling obtained by molecular dynamics simulation to incorporate the fourth dimension.[7]

In the present paper, the 4D-QSAR technique was used in the study of 77 NMT inhibitors of *Leishmania donovani* developed by Leatherbarrow *et al*.[8] to propose structural changes in these compounds in order to make them more potent. Subsequently, a docking study was carried out to understand the interaction mode of the proposed compounds in the NMT active site. In addition, an *in silico* modeling of absorption, distribution, metabolism

and excretion (ADME) properties was performed in order to estimate the pharmacokinetic properties of the compounds, which are extremely important so that drug candidates can be administered orally and reach their sites of interaction to exert the expected therapeutic effect.[9,10]

Methodology

Biological data

The 77 *Leishmania donovani* inhibitors studied were synthesized and pharmacologically tested by Leatherbarrow *et al.*[8] Among these compounds, 19 represent the test set (external validation). Three test sets were built to obtain greater reliability for the model and the choice of these compounds was performed in two different ways, using the Kennard-Stone algorithm and randomly.

The first group, test set I, formed by molecules **3**, **6**, **9**, **13**, **20**, **21**, **26**, **27**, **29**, **31**, **32**, **41**, **46**, **55**, **60**, **63**, **66**, **71** and **72** was evaluated using the Kennard-Stone algorithm that select the calibration samples by Euclidean distance from the spectra. Generally, the process starts by separating two samples, the nearest to the average and the farthest, then the two are removed and the process is repeated until the desired number is reached for calibration.[11]

The remaining 2 groups, test set II, constituted by the molecules **7**, **10**, **16**, **19**, **24**, **26**, **33**, **37**, **40**, **44**, **50**, **55**, **56**, **63**, **64**, **68**, **71**, **74** and **75**, and test set III consisted of the molecules **11**, **18**, **19**, **23**, **24**, **26**, **28**, **29**, **30**, **35**, **37**, **39**, **40**, **41**, **46**, **47**, **61**, **63** and **75** were randomly evaluated. This process consists in the selection of the 77 compounds, structurally related, in three subgroups according to ranges of biological activity values and then randomly selected 20% of the compounds of each subgroup to compose the test set and the remaining ones correspond to the training set and are used for the construction of QSAR-4D models.[12]

The biological activities of these compounds were reported as the concentration capable of inhibiting 50% of the enzyme activity (IC_{50}), measured using an adapted version of the sensitive fluorescence-based assay based on coenzyme A (CoA) detection by 7-diethylamino-3-(4-maleimido-phenyl)-4-methylcoumarin.[8] In addition, all pharmacological data were obtained from the same laboratory, eliminating the potential noise that might have been introduced by the pooling of data sets from different sources. The IC_{50} (µM) values were converted into molar units and then expressed in negative logarithmic units (pIC_{50}), and are represented in Table 1. The range of pIC_{50} values for the training and test set spans at least four orders of magnitude (3.87 to 8.00), and the biological activity values show a regular distribution over the whole range.

The three-dimensional (3D) models of the 77 compounds, Table 1, were constructed using the HyperChem 7.0 software.[13] The structures were geometry-optimized in vacuum, without any restriction, using the MM+ molecular mechanics force field (HyperChem), and latterly applying the semi-empirical RM1 Hamiltonian,[14] in order to assign the partial atomic charges.

Molecular dynamic simulation (MDS)

Molecular dynamics simulation (MDS) was carried out using the MOLSIM package in the 4D-QSAR program,[15] starting from the RM1 structures, in order to build the conformational ensemble profile (CEP) of each molecule. The temperature for MDS was set at 300 K, close to the temperature assays, with a simulation sampling time of 100 ps, and intervals of 0.001 ps. Thus, a total sample of 100,000 conformations of each ligand was produced. The MDS calculations were carried out employing a distance-dependent dielectric function, $\varepsilon_r = D*r_{ij}$, which was set to $3*r_{ij}$ in order to try to model the solvent effect.

Alignment definition

In this study, we will assume that all molecules bind to the enzyme in a similar way, since the compounds are structural analogues. In general, the alignments are chosen to span the common framework of the molecules in the training and test sets. Seven alignments were performed using atoms of the benzene ring. Three-ordered atom trial alignments were selected: (*i*) A-B-C, (*ii*) A-B-D, (*iii*) C-D-E, (*iv*) C-D-F, (*v*) B-A-C, (*vi*) C-B-A and (*vii*) D-A-F, using compound **76** as a reference (Figure 1). The CEP for each compound was obtained after the MDS step was overlaid onto a cubic lattice with grid cell size of 1 Å.

Interaction pharmacophore elements

Each atom was classified into seven types of interaction pharmacophore elements (IPE): (*i*) any type (any); (*ii*) nonpolar (np); (*iii*) polar-positive charge density (p+); (*iv*) polar-negative charge density (p-); (*v*) hydrogen bond acceptor (hba); (*vi*) hydrogen bond donor (hbd) and (*vii*) aromatic systems (ar).[13] The occupancy of the grid cells by each IPE type are recorded over the conformational assembly profile, and forms the set of grid cell occupancy descriptors (GCOD) to be utilized as the pool of trial descriptors in the model building and optimization process.[7]

Table 1. Structure of the 77 *Leishmania donovani* inhibitors and their pIC$_{50}$ values.[8] Compound numbers of test set I are underlined
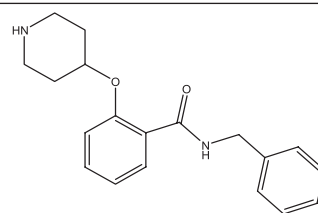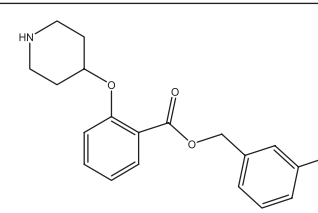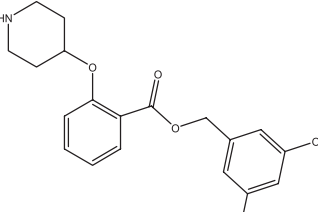
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
1		6.30	2		6.22
<u>3</u>		5.46	4		6.10
5		5.38	<u>6</u>		5.03
7		4.63	8		5.37
<u>9</u>		5.80	10		5.15
11		5.15	12		6.00
<u>13</u>		6.68	14		6.77
15		6.48	16		5.80

Table 1. Structure of the 77 *Leishmania donovani* inhibitors and their pIC_{50} values.[8] Compound numbers of test set I are underlined (cont.)

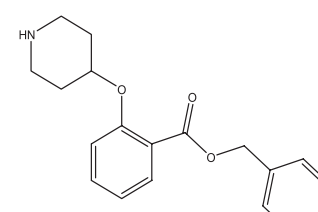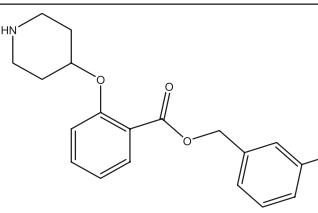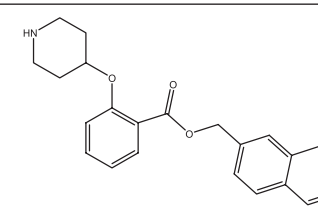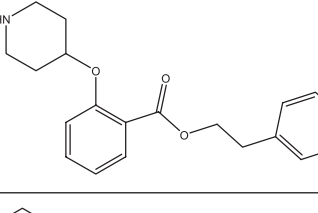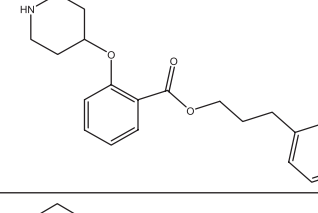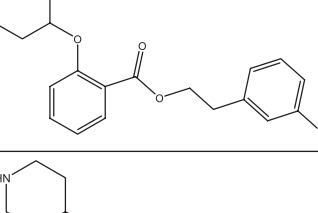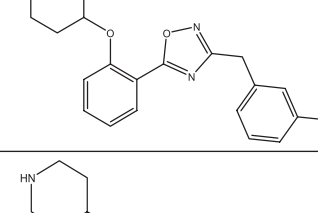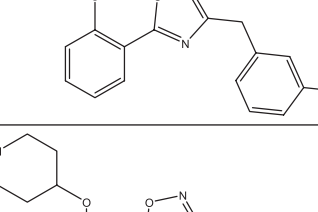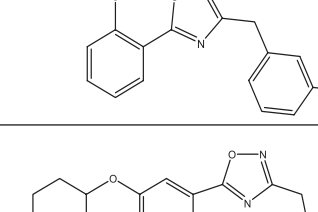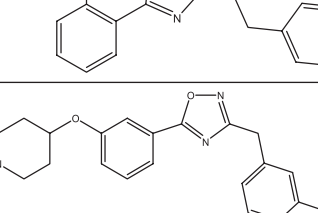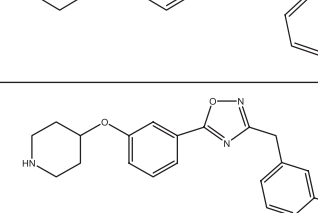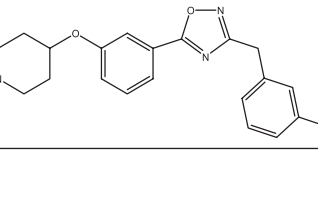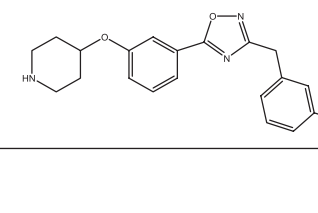| No.	Structure	pIC_{50}	No.	Structure	pIC_{50}
17		5.52	**18**		6.51
19		6.15	**20**		5.80
21		5.62	**22**		5.40
23		6.15	**24**		5.82
25		6.30	**26**		7.18
27		7.36	**28**		6.85
29		6.10	**30**		7.15
31		6.82	**32**		6.52

Table 1. Structure of the 77 *Leishmania donovani* inhibitors and their pIC$_{50}$ values.[8] Compound numbers of test set I are underlined (cont.)

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
33		6.00	**34**		5.68
35		7.34	**36**		7.51
37		7.14	**38**		6.41
39		7.08	**40**		6.62
<u>41</u>		6.55	**42**		4.47
43		3.87	**44**		5.24
45		4.52	**<u>46</u>**		4.29

Table 1. Structure of the 77 *Leishmania donovani* inhibitors and their pIC$_{50}$ values.[8] Compound numbers of test set I are underlined (cont.)

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
47		5.47	**48**		5.28
49		4.69	**50**		4.77
51		5.35	**52**		5.11
53		6.06	**54**		6.34
55		4.42	**56**		4.25
57		4.66	**58**		4.66
59		5.80	**60**		6.57

Table 1. Structure of the 77 *Leishmania donovani* inhibitors and their pIC$_{50}$ values.[8] Compound numbers of test set I are underlined (cont.)

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
61		7.70	**62**		7.07
<u>63</u>		7.30	**64**		6.41
65		7.00	**<u>66</u>**		6.77
67		5.32	**68**		6.17
69		5.68	**70**		7.00
<u>71</u>		5.51	**<u>72</u>**		6.89
73		6.19	**74**		5.35

Table 1. Structure of the 77 *Leishmania donovani* inhibitors and their pIC_{50} values.[8] Compound numbers of test set I are underlined (cont.)

No.	Structure	pIC_{50}	No.	Structure	pIC_{50}
75		7.76	**76**		8.00
77		7.15			

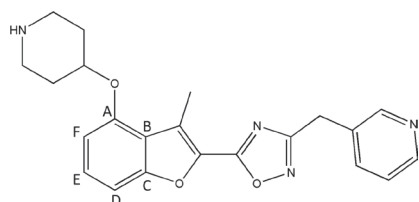pIC_{50}: negative logarithm concentration capable of inhibiting 50% of the enzyme activity.

Figure 1. Ordered atom letter codes (A, B, C, D, E and F) used in the 4D-QSAR analysis defines the three trial alignments. Compound **76** ($pIC_{50} = 8.00$) is used to define the atom letter code.

4D-QSAR model calculation

The two hundred GCODs that presented the highest weight of the data set were selected by least squares regression (PLS) method,[16] and then these GCODs were optimized using a combined genetic function approximation (GFA) approach,[17] implemented in the 4D-QSAR program.[18] The GFA measures the quality of the models from the statistical parameters, such as the least-square error of fit (LSE) and Friedman's lack-of-fit (LOF). This measure penalizes appropriately for the addition of terms to the equation (and consequent loss of degrees of freedom) in such a way to resist over-fitting.

Thus, the calculations were initiated using 100 randomly generated models and 10,000-100,000 crossover operations. Mutation probability over the crossover optimization cycle was set at 10-30%. The smoothing factor, the variable that specifies the number of descriptors in the QSAR models, was varied between 1.0 and 3.0 in order to obtain equations with no more than eleven terms.

Model validation

Validation is an important factor in QSAR modeling. The best models, resulting from the 4D-QSAR study were based on different criteria.[19-23]

(*i*) Leave-one-out cross-validation (LOOcv) correlation coefficient (q^2): estimating the performance of a predictive model;

(*ii*) Adjusted cross-validated squared correlation coefficient (q^2_{adj}): allows the comparison between models with different number of variables;

(*iii*) Correlation coefficient of external validation set (R^2_{pred}): reflects the degree of correlation between the observed ($Y_{Exp(test)}$) and predicted ($Y_{Pred(test)}$) activity data of the test set with equation 1:

$$R^2_{Pred} = 1 - \frac{\sum_1^n (Y_{Exp(test)} - Y_{Pred(test)})^2}{\sum_1^n (Y_{Exp(test)} - \bar{Y}_{training})^2} \qquad (1)$$

where $\bar{Y}_{training}$ is average value for the dependent variable for the training set;

(*iv*) Modified r^2 ($r^2_{m\,(test)}$) equation determining the proximity between the observed and predicted values with the zero axis intersection, equation 2:

$$r^2_{m\,(test)} = R^2 \left(1 - \left|R^2 - R^2_0\right|\right) \qquad (2)$$

(*v*) Y-Randomization (R^2_r): consists of the random exchange of the independent variable values. Thus, the R^2_r value must be less than the correlation coefficient of the non-randomized models;

(*vi*) R^2_p: penalizes the model R^2 for the difference between the squared mean correlation coefficient (R^2_r) of randomized models and the square correlation coefficient (R^2) of the non-randomized model, equation 3:

$$R^2_p = R^2 \sqrt{R^2 - R^2_r} \qquad (3)$$

Conformational selection

In the 4D-QSAR method, the conformation of

each compound can be postulated as the lowest-energy conformer state from the set sampled for each compound, which predicted the maximum activity using the optimum 4D-QSAR model.[15]

Docking studies

In order to investigate the intermolecular interactions between the inhibitors and the protein, the molecular docking method was used. The crystal structural coordinates of NMT at 1.42 Å resolution was obtained from the Protein Data Bank (PDB code: 2WUU),[5] and the protein has been crystallized in complex with the non-hydrolysable substrate analogue *S*-(2-oxo)pentadecyl-CoA. For the docking procedure the Molegro Virtual Docker (MVD) program was used.[24] Molecular docking uses the receptor structure as a template for the development of new ligands, estimating binding affinity between linker and receptor.[25]

Molecular docking simulations were performed using the MOLDOCK optimizer algorithm, implemented in the VMD, capable of accurately identifying the probable conformations and orientations of the ligand (poses) in the protein interaction site.[26]

In order to evaluate the quality of the docking pose, the MolDock score [GRID] was used as a scoring function, based on a linear potential by parts (PLP), a simplified potential whose parameters are fit to protein-ligand structures and binding data scoring functions.[27]

The binding sites were restricted within spheres of 12 Å radius for the study of docking. Moreover, multiple runs were performed for each compound, generating a total of 100 poses each, this procedure was necessary to avoid random results, due to the stochastic nature of the docking algorithm.

Results and Discussion

The QSAR models were built using the pIC_{50} as dependent variable and the descriptors from 4D-QSAR

analysis (the GCODs) as independent variables. The GCODs are represented by the Cartesian coordinates and the corresponding IPE - ("x, y, z, IPE"). The use of IPEs allows the compounds to be partitioned into substructures with respect to possible interactions that may occur with a common receptor, thereby allowing for the identification of relevant features responsible for the biological activity and, ultimately, the proposal of structural modifications in order to increase their potency. Nevertheless, the 4D-QSAR methodology generates a large number of GCODs due to the number of grid cells and IPEs. Therefore, the 4D-QSAR models were built from the most weighted GCODs based on the PLS-GA analysis.[15,16]

In the GFA methodology, the QSAR models are ranked by the Friedman's lack of fit (LOF) measure, equation 4:

$$LOF = \frac{LSE}{\left(1 - \dfrac{c + dp}{M}\right)^2} \qquad (4)$$

In equation 4, c is the number of basis function (terms), d is the smoothing factor (the only user-defined parameter, wherein larger values of d lead to model with fewer terms), p is the total number of features in all basis function, M is the number of compounds in the training set and LSE is the least-square error. The LOF act as a penalized LSE, i.e., when two models have the same LSE, the one that has the lowest number of terms will have the best (lowest) value of LOF, then resisting overfitting.[16,17]

Therefore, as mentioned above, in this study seven alignments were evaluated for the three test sets. The statistical parameters of each alignment for test set I, II and III are shown in Tables 2, 3 and 4, respectively.

All tested alignments showed q^2 and q^2_{adj} values higher than 0.5. This reveals that the model can be a useful tool for predicting affinities of new compounds based on these structures; r^2 greater than 0.7 indicates that the model is

Table 2. Statistical parameters evaluated in the 4D-QSAR analysis for the seven performed alignments of the test set I

Alignment	R^2	q^2	q^2_{adj}	LSE	LOF	R^2_{Pred}	$r^2_{m\,(test)}$	R^2_r	R^2_p	GCODs
A1	0.76	0.70	0.65	0.21	0.39	0.74	0.54	0.11	0.61	5
A2	0.73	0.66	0.61	0.22	0.49	0.82	0.48	0.09	0.68	5
A3	0.74	0.69	0.64	0.24	0.52	0.67	0.45	0.08	0.70	5
A4	0.69	0.69	0.64	0.24	0.52	0.67	0.45	0.09	0.70	5
A5	0.72	0.66	0.61	0.24	0.72	0.73	0.51	0.08	0.68	5
A6	0.74	0.68	0.63	0.23	0.41	0.75	0.45	0.07	0.71	5
A7	0.71	0.64	0.59	0.24	0.53	0.77	0.58	0.10	0.66	5

R^2: coefficient of determination; q^2: leave-one-out cross-validation correlation coefficient; q^2_{adj}: adjusted cross-validated squared correlation coefficient; LSE: least-square error of fit; LOF: Friedman's lack-of-fit; R^2_{pred}: correlation coefficient of external validation; $r^2_{m(test)}$: equation 2; R^2_r: Y-randomization; R^2_p: defined in equation 3; GCODs: grid cell occupancy descriptors.

Table 3. Statistical parameters evaluated in the 4D-QSAR analysis for the seven performed alignments of the test set II

Alignment	R^2	q^2	q^2_{adj}	LSE	LOF	R^2_{Pred}	$r^2_{m (test)}$	R^2_r	R^2_p	GCODs
B1	0.74	0.69	0.64	0.20	0.37	0.69	0.52	0.08	0.71	5
B2	0.72	0.66	0.61	0.23	0.50	0.78	0.55	0.08	0.68	5
B3	0.70	0.63	0.59	0.24	0.53	0.77	0.55	0.08	0.65	5
B4	0.64	0.57	0.54	0.28	0.59	0.76	0.44	0.08	0.60	4
B5	0.72	0.66	0.61	0.22	0.49	0.78	0.58	0.08	0.68	5
B6	0.70	0.64	0.59	0.22	0.41	0.83	0.63	0.08	0.66	5
B7	0.73	0.67	0.62	0.24	0.52	0.69	0.49	0.08	0.69	5

R^2: coefficient of determination; q^2: leave-one-out cross-validation correlation coefficient; q^2_{adj}: Adjusted cross-validated squared correlation coefficient; LSE: least-square error of fit; LOF: Friedman's lack-of-fit; R^2_{pred}: correlation coefficient of external validation; $r^2_{m(test)}$: equation 2; R^2_r: Y-randomization; R^2_p: defined in equation 3; GCODs: grid cell occupancy descriptors.

Table 4. Statistical parameters evaluated in the 4D-QSAR analysis for the seven performed alignments of the test set III

Alignment	R^2	q^2	q^2_{adj}	LSE	LOF	R^2_{Pred}	$r^2_{m (test)}$	R^2_r	R^2_p	GCODs
C1	0.74	0.68	0.63	0.21	0.38	0.74	0.57	0.07	0.70	5
C2	0.72	0.65	0.60	0.23	0.50	0.75	0.52	0.08	0.67	5
C3	0.71	0.64	0.59	0.25	0.60	0.65	0.33	0.11	0.65	5
C4	0.71	0.65	0.60	0.23	0.51	0.66	0.37	0.12	0.65	5
C5	0.73	0.66	0.61	0.22	0.49	0.72	0.46	0.09	0.68	5
C6	0.75	0.69	0.64	0.22	0.49	0.64	0.51	0.08	0.70	5
C7	0.72	0.65	0.60	0.24	0.52	0.73	0.56	0.08	0.67	5

R^2: coefficient of determination; q^2: leave-one-out cross-validation correlation coefficient; q^2_{adj}: Adjusted cross-validated squared correlation coefficient; LSE: least-square error of fit; LOF: Friedman's lack-of-fit; R^2_{pred}: correlation coefficient of external validation; $r^2_{m(test)}$: equation 2; R^2_r: Y-randomization; R^2_p: defined in equation 3; GCODs: grid cell occupancy descriptors.

correlated and may be considered to represent the training set in the same manner.[13] Thus, the alignments A4 and B4 were eliminated from the analysis because they had r^2 lower than 0.7.

Regarding external validation, R^2_{pred} values for all alignments were greater than 0.5. Analyzing $R^2_{m(test)}$ values, alignments A2, A3, A4, A6, B4, B7, C3, C4 and C5 were excluded, since they should be higher than 0.5.

Another parameter used in the validation of the models was R^2_p and all values are greater than 0.5, which means that R^2_p values are acceptable for a good QSAR model. In addition, all R^2_r values are well below those of R^2. As alignment A1 from test set I showed the highest R^2, q^2 and q^2_{adj} values, it was chosen as the best alignment in this 4D-QSAR study. In addition, alignment A1 presents good external validation values. We will only present the analysis of the best model derived from A1.

Equation 5 presents Model A1. The statistical measures, including the values of R^2, q^2, q^2_{adj}, LSE, LOF, R^2_{pred}, $r^2_{m(test)}$, R^2_p and R^2_r are presented below.

$$pIC_{50} = 5.76 + 1.19(0,-2,4,np) - 0.56(0,3,2,any) + 1.46(3,-2,-1,np) + 2.03(0,-3,2,any) - 0.95(0,4,-1,any) \quad (5)$$

n = 58, GCODs = 5, R^2 = 0.76, q^2 = 0.70, q^2_{adj} = 0.65, LSE = 0.21, LOF = 0.39,
R^2_{pred} = 0.74, $r^2_{m (test)}$ = 0.54, R^2_r = 0.11, R^2_p = 0.61

Model A1 generated five descriptors, where GCODs (0,−2,4,np), (3,−2,−1,np), (0,−3,2,any), present positive coefficients (equation 5) and correspond to favorable interactions between the molecule substituent and amino acid residues in the active site of NMT. Thus, substituents in these positions increase the effectiveness of the compounds. GCODs (0,3,2,any) and (0,4,−1,any) have negative coefficients and correspond to unfavorable interactions between the molecule substituent and amino acid residues in the active site of NMT. Thus, the occupation of GCODs (0,3,2,any) and (0,4,−1,any) decreases the compound potency.

The cross-correlation matrix of the GCODs from Model A1 (equation 5, Table 5) was computed in order to determine if two or more highly correlated GCODs appear in the same 4D-QSAR model. We can observe that there is no correlation (r > 0.7) between the GCODs. Therefore, each descriptor contributes in different ways to the 4D-QSAR models.[23]

Table 5. Cross-correlation matrix of the GCODs (grid cell occupancy descriptors) of Model A1 obtained in the 4D-QSAR analysis

|	(0,−2,4,1)	(0,3,2,0)	(3,−2,−1,1)	(0,−3,2,0)	(0,4,−1,0)
(0,−2,4,1)	1				
(0,3,2,0)	0.245	1			
(3,−2,−1,1)	−0.403	−0.416	1		
(0,−3,2,0)	−0.324	0.069	0.098	1	
(0,4,−1,0)	−0.27	0.222	−0.037	−0.216	1

Figure 2 shows William's plot for the Model A1. As can be observed, all compounds fall inside the red dashed area, thus ensuring the lack of influential samples and outliers in this dataset.
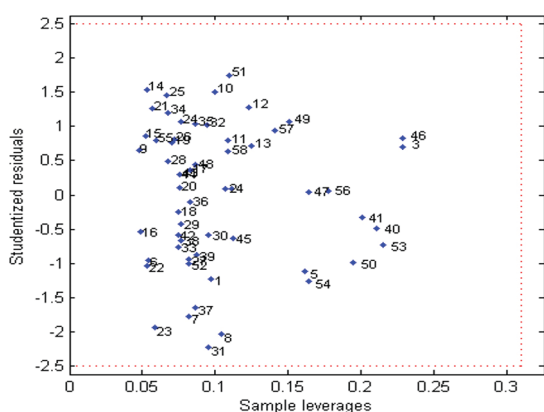
Figure 2. Plot of sample leverage *versus* Student residuals for the Model A1.

A graphic representation of the descriptors of Model A1 is shown in Figure 3 using compound **76** as a reference. Light and dark spheres represent GCODs with positive and negative coefficients, respectively, in accordance with equation 5 (Model A1).

GCOD-1 (0,-2,4,np), Figure 4, has a positive coefficient of 1.19 and has a nonpolar IPE. This grid cell shows the highest frequency of occupation for compounds having a benzyl group at this site of the molecule, such as molecules **2**, **11** and **15** (pIC$_{50}$ = 6.22, 5.15 and 6.48, respectively). In contrast, compounds with only one phenyl group have no occupancy in this grid cell. Furthermore, the molecules **61**, **75** and **76**, which are the most active in this series (pIC$_{50}$ = 7.70, 7.76 and 8.00, respectively), have no occupancy for this descriptor, which is favorable for the potency of the compounds.

GCOD-2 (0,3,2,any) (Figure 5) contributes to decrease compound potency and presents a coefficient of −0.56. This grid cell represents a non-specific IPE. The great majority of the compounds present a high GCOD-2 occupation of frequency and have a hydrogen atom in this grid cell. However, compounds **53**, **54**, **62** and **70** (pIC$_{50}$ = 6.06,
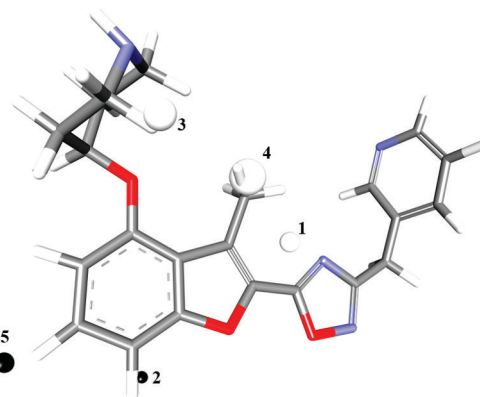
Figure 3. Graphic representation of compound **76** according to the 4D-QSAR Model A1. White spheres represent GCODs that positively contribute to the potency of the compounds, while black spheres represent GCODs that negatively contribute. Spheres sizes are depicted according to GCODs frequency of occupation, i.e., bigger spheres represent GCODs with higher frequency. The GCODs described are: (**1**) (0,-2,4,np), (**2**) (0,3,2,any), (**3**) (3,-2,-1,np), (**4**) (0,-3,2,any), (**5**) (0,4,-1,any).
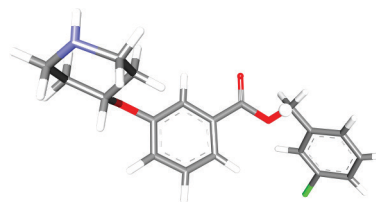
Figure 4. Representation of compound **15** and GCOD-1 (0,-2,4,np) (white sphere) obtained from Model A1.
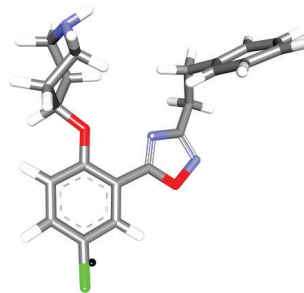
Figure 5. Representation of compound **54** and GCOD-2 (0,3,2,any) (black sphere) obtained from Model A1.

6.34, 7.07 and 7.00, respectively) present low occupancy values for this negative GCOD and in this position, have the following substitutions: CH$_3$, Cl, O−CH$_3$ and Br. Thus, the presence of a hydrogen atom in this grid cell results in

a decrease in the potency of the compounds. Among the most active compounds, molecules **75** and **76** (pIC$_{50}$ = 7.76 and 8.00) have the hydrogen atom in this position, which is detrimental to their activity.

GCOD-3 (3,-2,-1,np) shows a positive coefficient and corresponds to a nonpolar type (IPE) (Figure 6). The GCOD (3,-2,-1,np) is located near the hydrogen atom in the piperidine ring and has the highest frequency of occupation for compounds **61**, **70**, **73** and **75** (pIC$_{50}$ = 7.70, 7.00, 6.19 and 7.76, respectively). Although most of the compounds present the piperidine ring in the same position, not all of them show high occupancy for this descriptor, probably because during the molecular dynamics simulation they assumed different conformations, as exemplified by compound **17** (pIC$_{50}$ = 5.52). This is also demonstrated by the docking results (see below), once the potential binding site is too large, the compounds have a high degree of conformational freedom, especially in the portion near to the piperidine ring.
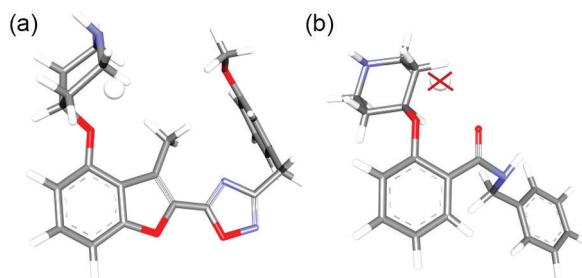

Figure 6. Representation of GCOD-3(3,-2,-1,np) (white sphere) and (a) compound **75**; (b) compound **17**.

The GCOD-4 (0,-3,-,2,any) (Figure 7) also has a positive coefficient of 2.03 and is the descriptor that contributes most to the increased effectiveness of the compounds. This grid cell represents a non-specific IPE and is located near the methyl group in benzofuran, 2,3-dihydro-3-methyl. This descriptor shows the highest occupation frequency for compounds **1**, **75**, **76** and **77** (pIC$_{50}$ = 6.30, 7.76, 8.00 and 7.15, respectively), which have benzofuran, 2,3-dihydro-3-methyl in their structures. Thus, potential inhibitors would benefit from the exploitation of this region with methyl groups.
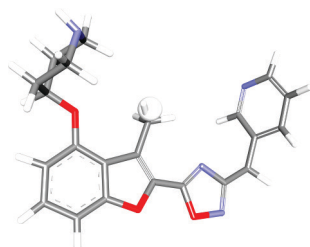

Figure 7. Representation of compound **76** and GCOD-4 (0,-3-,2,any) (white sphere) obtained from Model A1.

Finally, the GCOD-5 (0,4,-1,any) (Figure 8) corresponds to a non-specific (IPE) and presents the most negative coefficient, impairing the activity of the compounds. Unlike GCOD-2, the substitution by CH$_3$, Cl, F, O–CH$_3$ and Br atoms or another benzene ring forming a naphthalene group in the GCOD-5, is detrimental to the potency of the compounds.
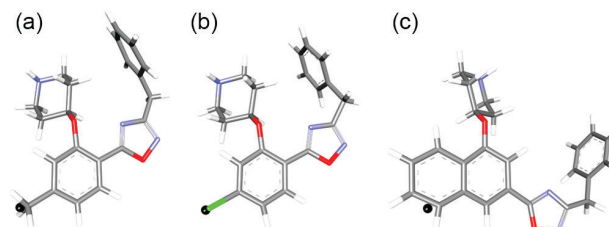

Figure 8. Representation of GCOD-5(0,4,-1,any) (black sphere) (a) compound **42**; (b) compound **48**; (c) compound **57**.

Based on these information, the structures of compounds **A**, **B** and **C** were proposed, and their activities were predicted using Model A1. The predicted activity of compound **76**, the most active in the series, was 7.57. The structure of the three compounds and their predicted pIC$_{50}$ values are shown in Figure 9 and all of them showed predicted activity values higher than compound **76**.
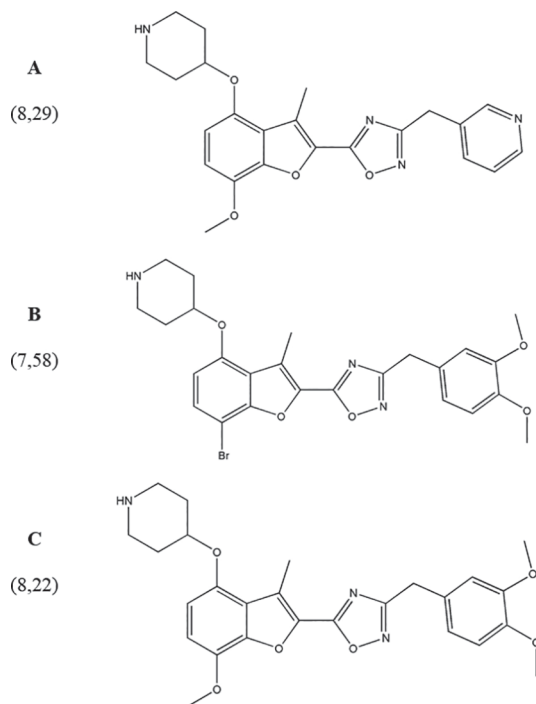

Figure 9. Proposed compounds using the Model A1 (predicted pIC$_{50}$ values are shown in parenthesis).

Afterwards, in order to further understand the behavior of the studied compounds inside the binding pocket of NMT,

docking studies were performed for proposed structures (**A**, **B** and **C**) as well as for compound **76**. The NMT structure were extracted from the Protein Database (PDBID:2WUU) and potential binding sites were automatically identified by MVD. A cavity of 351.65 Å^3 (surface = 1208.16 Å^2) around the amino acid residues Tyr80, Val81, Glu82, Ser86, Met87, Phe88, Arg89, Phe90, Tyr92, Phe96, Asn167, Phe168, Thr203, Ala204, Gly205, Val206, Tyr217, Phe218, His219, Phe232, Tyr326, Ile328, Pro329, Ser330, Leu341, Ala343, Tyr345, Val346, Val374, Asn376, Met377, Val378, Ile380, Leu381, Asn383, Gly395, Asp396, Gly397, His398, Leu399, Tyr401, Val419, Met420 and Leu421 was then selected as the binding site for docking the ligands. This cavity was selected based on structural information from other NMT analogous.[28]

By analyzing the hydrogen bond formed between the compounds and the NMT active site, we observed: (*i*) the proposed compounds **A** and **C** establish a hydrogen bond with: Tyr217, Tyr326 and Tyr345; however, compounds **76** and **B** do not realize a hydrogen bond with the amino acid Tyr326. This portion of the molecule must be related to the GCOD-2, in which, the substitution of a hydrogen atom by different groups (CH$_3$, Br, Cl and O–CH$_3$) is favorable, as we have seen. The docking analysis also allows concluding that the presence of a group capable of hydrogen bonding with Tyr326 makes the compound more active. The structural difference between compounds **A** and **76** is that compound **A** has a group capable of hydrogen bonding with the amino acid Tyr326, while compound **76** has a hydrogen atom,

which is unfavorable. Thus, the activity value predicted by the Model A1 for compound **A** is much higher than that of compound **76**, as well as the interaction energy of the docking (Table 6) Thus, molecules **A** and **C** have better values of predicted activity, which was validated by the energy values of the docking, which reflect the importance of the interaction with the amino acid Tyr326; (*ii*) compound **A** still makes a hydrogen bond with the amino acid Gly 397, near the pyridine; (*iii*) compounds **B** and **C** interacted with Ser330 by a hydrogen bond and, therefore, have better values of predicted activity and molecular anchor energy than molecule **76**, confirming the importance of groups such as O–CH$_3$ for the potency of the compounds. Figure 10 shows the docking poses for the compounds analyzed inside the NMT active site and the interacting amino acid residues.

Table 6. Predicted pIC$_{50}$ and interaction energy values of compound/protein

Compound	Predicted pIC$_{50}$	Interaction energy / (kcal mol^{-1})
76	7.57	−109.61
A	8.29	−126.19
B	7.58	−116.44
C	8.22	−127.32

pIC$_{50}$: negative logarithm concentration capable of inhibiting 50% of the enzyme activity.

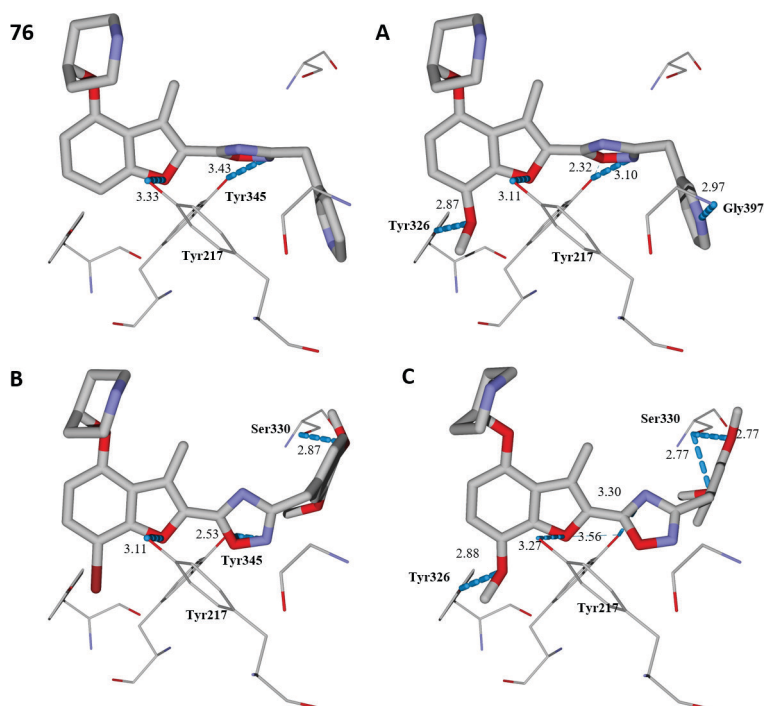In addition to the hydrogen bonding interactions, it was possible to note the presence of an apolar moiety which

Figure 10. Docking poses for the compounds **76**, **A**, **B**, **C** and main interactions (hydrogen bonds).
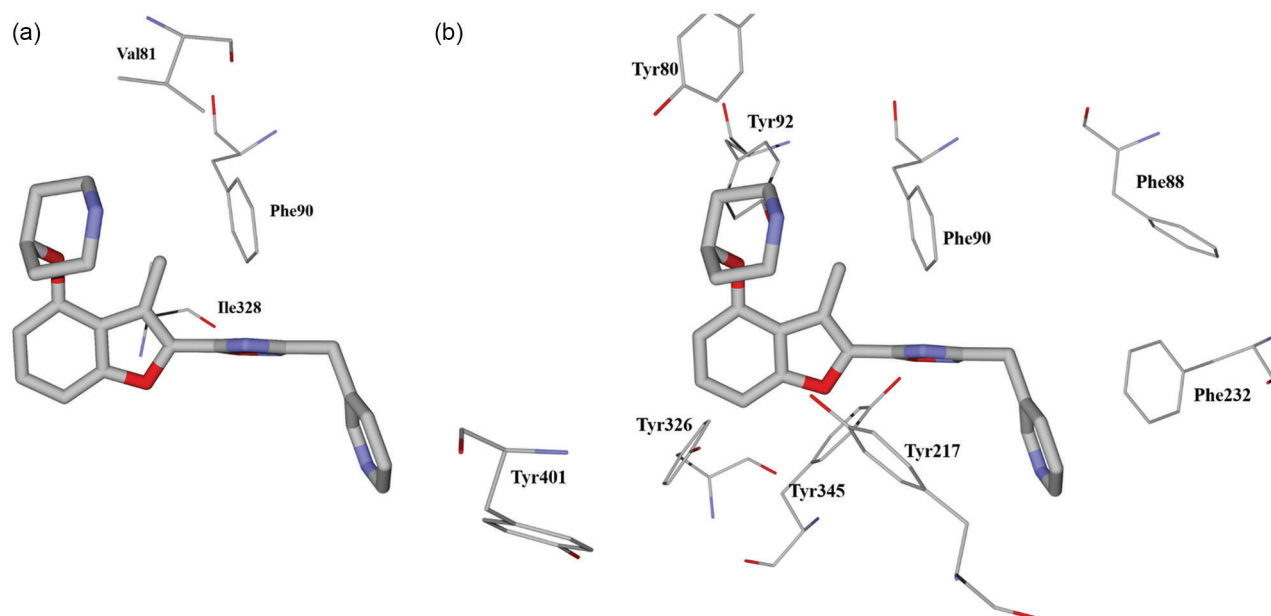
Figure 11. (a) Docking pose for the compound **76** and apolar moiety next to the methyl group in the NMT active site; (b) docking pose for the compound **76** and aromatic amino acids in the NMT active site.

includes the amino acids Val81, Ile328 and Phe90 located near the methyl in benzofuran, 2,3-dihydro-3-methyl group (Figure 11a), as indicated by the GCOD-4. The presence of several aromatic amino acids in the active site, Figure 11b, as Tyr80, Phe88, Phe90, Tyr92, Tyr217, Phe232, Tyr326, Tyr345 and Tyr401 also contributes to the π-π staking interaction between these amino acids and aromatic moieties of inhibitors stabilizing them.

The absorption, distribution, metabolism and excretion (ADME) are important pharmacokinetic properties that must be met in the elaboration of a new drug. In this respect, the Lipinski's Rule of Five (RO5) is a widely used test to estimate the drug likeness of a compound. If a compound fails in this test, it is likely that some oral bioavailability problem will be encountered.[29] According to the RO5, the compounds should present $logP \leq 5$, molecular weight ≤ 500, number of hydrogen bond acceptors (n_{ON}) ≤ 10, number of hydrogen bond donors (n_{OHNH}) ≤ 5, and number of rotatable bonds (n_{rotb}) ≤ 10, wherein compounds should not violate more than one rule. In this sense, molecules **A**, **B** and **C** were constructed in the Molinspiration Online Property Calculation Software Toolkit[30] for the evaluation of the Lipinski's Rule. As can be seen in Supplementary Information section (Table S1), the results for the proposed structures are encouraging because all of them satisfy the criteria established by RO5, there is no more than one violation.

Due to the fact that the proposed compounds presented better binding energy values and predicted activity than the evaluated series of compounds, as well as acceptable parameters in ADME *in silico* tests, they can be considered promising candidates in the treatment of leishmaniasis.

Conclusions

In this work, the methodology of 4D-QSAR was used for the study of a series of 77 inhibitors of *Leishmania donovani* that were selected from the literature. As an advantage over others, this methodology allows identification of groups that are important for the activity of the compounds, thus facilitating the design of new structures that can be more active and selective. This study was performed applying three test groups and seven alignments. The best model, Model A1, presents important features that can be applied in the development of new NMT inhibitors. The Model A1 showed R^2, q^2, and values of 0.76 and 0.70, respectively. In addition, it has external validation values of $R^2_{pred} = 0.74$ and $r^2_{m\,(test)} = 0.54$. Moreover, based on the descriptors obtained from Model A1, compounds **A**, **B** and **C** were proposed, which demonstrated biological activity values superior to those of the most active compound. These results were corroborated by docking studies and the ADME evaluation. Consequently, the proposed compounds may be considered as promising drug candidates for the treatment of leishmaniasis.

Supplementary Information

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

Acknowledgments

References

1. World Health Organization (WHO); *WHO and Gilead Sciences Extend Collaboration against Visceral Leishmaniasis*; Geneva, 2016.

2. de Menezes, J. P. B.; Guedes, C. E. S.; Petersen, A.; Fraga, D. B. M.; Veras, P. S. T.; *BioMed. Res. Int.* **2015**, *2015*, 815023.

3. Ghosh, S.; Kar, N.; Bera, T.; *Int. J. Biol. Macromol.* **2016**, *93*, 961.

4. Goldston, A. M.; Sharma, A. I.; Paul, K. S.; Engman, D. M.; *Trends Parasitol.* **2014**, *30*, 350.

5. Brannigan, J. A.; Smith, B. A.; Yu, Z. Y.; Brzozowski, A. M.; Hodgkinson, M. R.; Maroof, A.; Price, H. P.; Meier, F.; Leatherbarrow, R. J.; Tate, E. W.; Smith, D. F.; Wilkinson, A. J.; *J. Mol. Biol.* **2010**, *396*, 985.

6. Singh, N.; Shah, P.; Dwivedi, H.; Mishra, S.; Tripathi, R.; Sahasrabuddhe, A. A.; Siddiqi, M. I.; *Mol. BioSyst.* **2016**, *12*, 3711.

7. Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B. Q.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C.; *J. Am. Chem. Soc*. **1997**, *119*, 10509.

8. Leatherbarrow, R.; Tate, E.; Yu, Z.; Rackham, M.; *WO pat. 2013/083991* **2013**.

9. Rose, S.; Stevens, A.; *Curr. Opin. Chem. Biol*. **2003**, *7*, 331.

10. Yu, H.; Adedoyin, A.; *Drug Discovery Today* **2003**, *8*, 852.

11. Kennard, R. W.; Stone, L. A.; *Technometrics* **1969**, *11*, 148.

12. da Cunha, E. F. F.; Albuquerque, M. G.; Antunes, O. A. C.; de Alencastro, R. B.; *Mol. Inf.* **2005**, *2*, 25.

13. *HyperChem Release 7*; *HyperCube Inc*; Gainesville, Florida, 2002.

14. Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P.; *J. Comput. Chem.* **2006**, *27*, 1101.

15. Hopfinger, A.; *4D-QSAR Package User's Manual*; *The Chem21 Group, Inc.:1780 Wilson Drive*; Lake Forest, USA, 2001.

16. Romeiro, N. C.; Albuquerque, M. G.; de Alencastro, R. B.; Ravi, M.; Hopfinger, A. J.; *J. Comput.-Aided Mol. Des*. **2005**, *19*, 385.

17. Rogers, D.; Hopfinger, A. J.; *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.

18. Allinger, N. L.; *J. Am. Chem. Soc*. **1977**, *99*, 8127.

19. Kiralj, R.; Ferreira, M. M. C.; *J. Braz. Chem. Soc*. **2009**, *20*, 770.

20. Kar, S.; Roy, K.; *Indian J. Biochem. Biophys*. **2011**, *48*, 111.

21. Roy, K.; Paul, S.; *QSAR Comb. Sci.* **2009**, *28*, 406.

22. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P. K.; Kar, S.; Das, R. N.; *J. Comput. Chem*. **2013**, *34*, 1071.

23. Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C. P.; Agrawal, R. K.; *Int. J. Drug Des. Discovery* **2011**, *3*, 511.

24. Thomsen, R.; Christensen, M. H.; *J. Med. Chem.* **2006**, *49*, 3315.

25. Sodero, A. C. R.; Romeiro, N. C.; da Cunha, E. F. F.; Magalhaes, U. D.; de Alencastro, R. B.; Rodrigues, C. R.; Cabral, L. M.; Castro, H. C.; Albuquerque, M. G.; *Molecules* **2012**, *17*, 7415.

26. Assis, T. M.; Mancini, D. T.; Ramalho, T. C.; da Cunha, E. F. F.; *E-J. Chem.* **2014**, *2014*, 1.

27. Hehre, W. J.; Deppmeier, B. J.; Klunzinger, P. E.; *PC SPARTAN Pro, Wavefunction, Inc.*; Irvine, California, 1999.

28. Brannigan, J. A.; Roberts, S. M.; Bell, A. S.; Hutton, J. A.; Hodgkinson, M. R.; Tate, E. W.; Leatherbarrow, R. J.; Smith, D. F.; Wilkinson, A. J.; *IUCrJ* **2014**, *1*, 250.

29. Patel, H. M.; Sing, B.; Bhardwaj, V.; Palkar, M.; Shaikh, M. S.; Rane, R.; Alwan, W. S.; Gadad, A. K.; Noolvi, M. N.; Karpoormath, R.; *Eur. J. Med. Chem.* **2015**, *93*, 599.

30. http://www.molinspiration.com, accessed in December 2017.