

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/45798718>

# QSAR and docking studies of novel antileishmanial diaryl sulfides and sulfonamides

ARTICLE *in* EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY · NOVEMBER 2010

Impact Factor: 3.45 · DOI: 10.1016/j.ejmech.2010.07.060 · Source: PubMed

CITATIONS

22

READS

64

## 4 AUTHORS, INCLUDING:



[Mohammad Goodarzi](#)

Vrije Universiteit Brussel

107 PUBLICATIONS 922 CITATIONS

SEE PROFILE



[Matheus Freitas](#)

Universidade Federal de Lavras (UFLA)

109 PUBLICATIONS 1,111 CITATIONS

SEE PROFILE



[Teodorico C. Ramalho](#)

Universidade Federal de Lavras (UFLA)

188 PUBLICATIONS 1,593 CITATIONS

SEE PROFILE



## Original article

## QSAR and docking studies of novel antileishmanial diaryl sulfides and sulfonamides

Mohammad Goodarzi<sup>a</sup>, Elaine F.F. da Cunha<sup>b</sup>, Matheus P. Freitas<sup>b</sup>, Teodorico C. Ramalho<sup>b,\*</sup>

<sup>a</sup> Instituto de Investigaciones Físicoquímicas Teóricas y Aplicadas (INIFTA), UNLP, CCT La Plata-CONICET, Diag. 113 y 64, C.C. 16, Suc.4, 1900 La Plata, Argentina

<sup>b</sup> Departamento de Química, Universidade Federal de Lavras, C.P. 3037, 37200-000 Lavras, MG, Brazil

## ARTICLE INFO

## Article history:

Received 26 May 2010

Received in revised form

25 July 2010

Accepted 29 July 2010

Available online 7 August 2010

## Keywords:

Leishmaniasis

QSAR

Docking

Diaryl sulfide compounds

Sulfonamides

## ABSTRACT

Leishmaniasis is a neglected disease transmitted in many tropical and sub-tropical countries, with few studies devoted to its treatment. In this work, the activities of two antileishmanial compound classes were modeled using Dragon descriptors, and multiple linear (MLR) and support vector machines (SVM) as linear and nonlinear regression methods, respectively. Both models were highly predictive, with calibration, leave-one-out validation and external validation  $R^2$  of 0.79, 0.72 and 0.78, respectively, for the MLR-based model, improving significantly to 0.98, 0.93 and 0.90 when using SVM modeling. Therefore, novel compounds were proposed using the QSAR models built by combining the substructures of the main active compounds of both classes. The most promising structures were docked into the active site of *Leishmania donovani*  $\alpha$ , $\beta$  tubulin (*Ld-Tub*), demonstrating the high affinity of some new structures when compared to existing antileishmanial compounds.

© 2010 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

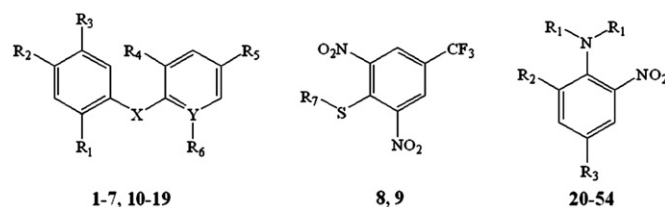
Leishmaniasis is a tropical and sub-tropical diseases transmitted by the bite of female phlebotomine sandflies infected with the pathogen *Leishmania*. It affects as many as 12 million people worldwide and the visceral form of leishmaniasis has an estimated incidence of 500,000 new cases and 60,000 deaths each year [1,2]. Current treatments involve the use of antimony-containing compounds and amphotericin, among others, but these therapies have shown to be not entirely efficient, due to parasite resistance and/or failure in some cases to treat visceral leishmaniasis (*Leishmania donovani*) [3–5]. Novel therapies to combat leishmaniasis are urgently needed, because for the treatment of Leishmaniasis the currently used drugs are limited to four [6]. The first line compounds are the two pentavalent antimonials, sodium stibogluconate and meglumine antimoniate. Therefore, in most of the countries where this disease is present Glucantime<sup>®</sup> and Pentostan<sup>®</sup> are still the drugs to fight against Leishmaniasis [7,8]. These kinds of drugs show severe limitations (parenteral administration, price, toxicity, variable efficacy). Furthermore, the resistance to drug treatment is common, especially in endemic regions, which further complicates the panorama of the disease. This series

outlook reveals the need to develop new and more selective drugs in order to combat Leishmaniasis. Recently, several efforts in this direction have been performed. For instance, 1-phenyl-4-glycosyl-dihydropyridines [9], 4-aryl coumarins [10], chalconoids [11] derivatives and a series of 5-nitro-2-heterocyclic benzyldiene hydrazides [12], 2,4,6-trisubstituted pyrimidines and 1,3,5-triazines [13] were prepared, revealing promising *in vitro* activity against Leishmaniasis. Furthermore, other theoretical studies involving docking [12] and QSAR [14] techniques have explored different molecular targets in order to design new antileishmanial compounds.

Two recent studies have been performed with the aim of understanding the action mechanism and to design new antileishmanial candidates with strong *in vitro* activity [15,16]. Aromatic nitro groups were found to have significant biological redox activity leading to the potentially cytotoxic production of reactive oxygen species; thus, subsequent compound screening in those works was based on nitro analogs [15,16]. A suitable way to design new drug leads and improves the potency and/or selectivity of existing congeners is based on the combination of substructures of the main derivatives of two or more different series to give a new compound. This procedure has been found to be successful at least in a couple of studies [17,18] and is proposed here to derive novel antileishmanial analogs of two compound classes previously described in the literature with activity against *L. donovani* [15,16]. In the first step, a QSAR modeling of the activities of series of diaryl sulfonyl and diaryl sulfonamide compounds was carried out. The

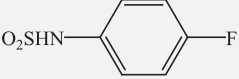
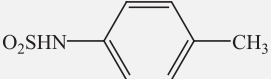
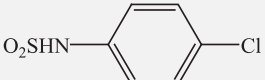
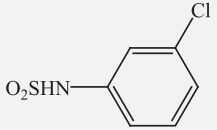

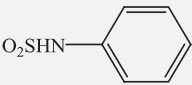
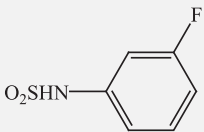
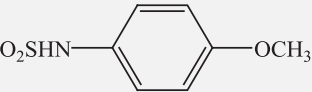
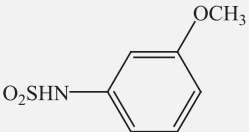
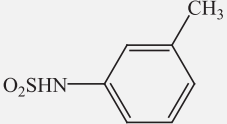
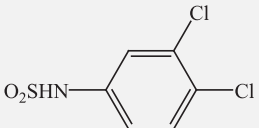
\* Corresponding author. Tel.: +55 35 3829 1891; fax: +55 35 3829 1271.

E-mail address: [teo@dqf.ufba.br](mailto:teo@dqf.ufba.br) (T.C. Ramalho).

**Table 1**Compounds used in the QSAR modeling and respective experimental pIC<sub>50</sub> (IC<sub>50</sub> in mol L<sup>-1</sup>).<sup>a</sup>

Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	X	Y	pIC <sub>50</sub>
<b>1<sup>b</sup></b>	Cl	Cl	CH <sub>3</sub>	NO <sub>2</sub>	CF <sub>3</sub>	NO <sub>2</sub>	S	C	6.252
<b>2</b>	H	H	H	NO <sub>2</sub>	CF <sub>3</sub>	NO <sub>2</sub>	S	C	6.018
<b>3</b>	H	H	CH <sub>3</sub>	NO <sub>2</sub>	CF <sub>3</sub>	NO <sub>2</sub>	S	C	5.796
<b>4</b>	Cl	Cl	H	NO <sub>2</sub>	CF <sub>3</sub>	NO <sub>2</sub>	S	C	6.398
<b>5</b>	H	OCH <sub>3</sub>	H	NO <sub>2</sub>	CF <sub>3</sub>	NO <sub>2</sub>	S	C	5.721
<b>6</b>	H	OH	H	NO <sub>2</sub>	CF <sub>3</sub>	NO <sub>2</sub>	S	C	5.481
<b>7</b>	CO <sub>2</sub> H	H	H	NO <sub>2</sub>	CF <sub>3</sub>	NO <sub>2</sub>	S	C	5.137
<b>8<sup>b</sup></b>	<b>R<sub>7</sub></b> = cyclohexyl group								4.456
<b>9<sup>b</sup></b>	<b>R<sub>7</sub></b> = ethyl group								4.009
<b>10<sup>b</sup></b>	Cl	Cl	CH <sub>3</sub>	NO <sub>2</sub>	H	NO <sub>2</sub>	S	C	4.638
<b>11</b>	Cl	Cl	CH <sub>3</sub>	NO <sub>2</sub>	CF <sub>3</sub>	H	S	C	4.229
<b>12</b>	Cl	Cl	CH <sub>3</sub>	NO <sub>2</sub>	CF <sub>3</sub>	—	S	N	4.008
<b>13</b>	Cl	Cl	CH <sub>3</sub>	H	NO <sub>2</sub>	H	S	C	4.055
<b>14</b>	Cl	Cl	CH <sub>3</sub>	NO <sub>2</sub>	NO <sub>2</sub>	CF <sub>3</sub>	S	C	6.207
<b>15</b>	Cl	Cl	CH <sub>3</sub>	NO <sub>2</sub>	CN	NO <sub>2</sub>	S	C	6.000
<b>16</b>	Cl	Cl	CH <sub>3</sub>	CF <sub>3</sub>	CF <sub>3</sub>	CF <sub>3</sub>	S	C	4.000
<b>17</b>	Cl	Cl	H	NO <sub>2</sub>	CF <sub>3</sub>	NO <sub>2</sub>	O	C	5.468
<b>18</b>	Cl	Cl	H	NO <sub>2</sub>	CF <sub>3</sub>	NO <sub>2</sub>	NH	C	4.602
<b>19</b>	H	OCH <sub>3</sub>	H	NO <sub>2</sub>	CN	NO <sub>2</sub>	S	C	6.174
<b>20 (Oryzalin)</b>	n-propyl	NO <sub>2</sub>	SO <sub>2</sub> NH <sub>2</sub>						4.187
<b>21</b>	n-propyl	NO <sub>2</sub>							5.301
<b>22</b>	n-propyl	NO <sub>2</sub>							5.432
<b>23</b>	n-propyl	NO <sub>2</sub>							4.222
<b>24</b>	n-propyl	NO <sub>2</sub>							4.677
<b>25</b>	n-propyl	NO <sub>2</sub>	SO <sub>2</sub> N(CH <sub>2</sub> CH <sub>3</sub> ) <sub>2</sub>						4.568
<b>26</b>	n-propyl	NO <sub>2</sub>	SO <sub>2</sub> NHCH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>						4.267
<b>27</b>	n-propyl	NO <sub>2</sub>	SO <sub>2</sub> N(CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub> ) <sub>2</sub>						4.259
<b>28</b>	n-propyl	NO <sub>2</sub>	SO <sub>2</sub> NH(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>						4.301
<b>29</b>	n-propyl	NO <sub>2</sub>							4.366
<b>30</b>	n-propyl	NO <sub>2</sub>							4.301
<b>31</b>	n-butyl	NO <sub>2</sub>							5.585

Table 1 (continued)

Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	X	Y	pIC <sub>50</sub>
<b>32<sup>b</sup></b>	n-butyl	NO <sub>2</sub>							5.252
<b>33</b>	n-propyl	NO <sub>2</sub>							4.495
<b>34</b>	n-propyl	NO <sub>2</sub>							4.886
<b>35<sup>b</sup></b>	n-propyl	NO <sub>2</sub>							5.259
<b>36<sup>b</sup></b>	n-propyl	NO <sub>2</sub>	SO <sub>2</sub> N((CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub> ) <sub>2</sub>						4.318
<b>37</b>	n-propyl	NO <sub>2</sub>	SO <sub>2</sub> NH(CH <sub>2</sub> ) <sub>4</sub> CH <sub>3</sub>						4.366
<b>38</b>	n-propyl	NO <sub>2</sub>	SO <sub>2</sub> NH(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>						4.585
<b>39</b>	n-propyl	NO <sub>2</sub>							4.328
<b>40</b>	n-propyl	H							4.366
<b>41</b>	n-butyl	NO <sub>2</sub>							5.244
<b>42</b>	H, n-propyl	NO <sub>2</sub>	SO <sub>2</sub> NH <sub>2</sub>						4.174
<b>43<sup>b</sup></b>	ethyl	NO <sub>2</sub>	SO <sub>2</sub> NH <sub>2</sub>						4.161
<b>44<sup>b</sup></b>	n-butyl	NO <sub>2</sub>	SO <sub>2</sub> NH <sub>2</sub>						4.699
<b>45</b>	n-propyl	H	SO <sub>2</sub> NH <sub>2</sub>						4.046
<b>46<sup>b</sup></b>	n-propyl	NO <sub>2</sub>	CONH <sub>2</sub>						4.495
<b>47</b>	n-propyl	NO <sub>2</sub>							4.119
<b>48</b>	n-propyl	NO <sub>2</sub>							5.091
<b>49</b>	n-propyl	NO <sub>2</sub>							4.638
<b>50<sup>b</sup></b>	n-propyl	NO <sub>2</sub>							5.301

(continued on next page)

Table 1 (continued)

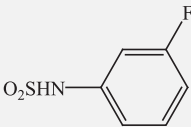
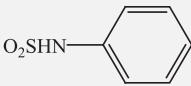
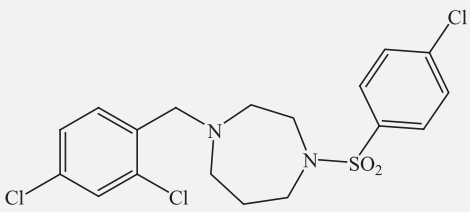
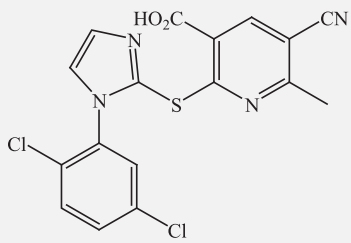
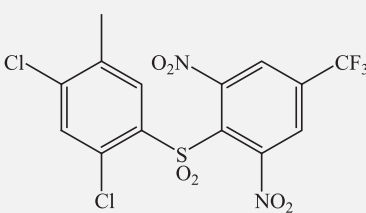
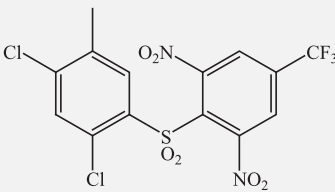
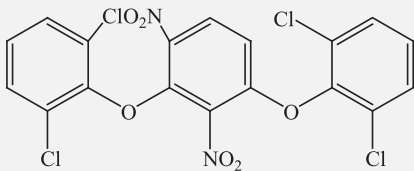
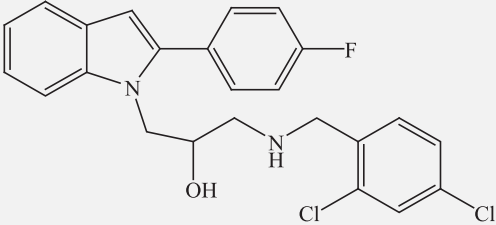
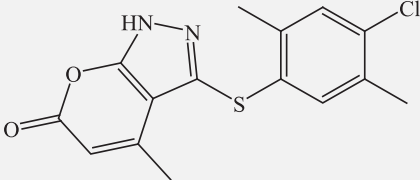
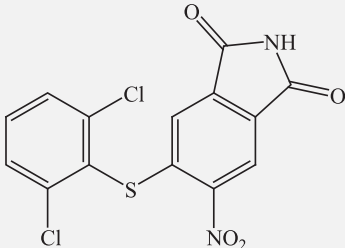
Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	X	Y	pIC <sub>50</sub>
51	n-propyl	NO <sub>2</sub>							5.602
52 <sup>b</sup>	ethyl	NO <sub>2</sub>							4.958
53	n-pentyl	NO <sub>2</sub>	SO <sub>2</sub> NH <sub>2</sub>						5.046
54	n-hexyl	NO <sub>2</sub>	SO <sub>2</sub> NH <sub>2</sub>						4.921
	Structure =								
55									4.143
	Structure =								
56									4.677
	Structure =								
57									6.301
	Structure =								
58									5.638
	Structure =								
59 <sup>b</sup>									4.097

Table 1 (continued)

Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	X	Y	pIC <sub>50</sub>
60 <sup>b</sup>	Structure =								4.432
61	Structure =								4.409
62	Structure =								4.222

<sup>a</sup> 1–19 from Ref. [6], and 20–62 from Ref. [7].<sup>b</sup> Test set compounds.

antileishmanial activities of novel compounds were then predicted using the calibration models built, and the most promising candidates were docked into the active site of *L. donovani*  $\alpha,\beta$  tubulin (Ld-Tub). Absorption, distribution, metabolism, excretion and toxicity (ADME–Tox) profiles were also calculated for the proposed compounds to test their druglikeness and safety.

## 2. Computational methods

### 2.1. QSAR modeling

The 2D structures of the molecules of Table 1 were drawn using HyperChem 7 software [19]. The final geometries were obtained with the semi-empirical AM1 method in the Hyperchem program. All calculations were carried out at the restricted Hartree–Fock level with no configuration interaction. The molecular structures were optimized using the Polak–Ribiere algorithm until the root-mean-square gradient reached  $0.001 \text{ kcal mol}^{-1}$  [20]. The resulting geometry was transferred into the Dragon program package [21,22] in order to obtain 1497 descriptors in Constitutional, Topological,

Geometrical, Charge, GETAWAY (Geometry, Topology and Atoms-Weighted Assembly), WHIM (Weighted Holistic Invariant Molecular descriptors), 3D-MorSE (3Dmolecular Representation of Structure based on Electron diffraction), Molecular Walk Count, BCUT, 2D Autocorrelation, Aromaticity Index, Randic molecular profile, Radial Distribution Function, Functional group and Atom-Centred Fragment classes. Genetic algorithm (GA) was used as the feature selection method to search for those descriptors with minimum collinearity among each other and with significant explained variance; the selected descriptors were T(N...N) (sum of topological distances between N...N), MATS6e (Moran autocorrelation weighted by atomic Sanderson electronegativities 2D autocorrelations), Mor20m (3D-MorSE weighted by atomic masses 3D-MorSE descriptors), E2m (2nd component accessibility directional WHIM index weighted by atomic masses WHIM descriptors), R1v<sup>+</sup> (maximal autocorrelation of lag 1 weighted by atomic van der Waals volumes GETAWAY descriptors), and Ui (unsaturation index empirical descriptors). The activity data, pIC<sub>50</sub> (IC<sub>50</sub> in  $\text{mol L}^{-1}$ ), were obtained from the literature [6,7] and then used for subsequent QSAR analyses as the response variables. Linear (multiple

Table 2

Descriptor correlation matrix and collinearity indicators.

	T(N...N)	MATS6e	Mor20 m	E2 m	R1v+	Ui	Tolerance	VIF
T(N...N)	1	0.004	0.0041	0.0202	0.0083	0.0024	0.358	2.795
MATS6e		1	0.1118	0.2745	0.0128	0.3677	0.944	1.059
Mor20 m			1	0.0161	0.0019	0.0629	0.439	2.279
E2 m				1	0.0119	0.2444	0.811	1.233
R1v+					1	0.1563	0.612	1.634
Ui						1	0.558	1.792

linear regression, MLR) and nonlinear (support vector machines, SVM) [13] regression methods were used to correlate dependent and independent variables. The QSAR models were validated through leave-one-out and external validation (for a test set). The models were statistically evaluated by the squared correlation

coefficient of the experimental *versus* fitted and predicted  $\text{pIC}_{50}$  values ( $R^2$ ), and the standard deviations in calibration and external validation (SD).

## 2.2. Docking studies

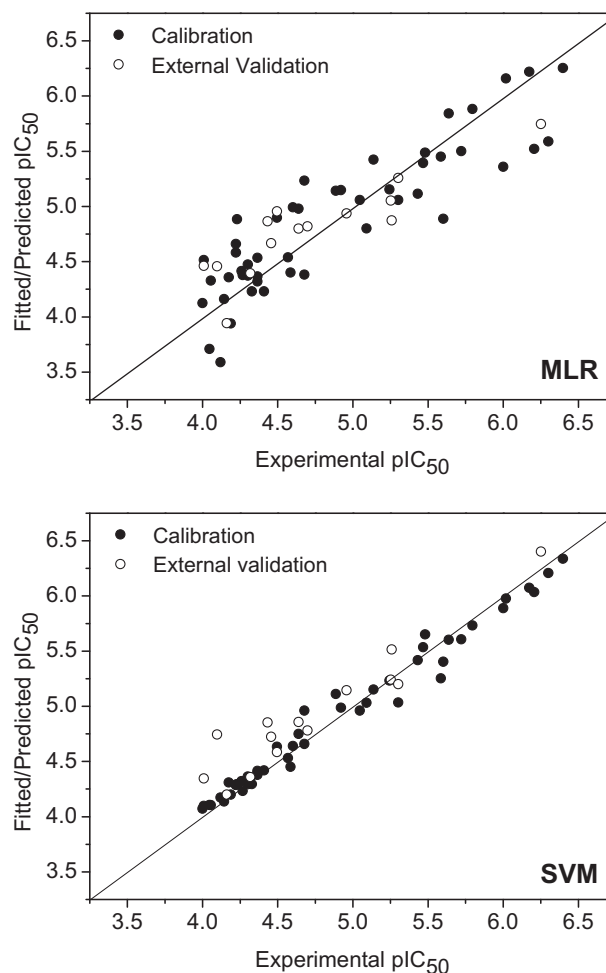
The amino acid primary sequence of *L. donovani*  $\alpha,\beta$  tubulin (*Ld-Tub*) was retrieved from the Swiss-Model Data Bank [23–26]. The search for sequences similar to *Ld-Tub* within the Protein Data Bank (PDB) [27] was performed with the Basic Local Alignment Search Tool program (BLAST) [28]. The search for the best template for modeling was carried out by choosing structures possessing a high degree of sequence similarity with *Ld-Tub*. The crystal structural coordinates of bovine brain tubulin at 3.7 Å resolution (PDB code: 1tub) [29] was used as template structure to build a three-dimensional model of *Ld-Tub*. Sequence alignment was performed using ClustalW [30] and then manually adjusted. The alignment was submitted to the Swiss-Model server and the three-dimensional model was built and energy minimized. Validation of the first generated model was done with the Procheck program available in the Biotech Validation Suite for Protein Structure [31,32]. The superimposition of the template backbone and target enzymes was feasible using the Swiss-Pdb Viewer 3.7 program [33].

The docking studies were carried out using the Molegro Virtual Docker (MVD), [34–39] a program for predicting the most likely conformation of how a ligand will bind to a macromolecule. The

**Table 3**

Experimental, fitted and predicted  $\text{pIC}_{50}$  for the antileishmanial compounds used in the training and test sets.

	Exp.	MLR	Residuals	SVMR	Residuals
<i>Training set</i>					
2	6.018	6.160	−0.142	5.977	0.041
3	5.796	5.883	−0.087	5.732	0.064
4	6.398	6.254	0.144	6.338	0.060
5	5.721	5.501	0.220	5.608	0.113
6	5.481	5.488	−0.007	5.652	−0.171
7	5.137	5.424	−0.287	5.152	−0.015
11	4.229	4.885	−0.656	4.294	−0.065
12	4.008	4.515	−0.507	4.098	−0.090
13	4.055	4.329	−0.274	4.105	−0.050
14	6.207	5.522	0.685	6.035	0.172
15	6.000	5.359	0.641	5.890	0.110
16	4.000	4.126	−0.126	4.073	−0.073
17	5.468	5.393	0.075	5.535	−0.067
18	4.602	4.992	−0.390	4.641	−0.039
19	6.174	6.220	−0.046	6.074	0.100
20	4.187	3.941	0.246	4.200	−0.013
21	5.301	5.058	0.243	5.035	0.266
22	5.432	5.115	0.317	5.419	0.013
23	4.222	4.582	−0.360	4.288	−0.066
24	4.677	5.233	−0.556	4.962	−0.285
25	4.568	4.540	0.028	4.531	0.037
26	4.267	4.379	−0.112	4.232	0.035
27	4.259	4.416	−0.157	4.322	−0.063
28	4.301	4.372	−0.071	4.293	0.008
29	4.366	4.535	−0.169	4.415	−0.049
30	4.301	4.475	−0.174	4.364	−0.063
31	5.585	5.450	0.135	5.253	0.332
33	4.495	4.898	−0.403	4.634	−0.139
34	4.886	5.143	−0.257	5.111	−0.225
37	4.366	4.319	0.047	4.396	−0.030
38	4.585	4.401	0.184	4.452	0.133
39	4.328	4.230	0.098	4.296	0.032
40	4.366	4.367	−0.001	4.378	−0.012
41	5.244	5.155	0.089	5.234	0.010
42	4.174	4.358	−0.184	4.310	−0.136
45	4.046	3.709	0.337	4.107	−0.061
47	4.119	3.589	0.530	4.173	−0.054
48	5.091	4.801	0.290	5.033	0.058
49	4.638	4.980	−0.342	4.751	−0.113
51	5.602	4.888	0.714	5.404	0.198
53	5.046	5.059	−0.013	4.962	0.084
54	4.921	5.149	−0.228	4.988	−0.067
55	4.143	4.162	−0.019	4.138	0.005
56	4.677	4.382	0.295	4.659	0.018
57	6.301	5.589	0.712	6.207	0.094
58	5.638	5.843	−0.205	5.603	0.035
61	4.409	4.231	0.178	4.419	−0.010
62	4.222	4.660	−0.438	4.289	−0.067
<i>Test set</i>					
1	6.252	5.747	0.505	6.403	−0.151
8	4.456	4.667	−0.211	4.723	−0.267
9	4.009	4.462	−0.453	4.346	−0.337
10	4.638	4.799	−0.161	4.857	−0.219
32	5.252	5.053	0.199	5.242	0.010
35	5.259	4.875	0.384	5.516	−0.257
36	4.318	4.396	−0.078	4.359	−0.041
43	4.161	3.943	0.218	4.200	−0.039
44	4.699	4.819	−0.120	4.781	−0.082
46	4.495	4.955	−0.460	4.586	−0.091
50	5.301	5.258	0.043	5.201	0.100
52	4.958	4.938	0.020	5.147	−0.189
59	4.097	4.458	−0.361	4.744	−0.647
60	4.432	4.865	−0.433	4.853	−0.421



**Fig. 1.** Plots of experimental *versus* fitted/predicted  $\text{pIC}_{50}$  using the MLR and SVM-based QSAR models.

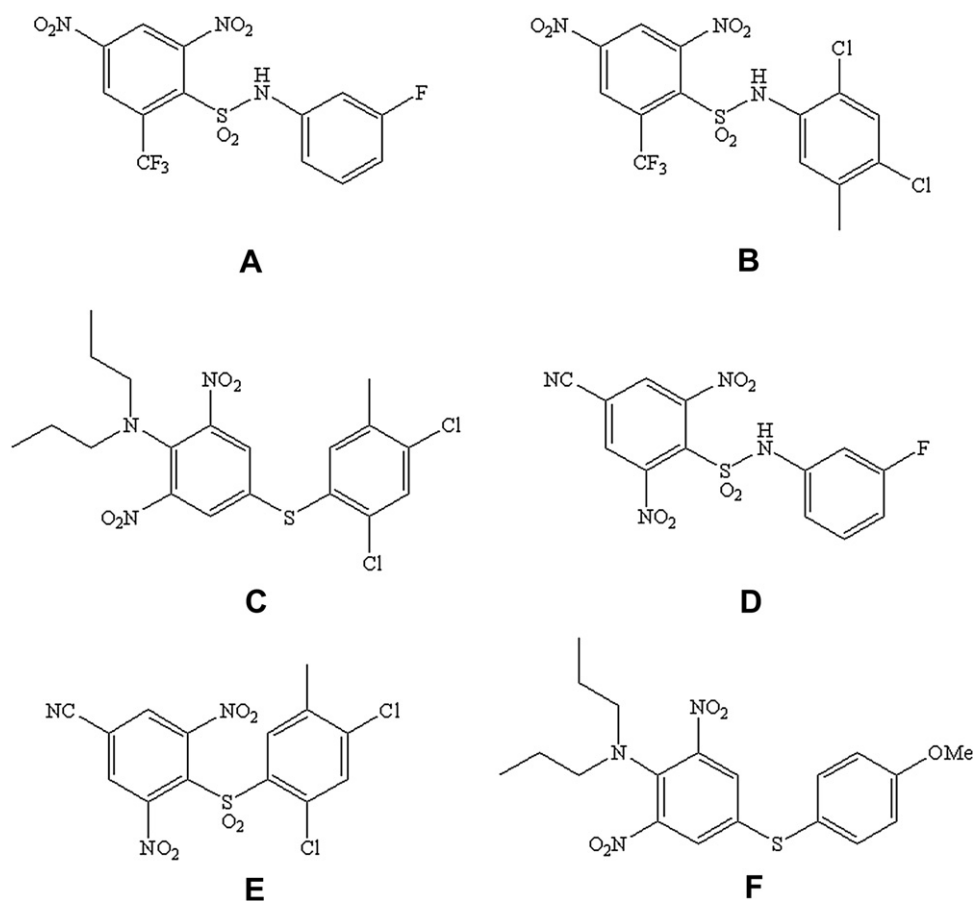


Fig. 2. Proposed compounds using the QSAR approach.

MolDock scoring function ( $E_{\text{score}}$ ) used by MVD program, is defined by the following energy terms:

$$E_{\text{score}} = E_{\text{inter}} + E_{\text{intra}} \quad (1)$$

where  $E_{\text{inter}}$  is the ligand–protein interaction energy and  $E_{\text{intra}}$  is the internal energy of the ligand. Ligands were docked into the *Ld*-Tub. The active site exploited in docking studies was defined through the calculated cavity. The interaction modes of each ligand with the *Ld*-Tub active site were determined as the highest protein–ligand complex energy score used during docking.

### 3. Results and discussion

#### 3.1. QSAR modeling and proposed structures

Dragon [21] descriptors were generated for the series of 62 compounds obtained from the literature [15,16], which was split into training (48 compounds) and test (14 compounds) set compounds with activity against *L. donovani*. Only six representative descriptors were selected through genetic algorithm from the pool of 1497 original descriptors, and their non-collinearity was tested through Variance Inflation Factor (VIF) and Tolerance, where:

$$\text{VIF} = \frac{1}{1 - R^2} \quad (2)$$

$$\text{Tolerance} = \frac{1}{\text{VIF}} \quad (3)$$

In practice, when  $\text{VIF} > 5$  or higher, or if the tolerance remains under the value 0.20, then this would indicate that there exists

multicollinearity among the descriptors. Table 2 shows that there is no significant correlation between the selected descriptors; then, there is no multicollinearity problem on the selected subset of molecular descriptors.

The selected descriptors are related to N...N distances, atomic electronegativities, atomic masses and volumes, and unsaturation index; therefore, it is supposed that steric hindrance and dipolar interactions, in addition to the presence of unsaturation, is ruling the enzyme–substrate interaction mode and, consequently, the bioactivities. Such descriptors were subsequently regressed against the bioactivities of the training set compounds by means of multiple linear regression (MLR). The following MLR equation was obtained, and the signals give insight about the positive or negative effect of the descriptors on the biological activities:

$$\begin{aligned} \text{pIC}_{50} = & -3.05891 (\pm 0.798) - 0.02769 (\pm 0.007) \times T(N...N) \\ & + 1.80863 (\pm 0.352) \times \text{MATS6e} - 0.61676 (\pm 0.198) \\ & \times \text{Mor20m} - 2.31161 (\pm 0.508) \times \text{E2m} \\ & + 19.00113 (\pm 4.390) \times \text{R1v} + 2.41136 (\pm 0.283) \times \text{Ui} \end{aligned}$$

The above MLR model gave a reasonable calibration with  $R^2$  of 0.792 (SD = 0.329). The model predictability was tested through leave-one-out and external validations, giving  $R^2_{\text{LOO}}$  of 0.724 (SD<sub>LOO</sub> = 0.408) and  $R^2_{\text{test}}$  of 0.777 (SD<sub>test</sub> = 0.316), respectively. The fitted and predicted results are depicted in Table 3. Despite the satisfactory results obtained by the MLR model, accounting for nonlinearities has provided advantages in terms of prediction performance [40,41]. Support vector machines (SVM) have appeared as a powerful regression tool to take nonlinearity into



Alignment between  $\alpha$ -tubulin\_ *L. donovani* and 1TUB

Ldonovani 1TUB	MREAICIHIGQAGCQVGNACWELFCLEHGIQPDGSMPSDKCIGVEDDAFNTFFSETGAGK MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGGGDDSFNTFFSETGAGK ***.*:***:***:*****:*****:*****:*****:*** ** *:*****:*****
Ldonovani 1TUB	HVPRCIFLDLEPTVVDEVRTGTYRQLFNPEQLVSGKEDAANNYARGHYTIGKEIVDLALD HVPRAVFVDLEPTVIDEVRTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIIDLVLD ***.*:***:*****:*****:*****:*****:*****:*****:***.*
Ldonovani 1TUB	RIRKLADNCTGLQGFMVFHAVGGGTGSGGLGALLLERLSVDYGKSKLGYTVPSQPQVSTA RIRKLADQCTGLQGFVSFHSFGGTGSGFTSLLMERLSVDYGKSKLEFSIYPAPQVSTA *****:***** ***:.*:*****: ***:*****:*****:*****:***:*****
Ldonovani 1TUB	VVEPYNCVLSTHSLEHTDVATMLDNEAIYDLTRRLDIERPSYTNVNLIGQVVSSTA VVEPYNSILTHTTLEHSDCAFMDNEAIYDICRRNLDIERPTYTNLNLIGQIVSSITA *****.*:***: ***: * *:*****: ***.*****:***:*****:***:***
Ldonovani 1TUB	SLRFDGALNVDLTFQTNLVPYPRIFHVLTSYAPVVSAAEKAYHEQLSVADITNSVFEPAG SLRFDGALNVDLTFQTNLVPYPRGHFPLATYAPVISAEEKAYHEQLSVAEITNACFEPAN *****:***** ** *:***:*****:*****:***: ***,
Ldonovani 1TUB	MLTKCDPRHGKYMCCLLMYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKCGINYQPP QMVKCDPRHGKYMCCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYEP :.*:*****:***:*****:*****:*****:*****:*****:***:***
Ldonovani 1TUB	TVVPGGDLAKVQRAVCMIANSTAEVFAIDHKFDLMYSKRAVFWHWYVGEEMEEGEFSE VVPGGDLAKVQRAVCMISNTTAEAWARLDHKFDLMYAKRAVFWHWYVGEEMEEGEFSE *****:*****:***:*****:***:*****:*****:*****:*****:*****
Ldonovani 1TUB	AREDLAAVEKDYEEVGAESADDMGEEDVEEY AREDMAALEKDYEEVGVDV----- ***.*:***:*****.*:.*.

Alignment between  $\beta$ -tubulin\_ *L. donovani* and 1TUB

Ldonovani 1TUB	MREIVSCQAGQCGNQIGSKFWEVIADHEGVDP TGSYQGDSDLQ-----ESAGGRYV MREIVHIQAGQCGNQIGAKFWEVISDEHGDIDPTGSYHGSDQLERINVYNEAAGNKYV ***** *****:*****:***:*****:***** *:***:***
Ldonovani 1TUB	PRAVLMDLEPGTMDSVRAGPYGQLFRPDNFI FGQSGAGNNWAKGHYTEGAELIDSVLDVC PRAILVDLEPGTMDSVRSGPFGQIFRPDNFVFGQSGAGNNWAKGHYTEGAELVDSVLDV ***.*:*****:***:***:*****:*****:*****:*****:*****
Ldonovani 1TUB	RKEAESCDCLQGQFQLSHSLGGGTGSGMGTLLISKLREEYPDRIMMTFSVIPSPRVSDTVV RKESESCDCLQGQFQLTHSLGGGTGSGMGTLLISKIREEYPDRIMNTFSVVPSPKVS DTVV ***:*****:*****:*****:*****:***** *****:***:*****
Ldonovani 1TUB	EPYNTTSLSVHQLVENSDESMCIDNEALYDICFRTLKLTPTFGDLNHLVAAMVSGVTCCCL EPYNATLSVHQLVENTDETYCIDNEALYDICFRTLKLTPTTYGDLNHLVSATMSGVTTCL ***:*****:***: *****:*****:*****:*****:***:***** **
Ldonovani 1TUB	RFPQQLNSDLRKLAVNLVFPRLHFFMMGFAPLTSRGSHQYPGLSVAELTQQMFDKNNMM RFPQQLNADLRKLAVNMVFPRLHFFMPGFAPLTSRGSQYRALTVPELTQQMFDKNNMM *****:*****:*****:***** *****:***:***:*****:*****
Ldonovani 1TUB	QAADPRHGRYLTASALFRGRMSTKEVDEQMLNVQKNSSYFIEWIPNNIKSSICDIPPKG AACDPRHGRYLTVAAVFRGRMSMKEVDEQMLNVQKNSSYFVEWI PNNVKTAVCDIPPRG *.*:*****.*:***** *****:*****:*****:***:***:***
Ldonovani 1TUB	LKMSVTFIGNNTCIQEMFRRVGEQFTGMFRRKACLHWYTGEGMDE----- LKMSATFIGNSTAIQELFKRISEQFTAMFRRKAF LHWYTGEGMDEMEFTEAESNMNDLVS ***.*:*****.*:***:***:***:***** ***** *****
Ldonovani 1TUB	----- EYQQYQD

**Fig. 3.** Clustal alignment of  $\alpha$ ,  $\beta$ -tubulin *L. donovani* amino acid sequences with the sequence of the crystallographic structure of 1TUB, available in PDB. Regions where the secondary structure was predicted:  $\alpha$ -helix: H1 (Gly10–Cys25), H2 (Pro72–Arg79), H3 (Ile110–Lys124), H4 (Ala147–Ser158), H5 (Val181–Leu194), H6 (Ser223–Val234), H7 (Val324–Ile332), H8 (Ala383–His393), H9 (Phe418–Asp431),  $\beta$ -sheet: B1 (Ile7–Gly10), B2 (Gly134–His139), B3 (Ser165–Tr172), B3 (Asp199–Leu201), B4 (Val268–Ser271), B5 (Leu302–Tyr319), B6 (Val375–Asn380). (\*) amino acid identity; (:) amino acid strongly similar; (.) amino acid weakly similar.

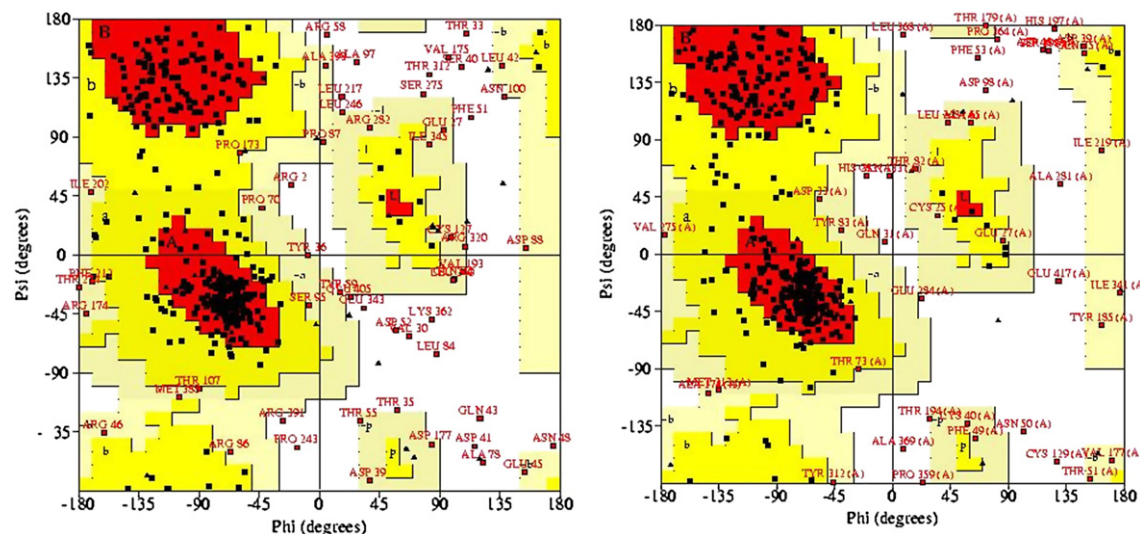


Fig. 4. Ramachandran plot for the *Ld-Tub*:  $\alpha$ -tubulin is on the left and  $\beta$ -tubulin is on the right. Glycine residues are separately identified by triangles.

account and thus it was applied in this study for calibration and prediction. The calibration results were greatly improved when using SVM ( $R^2 = 0.979$ ,  $SD = 0.113$ ), as well as the prediction ability for the external set of compounds ( $R_{LOO}^2 = 0.931$ ,  $SD_{LOO} = 0.169$ ;  $R_{test}^2 = 0.901$ ,  $SD_{test} = 0.204$ ), as illustrated in Fig. 1. Thus, both models, especially the SVM, may be reliably used to predict the biological activities of novel congeners of the diaryl sulfides and sulfonamides evaluated here.

The strategy to obtain novel active structures from the two compound classes consisted in combining substructures of two highly active compounds of each class, similarly to that previously reported for other series of compounds [42]. The selected compounds having mixed substructures were **4**, **19**, **57** and **58**, giving the six proposed compounds of Fig. 2. Four (**A**, **B**, **D** and **E**) exhibited predicted  $pIC_{50}$  higher than 5 (within the range of the most active training compounds) when using both MLR and SVM regression parameters. Therefore, the proposed structures were submitted to docking evaluation to confirm their high affinity to the active site of *Ld-Tub*.

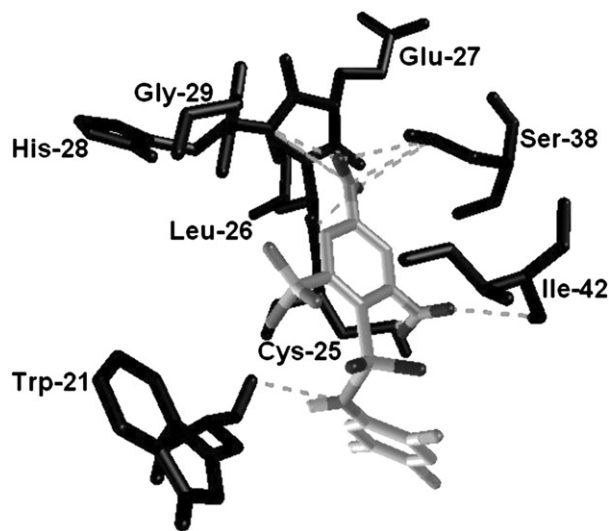


Fig. 5. Structure of compound **B** docked into the *Ld-Tub* active site. The residues shown are involved in electrostatic interaction (Asp-224), hydrogen bonding (Leu-228, His-227 and Thr-274) and cation  $\pi$ - $\pi$  interaction (Arg-286) with the proposed compound.

### 3.2. Tubulin homology-modeling

As a first step, the alignment of the *L. donovani*  $\alpha$ - $\beta$  tubulin primary structure with the chosen template sequences of 1TUB was performed. The alignment using  $\alpha$ -tubulin showed 85.0% and  $\beta$ -tubulin showed 79.0% of congruence (Fig. 3). The proposed models were validated using the program PROCHECK. The Ramachandran plot of the alpha and beta subunits *Ld-Tub* (Fig. 4) satisfied the tests with: i) 74.4% and 66.5% of the residues in the most favored regions, respectively, ii), 15.7% (alpha) and 20.0% (beta) of the residues are in additional allowed regions, iii) 5.7% and 7.0% of the residues are in generously allowed regions, iv) 4.2% and 6.5% of the residues in disallowed regions. The dimmer was modeled based on the 1TUB crystal structure. The compatibility between the active site of the *Ld-Tub* homology model and the 3D-structures of the template used (1TUB) was checked by superimposing the backbone of the two enzymes. The superimposition gave root-mean-square deviation (*rmsd*) values equal to 0.26 Å ( $\alpha$ -tubulin) and 0.34 Å ( $\beta$ -tubulin). GDP and GTP cofactor were added to the model and the complex was slowly relaxed to clean the starting structure.

#### 3.2.1. Inhibitors docking orientation into *Ld-Tub*

The docked binding mode is used to establish a link between the MolDock scoring function, structural properties of these compounds [43–45] and their biological activity against the *Ld-Tub* [46]. Evaluation of the docking results was based on protein–ligand complementarity considering steric and electrostatic properties. The site at which oryzalin analogs bind has yet to be characterized [47].

Table 4  
Interaction energy values ( $\text{kcal mol}^{-1}$ ) between *Ld-Tub* and some inhibitors of Fig. 2.

Compound	Interaction ( $\text{kcal mol}^{-1}$ )	$pIC_{50}$
<b>A</b>	−158.18	5.175 (pred.)
<b>B</b>	−169.20	5.054 (pred.)
<b>C</b>	−134.24	4.479 (pred.)
<b>D</b>	−136.51	5.077 (pred.)
<b>E</b>	−136.15	5.882 (pred.)
<b>F</b>	−119.91	4.330 (pred.)
<b>20</b> (Oryzalin)	−119.30	4.187 (exp.)
<b>4</b>	−136.80	6.398 (exp.)
<b>19</b>	−131.95	6.174 (exp.)
<b>57</b>	−142.30	6.301 (exp.)
<b>58</b>	−140.19	5.638 (exp.)

**Table 5**  
ADME related parameters<sup>a</sup> obtained from the Molinspiration program.

Compound	TPSA	log <i>P</i>	MW	<i>n</i> <sub>ON</sub>	<i>n</i> <sub>OHNH</sub>	Volume
<b>20</b> (Oryzalin)	155.05	3.38	346.36	10	2	286.54
<b>4</b>	91.65	6.73	413.16	6	0	278.62
<b>19</b>	124.67	3.56	331.31	8	0	262.65
<b>57</b>	91.65	7.19	427.19	6	0	295.18
<b>58</b>	125.79	5.72	459.18	8	0	308.48
<b>A</b>	137.82	4.39	409.27	9	1	282.18
<b>B</b>	137.82	6.44	474.2	9	1	320.88
<b>C</b>	94.89	8.66	458.37	7	0	376.99
<b>D</b>	161.61	2.51	366.29	10	1	267.74
<b>E</b>	149.58	4.08	416.20	9	0	294.04
<b>F</b>	104.12	6.53	405.48	8	0	358.91

<sup>a</sup> TPSA: total polar surface area; log *P*: logarithm of the octanol/water partition coefficient; MW: molecular weight; *n*<sub>ON</sub>: number of hydrogen bond acceptors; *n*<sub>OHNH</sub>: number of hydrogen bond donors.

However, studies on *Toxoplasma gondii* indicate that low-level resistance to these compounds is often associated with point mutations in  $\alpha$ -tubulin [48]. Thus, the potential binding sites ( $\alpha$ -tubulin) were calculated and a cavity of 51.7 Å<sup>3</sup> (surface = 170.2 Å<sup>2</sup>) was observed close to Ala19, Cys20–Gly29, Ser38–Gly43, His61, Pro63, Arg84, Gln85 and Phe87 (radius = 10 Å). By analyzing the hydrogen bond formed between the eleven compounds and the Ld-Tub active site we observed: (i) only compound **20** (Oryzalin) exhibited hydrogen bonds with Lys-40; (ii) compounds **A**, **B**, **D**, **19** and **20** interacted with Glu27 and His28; (iii) compounds **A**, **B** and **D** interacted with Trp-21; (iv) compounds **A** and **B** interacted with Leu26; (v) all compounds except **C** interacted with Cys25; (vi) compounds **A**, **B**, **D**, **E**, **57** and **58** interacted with Ile42; (vii) compounds **A**, **B**, **D**, **E**, **19** and **20**, interacted with Ser38; (viii) compounds **A** and **E** interacted with Gly29. Fig. 5 shows compound **B** inside the Ld-Tub active site and the interacting amino acid residues. Compounds **A** and **B** experience a lower intermolecular energy (more stable complex) because the distance between the two aromatic rings is larger. When the nitrogen atom is added close to the sulfur atom, the compounds **A** and **B** interact with Trp-21 through the hydrogen bond, stabilizing the ligand–protein complex. Compound **D** exhibits similar characteristics to the compounds **A** and **B**, but its lower intermolecular energy is due to different conformation adopted for compound **D** when the substituent group changes from CF<sub>3</sub> (compounds **A** and **B**) to NO<sub>2</sub>. The neat results of the above interactions are given in terms of affinity (interaction energy) toward the active site of Ld-Tub, as depicted in

Table 4. According to these values, compounds **A** and **B** presented an estimated affinity to the Ld-Tub active site higher than the standard compound Oryzalin and the most promising reference compounds of the training set. Therefore, druglikeness for **A–F** was tested and compared to the reference compounds by performing an ADME–Tox evaluation.

### 3.3. ADME–Tox evaluation

A druglike molecule has properties like optimal solubility in both water and fat (because an orally administered drug has to go through the intestinal lining, carried in aqueous blood and penetrate the lipid cellular membrane to reach the inside of a cell), transportation in aqueous media like blood and intracellular fluid, low molecular weight (because diffusion is directly affected), and low toxicity. The logarithm of the octanol/water partition coefficient (log *P*) is usually used to estimate solubility, as well as the number of hydrogen bond donors, whilst the total polar surface area (TPSA) of a molecule is directly related to its oral bioavailability. One test for druglikeness is Lipinski's Rule of Five, [49] a series of rules imposing limitations on log *P* (the logarithm of octanol/water partition coefficient), molecular weight, and the number of hydrogen bond acceptors and donors, known as “rule of five”. The rule states that most druglike molecules have log *P* ≤ 5, molecular weight ≤ 500, number of hydrogen bond acceptors ≤ 10, and number of hydrogen bond donors ≤ 5. Molecules violating more than one of these rules may have problems with bioavailability. TPSA values and the rule of five were obtained for the proposed and reference compounds using the Molinspiration program [50], as depicted in Table 5. The proposed compounds **A**, **D** and **E** did not violate any rule, such as compounds **19** and **20**. Compound **B**, a promising inhibitor of Lb-Tub according to the docking studies, is supposed to have low aqueous solubility, due to its large log *P* value. Druglikeness of the most promising compound **A**, highly active according to QSAR and docking studies, and suitable according to the ADME evaluation, was then scrutinized using the ADME and Tox-Boxes of the PharmaAlgorithm program [51]. The ADME–Tox properties of compound **A** were comparable to the reference compounds (Table 6), although the smaller predicted volume of distribution, but improved toxicity estimations, particularly the carcinogenicity (Ames) test. Furthermore, its interaction with Lb-Tub was more attractive, revealing its potential for driving synthesis as a potent antileishmanial compound.

**Table 6**  
ADME–Tox parameters for reference and proposed (**A–F**) compounds.

Parameter	<b>20</b> (Oryzalin)	<b>4</b>	<b>19</b>	<b>57</b>	<b>58</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
%F (oral) > 30%	0.716	0.884	0.775	0.884	0.716	0.775	0.775	0.465	0.775	0.775	0.465
%F (oral) > 70%	0.269	0.432	0.269	0.432	0.168	0.269	0.168	0.099	0.269	0.269	0.099
Absorption rate (min <sup>−1</sup> )	0.095	0.100	0.099	0.100	0.100	0.097	0.100	0.100	0.095	0.098	0.100
Vol. of distribution (L kg <sup>−1</sup> )	2.02	4.30	2.10	4.63	1.98	1.05	1.42	6.90	0.84	1.88	5.63
Lo gS (stomach)	−4.51	−6.32	−4.60	−6.44	−6.03	−5.41	−6.24	−6.81	−4.55	−5.09	−5.37
Log S (duodenum)	−4.51	−6.32	−4.60	−6.44	−6.03	−5.41	−6.24	−6.81	−4.55	−5.09	−5.37
Log S (jejunum and ileum)	−4.51	−6.32	−4.60	−6.44	−6.03	−5.40	−6.23	−6.81	−4.55	−5.09	−5.37
Log S (blood)	−4.50	−6.32	−4.60	−6.44	−5.99	−5.36	−6.20	−6.81	−4.51	−5.09	−5.37
Log S (colon)	−4.50	−6.32	−4.60	−6.44	−5.89	−5.26	−6.09	−6.81	−4.41	−5.09	−5.37
Log S <sub>w</sub>	−4.51	−6.32	−4.60	−6.44	−6.03	−5.41	−6.24	−6.81	−4.55	−5.09	−5.37
Toxicity (prob. Ames)	0.822	0.364	0.951	0.267	0.111	0.007	0.005	0.853	0.099	0.374	0.965
Prob. effect on blood	0.85	0.60	0.35	0.63	0.83	0.54	0.65	0.92	0.62	0.72	0.70
Prob. cardiovascular effect	0.92	0.92	0.91	0.90	0.57	0.72	0.99	0.95	0.72	0.95	0.84
Prob. gastrointestinal effect	0.91	0.93	0.79	0.97	0.90	0.89	0.97	1.00	0.97	0.99	0.83
Prob. effect on kidney	0.17	0.75	0.49	0.74	0.55	0.25	0.53	0.78	0.36	0.38	0.69
Prob. effect on liver	0.16	0.47	0.68	0.49	0.49	0.27	0.57	0.54	0.67	0.43	0.44
Prob. effect on lungs	0.99	0.62	0.46	0.78	0.99	0.45	0.39	0.98	0.90	0.38	0.98
Oral LD <sub>50</sub> (mouse, mg kg <sup>−1</sup> )	1600	780	1100	810	1800	880	650	1800	1300	1100	1600
Oral LD <sub>50</sub> (rat, mg kg <sup>−1</sup> )	1400	600	1400	620	1200	620	350	1600	1200	1000	1900

#### 4. Conclusion

The QSAR method used for modeling antileishmanial compounds showed to be robust, with reliable prediction ability. Given the synergistic effect of some substituents contained in the training set, it was possible to predict novel compounds (miscellaneous of substructures of both classes in training set) with four proposal structures (A, B, D and E) with exhibited predicted pIC<sub>50</sub> higher than 5 (within the range of the most active training compounds). These results were corroborated by docking studies, which revealed the same trend found through the ligand-based approach. Overall, the joint use of MIA-QSAR method and docking studies allowed us to propose at least four highly potent antileishmanial compounds.

#### Acknowledgements

Authors are grateful to FAPEMIG and CNPq for the financial support of this research, as well as to CNPq for the fellowships.

#### References

- [1] WHO, <http://www.who.int/leishmaniasis/burden/en>.
- [2] Hope for tropical disease vaccine, BBC News April 23 (2006).
- [3] J. Soto, J.T. Toledo, *Lancet Infect. Dis.* 7 (2007) 7–14.
- [4] S. Sundar, J. Chakravarty, V.K. Rai, N. Agrawal, S.P. Singh, V. Chauhan, H.W. Murray, *Clin. Infect. Dis.* 45 (2007) 556–563.
- [5] M. Mueller, K. Ritmeijer, M. Balasegaram, Y. Koummuki, M.R. Santana, R. Davidson, *Trans. R. Soc. Trop. Med. Hyg.* 101 (2007) 19–28.
- [6] E.F.F. da Cunha, E.F. Barbosa, A.A. Oliveira, T.C. Ramalho, J. Biomol. Struct. Dyn. 27 (2010) 619–625.
- [7] S.L. Croft, *Mem. Inst. Oswaldo Cruz.* 94 (1999) 215–220.
- [8] E.S. Coimbra, R. Carvalhaes, R.M. Grazul, P.A. Machado, M.V.N. De Souza, A.D. da Silva, *Chem. Biol. Drug Design* 75 (2010) 628–631.
- [9] V.P. Pandey, S.S. Bisht, M. Mishra, A. Kumar, M.I. Siddiqi, A. Verma, M. Mittal, S.A. Sane, S. Gupta, R.P. Tripathi, *Eur. J. Med. Chem.* 45 (2010) 2381–2388.
- [10] J.T. Pierson, A. Dumetre, S. Hutter, F. Delmas, M. Laget, J.P. Finet, N. Azas, S. Combes, *Eur. J. Med. Chem.* 45 (2010) 864–869.
- [11] Z. Nazarian, S. Emami, S. Heydari, S.K. Ardestani, M. Nakhjiri, F. Poorrajab, A.A. Shafiee, A. Foroumadi, *Eur. J. Med. Chem.* 45 (2010) 1424–1429.
- [12] Z. Garkani-Nejad, B. Ahmadi-Roudi, *Eur. J. Med. Chem.* 45 (2010) 719–726.
- [13] N. Sunduru, S. Palne, P.M.S. Chauhan, S. Gupta, *Eur. J. Med. Chem.* 44 (2009) 2473–2481.
- [14] K.M.G. Oliveira, Y. Takahata, *QSAR & Comb. Sci.* 27 (2008) 1020–1027.
- [15] D.A. Delfin, R.E. Morgan, X. Zhu, K.A. Werbovetz, *Bioorg. Med. Chem.* 17 (2009) 820–830.
- [16] D.A. Delfin, A.K. Bhattacharjee, A.J. Yakovich, K.A. Werbovetz, *J. Med. Chem.* 49 (2006) 4196–4208.
- [17] J.R. Pinheiro, M. Bitencourt, E.F.F. da Cunha, T.C. Ramalho, M.P. Freitas, *Bioorg. Med. Chem.* 16 (2008) 1683–1691.
- [18] E.F.F. da Cunha, W. Sippl, M.G. Albuquerque, R.B. de Alencastro, *Eur. J. Med. Chem.* 44 (2009) 4344–4352.
- [19] HyperChem version 7.0. Hypercube, Inc., Gainesville, 2007.
- [20] D.C. Young, *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*. John Wiley & Sons, New York, 2001.
- [21] R. Todeschini, V. Consonni, M. Pavan, Dragon Software, Milano, 2002.
- [22] H. Li, Y. Liang, Q. Xu, *Chemom. Intell. Lab. Sys.* 95 (2009) 188–194.
- [23] M.C. Peitsch, *Biol. Technol.* 13 (1995) 658–664.
- [24] M.C. Peitsch, *Biochem. Soc. Trans.* 24 (1996) 274–283.
- [25] M.C. Peitsch, N. Guex, *Electrophoresis* 18 (1997) 2714–2721.
- [26] T. Schwede, J. Kopp, N. Guex, M.C. Peitsch, *Nucleic Acids Res.* 31 (2003) 3381–3394.
- [27] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Res.* 28 (2000) 235–243.
- [28] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, *J. Mol. Biol.* 215 (1990) 403–410.
- [29] E. Nogales, S.G. Wolf, K.H. Downing, *Nature* 391 (1998) 199–205.
- [30] <http://www.ch.embnet.org/software/ClustalW.html>.
- [31] <http://biotech.ebi.ac.uk:8400/>.
- [32] G. Vriend, *J. Mol. Graph.* 8 (1990) 52–63.
- [33] N. Guex, A. Diemand, M.C. Peitsch, T. Schwede, GlaxoSmithKline R&D & the Swiss Institute of Bioinformatics.
- [34] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, Berlin, 1992.
- [35] Z. Michalewicz, D.B. Fogel, *How to Solve It: Modern Heuristics*. Springer-Verlag, Berlin, 2000.
- [36] R. Thomsen, M.H. Christensen, *J. Med. Chem.* 49 (2006) 3315–3324.
- [37] E.F.F. da Cunha, R.C.A. Martins, M.G. Albuquerque, *J. Mol. Mod.* 10 (2004) 297–304.
- [38] E.F.F. da Cunha, E.F. Barbosa, A.A. Oliveira, T.C. Ramalho, J. Biom. Struct. Dyn. 27 (2010) 619–625.
- [39] T.C. Ramalho, M.V.J. Rocha, E.F.F. da Cunha, M.P. Freitas, *Expert Opin. Ther. Pat.* 19 (2009) 1193–1228.
- [40] R.A. Cormanich, M. Goodarzi, M.P. Freitas, *Chem. Biol. Drug. Des.* 73 (2009) 244–256.
- [41] M. Goodarzi, M.P. Freitas, R. Jensen, *J. Chem. Inf. Model.* 49 (2009) 824–830.
- [42] M.P. Freitas, *Org. Biomol. Chem.* 4 (2006) 1154–1161.
- [43] G. Vriend, *Protein Eng.* 4 (1990) 221–223.
- [44] R.W.W. Hooft, G. Vriend, C. Sander, E.E. Abola, *Nature* 381 (1996) 272–284.
- [45] R. Rodriguez, G. Chinea, N. Lopez, T. Pons, G. Vriend, *CABIOS* 14 (1998) 523–531.
- [46] D. Josa, E.F.F. da Cunha, T.C. Ramalho, T.C.S. Souza, M.S. Caetano, J. Biomol. Struct. Dyn. 25 (2008) 373–378.
- [47] B.J. Fennell, J.A. Naughton, J. Barlow, G. Brennan, I. Fairweather, E. Hoey, N. McFerran, A. Trudgett, A. Bell, *Expert Opin. Drug Discov.* 3 (2008) 501–518.
- [48] N.S. Morrisette, A. Mitra, D. Sept, L.D. Sibley, *Mol. Biol. Cell.* 15 (2004) 1960–1968.
- [49] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, *Adv. Drug Deliv. Rev.* 23 (1997) 3–9.
- [50] Molinspiration Cheminformatics, Bratislava, Slovak Republic, <http://www.molinspiration.com>.
- [51] Pharma Algorithms, Toronto, Canada, 2008.