



OPEN

Quinoxalines against *Leishmania amazonensis*: SAR study, proposition of a new derivative, QSAR prediction, synthesis, and biological evaluation

Anna Carolina Silva de Jesus Passaes¹, Juliana Arantes Dantas^{1,2}, Fernanda Landim Lopes¹, Diego Pereira Sangi¹, Magaly Girão Albuquerque³, Celso Vataru Nakamura⁴ & Julliane Yoneda¹✉

Neglected tropical diseases, such as leishmaniasis, lead to serious limitations to the affected societies. In this work, a structure–activity relationship (SAR) study was developed with a series of quinoxaline derivatives, active against the promastigote forms of *Leishmania amazonensis*. As a result, a new quinoxaline derivative was designed and synthesized. In addition, a quantitative structure–activity relationship (QSAR) model was obtained [$\text{pIC}_{50} = -1.51 - 0.96 (E_{\text{HOMO}}) + 0.02 (\text{PSA})$; $N = 17$, $R^2 = 0.980$, $R^2_{\text{Adj}} = 0.977$, $s = 0.103$, and $\text{LOO-cv-R}^2 (Q^2) = 0.971$]. The activity of the new synthesized compound was estimated ($\text{pIC}_{50} = 5.88$) and compared with the experimental result ($\text{pIC}_{50} = 5.70$), which allowed to evaluate the good predictive capacity of the model.

According to the World Health Organization (WHO), the neglected tropical diseases (NTDs) are a diverse group of 20 conditions that are mainly prevalent in the tropical and subtropical regions of the world, such as Latin America, Africa, and Asia, predominantly in developing countries. The NTDs are considered endemic in low-income populations, as they affect vulnerable people, who in most cases have limited access to clean and potable water, and poor hygiene and sanitation conditions¹.

Among the 20 NTDs, tuberculosis, Chagas disease, leprosy, malaria, dengue, schistosomiasis, and leishmaniasis are included in the Brazilian National Agenda of Priorities in Health Research². Although they have been present on our planet for thousands of years, they continue without being eradicated, and impose serious limitations on the affected societies, leading to a panorama of illness, suffering, disability, and death, with serious social, economic, and psychological consequences, affecting more than 1 billion people worldwide, according to the WHO data³.

In January 2021, WHO launched its new roadmap for tackling NTDs from 2021 to 2030. Targets include achieving prevention, control, elimination, and eradication of a diverse set of diseases until 2030. The goal includes a 90% decrease in the number of people in need of treatment, that at least one hundred countries eliminate at least one neglected disease present in their nation, and a 75% reduction in years of life lost due to disability caused by these diseases³.

The treatments available for the NTDs are very limited and insufficient, in addition to presenting a series of problems, such as low efficacy, high toxicity, and the emergence of resistant strains⁴.

Leishmaniasis are a set of diseases caused by protozoan of the *Leishmania* genus and the Trypanosomatidae family, and they are transmitted to humans by the bite of infected female phlebotomine sandflies. Leishmaniasis can occur in three different clinical forms: (i) visceral leishmaniasis (VL), which is generally fatal without treatment; (ii) cutaneous leishmaniasis (CL) that causes skin ulcers; and (iii) mucocutaneous leishmaniasis

¹Departamento de Química, Instituto de Ciência Exatas, Universidade Federal Fluminense, Volta Redonda 27213-145, Brazil. ²Programa de Pós-Graduação em Química, Universidade Federal de São Carlos, São Carlos 13565-905, Brazil. ³Programa de Pós-Graduação em Química, Departamento de Química Orgânica, Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-909, Brazil. ⁴Programa de Pós-Graduação em Ciências Farmacêuticas, Universidade Estadual de Maringá, Maringá 87020-900, Brazil. ✉email: jullianeyoneda@id.uff.br

(MCL), affecting nose, mouth, and throat. The WHO estimates that 700,000–1 million new cases occur each year worldwide⁵.

The antileishmanial treatment is performed, as a first choice, with pentavalent antimonial agents, such as sodium stibogluconate. Meanwhile, amphotericin B, pentamidine, miltefosine, and paromomycin have also been used with varying results. However, the current drugs have undesirable side effects and important toxicities. For example, amphotericin and pentavalent antimonial drugs cause high nephrotoxicity and cardiac effects, respectively, in addition to being administered intravenously, which sometimes hinders the patient's adherence to the treatment, resulting in therapeutic failures and favoring the parasite's resistance to the drugs^{5,6}.

Some N-heterocyclic nucleus are considered privileged scaffolds due to their broad spectrum of biological activity highlighted in the literature⁷. Privileged scaffolds generally exhibit physicochemical properties that allow a single class of molecules to provide potent and selective ligands for different biological targets⁸. The N-heterocycle quinoxaline (1,4-naphthyridine) (Fig. 1) is considered a privileged scaffold.

Quinoxaline derivatives represent a class of biologically active compounds, showing anti-inflammatory, anticancer, antibacterial, antimicrobial, antifungal, antiviral, and antileishmanial activities, among others^{8–12}. The quinoxaline heteroaromatic scaffold is found in more than 30 drugs available in the DrugBank (DB, <https://go.drugbank.com/>), including approved, nutraceuticals, investigational or experimental ones, such as brimonidine (DB00484), chlorsulfaquinoxaline (DB12921), erdafitinib (DB12147), rabeximod (DB05772), riboflavin (DB00140), and varenicline (DB01273).

On the search of new drug leads, there is a need for efficient and robust procedures that can be used to screen chemical databases against molecules with known activities. To this end, quantitative structure–activity relationships (QSAR) studies provide a mean for rationalizing the relationship between chemical structure and its biological action towards the development of new drug candidates¹³.

Cherkasov et al.¹⁴ have described several QSAR studies where computational and medicinal chemists worked together to discover novel molecules with unique biological activities.

In this context, this work aims to evaluate by an *in silico* approach which physicochemical properties of the quinoxaline derivatives (**2a–2i**, **3a–3i**, and **4a–4d**) (Table 1)¹¹ contribute to their *in vitro* inhibitory activity against the promastigote forms of *Leishmania amazonensis*, to propose and to synthesize a new potential antileishmanial agent, and to build a QSAR model able to predict its activity.

Methods

Computational chemistry

The three-dimensional (3D) structures of the quinoxaline derivatives (**2a–2i**, **3a–3i**, and **4a–4d**) (Table 1) were constructed using the Spartan'10 software (Wavefunction, Inc.)¹⁵. Each structure was submitted to a full geometry optimization step by a molecular mechanics model, using the Merck molecular force field (MMFF), available in the Spartan software. Then, each optimized structure was submitted to the default systematic conformational analysis at Spartan, using the same molecular mechanics force field. The lowest-energy conformer for each quinoxaline derivative was submitted to a full geometry optimization (energy minimization) step by a semi-empirical model, using the Austin Method 1 (AM1) Hamiltonian at Spartan. Then, each optimized conformer was submitted to a single-point energy calculation by a density functional theory (DFT) model, using the B3LYP hybrid DFT method at Spartan, considering the 6-311 + + G(*d,p*) basis set. For each energy minimized DFT structure, the following thirteen physicochemical properties were obtained: total energy (E_T , au), energy of the highest occupied molecular orbital (E_{HOMO} , eV), energy of the lowest unoccupied molecular orbital (E_{LUMO} , eV), HOMO–LUMO energy gap (GAP, eV), dipole moment (μ , Debye), base-10 logarithm of the partition coefficient (LogP), surface area (SA, Å²), molecular volume (MV, Å³), molecular weight (MW, amu), polarizability (P, 10^{–30} m³), number of hydrogen bond donors (HBD), number of hydrogen bond acceptors (HBA), and polar surface area (PSA, Å²).

A linear cross-correlation matrix was constructed with the calculated thirteen physicochemical properties as a criterion to exclude at least one from the two highly correlated pair of properties and generate a subset of properties to be used in the QSAR equations construction. Therefore, the calculated values of a set of selected properties were set as the independent variables (X) used to calculate the QSAR equations along with the values of the dependent variable (Y), i.e., the biological activity values, which were converted from IC₅₀ (μM) (Table 1) to the corresponding pIC₅₀ (M) values, before the QSAR equations generation. Then, the QSAR equations were obtained by multiple linear regression (MLR) analysis, using the Microsoft Excel[®] program (Microsoft Inc.).

In addition, using the OSIRIS Property Explorer server¹⁶, the toxicity risks of the quinoxaline derivatives were evaluated *in silico* and fragment-based drug-likeness score was calculated in the same server.

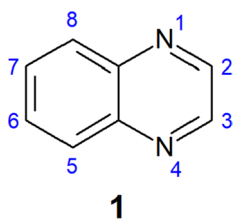
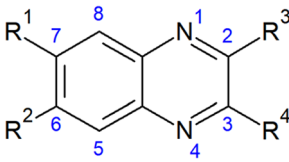


Figure 1. Chemical structure and atom numbering of the quinoxaline nucleus (**1**).



2a-2i, 3a-3i, and 4a-4d

#	R ¹	R ²	R ³	R ⁴	IC ₅₀ (μM)	pIC ₅₀ (M)
2a	H	OMe	NMe ₂	SMe	42.8	4.37
2b	H	OMe	NH(<i>n</i> -Bu)	SMe	35.2	4.45
2c	H	OMe	NH(cyclohexyl)	SMe	29.8	4.53
2d	OMe	H	NH(isopentyl)	SMe	27.1	4.57
2e	Br	H	NH(<i>n</i> -Bu)	SMe	25.2	4.60
2f	H	OMe	NH(isobutyl)	SMe	27.6	4.56
2g	Cl	H	NH(<i>n</i> -Bu)	SMe	24.4	4.61
2h	OMe	H	NH(isobutyl)	SMe	26.9	4.57
2i	OMe	H	NH(<i>n</i> -Bu)	SMe	30.2	4.52
3a	H	OMe	NH(<i>n</i> -Bu)	SO ₂ Me	2.5	5.60
3b	H	OMe	NH(cyclohexyl)	SO ₂ Me	2.9	5.54
3c	Br	H	NH(<i>n</i> -Bu)	SO ₂ Me	1.6	5.80
3d	H	OMe	NH(isobutyl)	SO ₂ Me	2.6	5.59
3e	Cl	H	NH(<i>n</i> -Bu)	SO ₂ Me	1.4	5.85
3f	Cl	H	NH(cyclohexyl)	SO ₂ Me	2.2	5.66
3g	H	H	NH(<i>n</i> -Bu)	SO ₂ Me	2.9	5.54
3h	Br	H	NH(EtOH)	SO ₂ Me	0.8	6.10
3i	Br	H	Cl	SO ₂ Me	0.2	6.70
4a	Cl	Cl	Ph	Ph	5.3	5.28
4b	H	H	4-OMe-Ph	4-OMe-Ph	30.0	4.52
4c	H	H	4-Me-Ph	Ph	8.9	5.05
4d	H	H	4-OMe-Ph	Ph	12.8	4.89

Table 1. Chemical structures of the quinoxaline derivatives (**2a–2i**, **3a–3i**, and **4a–4d**) and the corresponding *in vitro* inhibitory activities (IC₅₀, μM) against the promastigote forms of *Leishmania amazonensis*¹¹.

Organic synthesis

The proposed quinoxaline derivative (**5**) was synthesized following the route used by Cogo et al.¹¹ in the synthesis of 2-amino-3-sulfonylquinoxalines. This synthetic route consists in four steps: (i) vinylic substitution of 1,1-bis(methylsulfanyl)-2-nitroethene using 4-chloroaniline as nucleophile and ethanol as solvent to obtain the 4-chloro-*N*-(1-methylsulfanyl-2-nitroethenyl)aniline intermediate; (ii) cyclization of the intermediate with phosphoryl chloride (POCl₃) to produce the pyrazine ring of the quinoxaline nucleus of the 3,6-dichloro-2-methylsulfanylquinoxaline intermediate, using acetonitrile as solvent; (iii) microwave assisted nucleophilic substitution, using ethanol as solvent to install the methylamino substituent in the 2-position of the quinoxaline nucleus; and (iv) oxidation of the methylsulfanyl group with 3-chloroperbenzoic acid (mCPBA) to obtain the sulfone (**5**) in dichloromethane as solvent (Fig. 2).

The ¹H NMR spectra of all intermediates and final product were obtained by using a Bruker ARX-400 equipment (400 MHz).

In vitro growth inhibition assay

Promastigote (1 × 10⁶ cells/mL) cultures were inoculated in a 24-well plate in the absence or presence of different concentrations of the quinoxaline derivatives (0.1 and 100 μM). The inhibitory activity was evaluated after 72 h. The cell density for each concentration was determined by counting in a hemocytometer (Improved Double Neubauer). The concentration that inhibited cell growth in 50% (IC₅₀) was determined by nonlinear regression analysis¹¹.

Results and discussion

SAR analysis of the quinoxaline derivatives and design of a new derivative

Many descriptors reflect simple molecular properties give an insight referent to physicochemical nature of the observed biological activity¹⁷.

Table 2 shows the physicochemical descriptor values calculated at the DFT(B3LYP)/6-311 + +G(*d,p*) level of theory for the quinoxaline derivatives (**2a–2i**, **3a–3i**, and **4a–4d**). All the most active quinoxaline derivatives

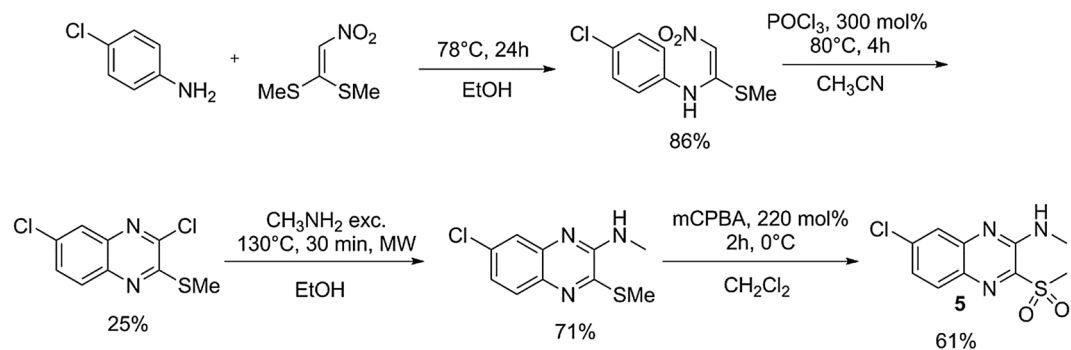


Figure 2. Synthetic route of 7-chloro-*N*-methyl-3-(methylsulfonyl)quinoxalin-2-amine (**5**).

#	E_T	E_{HOMO}	E_{LUMO}	GAP	μ	LogP	SA	MV	MW	P	HBD	HBA	PSA
2a	-1104.17	-5.80	-1.84	3.96	1.88	2.65	257.98	253.34	249.34	60.98	0	4	22.26
2b	-1182.83	-5.59	-1.69	3.90	2.35	3.51	319.31	289.61	277.39	63.94	0	4	30.50
2c	-1260.28	-5.56	-1.65	3.91	2.28	3.82	336.25	313.11	303.43	65.85	0	4	30.07
2d	-1222.16	-5.60	-1.58	4.02	2.08	4.03	334.48	306.94	291.42	65.32	0	4	28.70
2e	-3641.82	-6.01	-1.97	4.04	3.75	4.47	309.23	280.13	326.26	63.14	0	0	23.30
2f	-1182.83	-5.59	-1.67	3.92	2.34	3.14	314.10	289.20	277.39	62.91	0	4	28.27
2g	-1527.90	-5.99	-1.94	4.05	3.60	4.19	304.46	275.52	281.81	62.77	0	3	23.29
2h	-1182.83	-5.63	-1.62	4.01	1.79	3.14	314.10	289.20	277.39	63.88	0	4	28.15
2i	-1182.83	-5.61	-1.59	4.02	2.05	3.51	319.28	289.58	277.39	63.91	0	4	29.85
3a	-1333.23	-6.01	-2.54	3.47	4.19	2.48	333.49	302.54	309.39	65.09	1	7	60.56
3b	-1410.68	-5.95	-2.52	3.43	3.88	2.78	350.48	326.03	335.43	67.01	1	7	59.98
3c	-3792.22	-6.49	-2.79	3.70	2.93	3.43	323.38	293.04	358.26	64.27	1	6	53.34
3d	-1333.24	-6.03	-2.54	3.49	4.36	2.46	331.71	302.28	309.39	65.07	1	7	60.01
3e	-1678.30	-6.48	-2.76	3.72	2.99	3.16	318.58	288.42	313.81	63.89	1	6	53.32
3f	-1755.75	-6.41	-2.74	3.67	2.84	3.47	335.55	311.91	339.85	65.81	1	6	52.73
3g	-1218.68	-6.33	-2.62	3.71	3.94	2.60	303.53	275.14	279.36	62.82	1	6	53.38
3h	-3788.81	-6.62	-2.84	3.78	1.83	1.67	292.75	263.92	346.21	61.89	2	7	73.77
3i	-4039.16	-7.48	-3.22	4.26	5.52	2.88	243.84	220.52	321.58	58.26	0	5	46.79
4a	-1799.52	-6.65	-2.60	4.05	3.36	6.52	341.44	332.33	351.24	67.37	0	2	14.52
4b	-1109.39	-5.94	-2.16	3.78	2.35	5.15	371.49	360.45	342.40	69.72	0	4	28.79
4c	-919.61	-6.37	-2.24	4.13	1.14	5.89	332.14	324.08	296.37	66.68	0	2	14.52
4d	-994.84	-6.11	-3.22	2.89	1.86	5.28	341.21	333.16	312.37	67.48	0	3	21.65

Table 2. Physicochemical descriptors calculated at the DFT(B3LYP)/6-311 + + G(d,p) level of theory for the quinoxaline derivatives (**2a–2i**, **3a–3i**, and **4a–4d**) using the Spartan'10 software. Data for the most active compounds (**3a–3i**) ($IC_{50} < 3 \mu M$, see Table 1) are highlighted in italic. E_T , total energy (au); E_{HOMO} , energy (eV) of the highest occupied molecular orbital; E_{LUMO} , energy (eV) of the lowest unoccupied molecular orbital; GAP, HOMO–LUMO gap energy (eV); μ , dipole moment (Debye); LogP, base-10 logarithm of the partition coefficient; SA, surface area (\AA^2); MV, molecular volume (\AA^3); MW, molecular weight (amu); P, polarizability (10^{-30} m^3); HBD, number of hydrogen bond donors (NH + OH), HBA, number of hydrogen bond acceptors (N + O); PSA, polar surface area (\AA^2).

($IC_{50} < 3 \mu M$, i.e., pIC_{50} from 5.54 to 6.70 M, compounds **3a–3i**, see Table 1) presented the number of hydrogen bond acceptors (HBA) ranging from 5 to 7, the polar surface area (PSA) values ranging from 46 to 74 \AA^2 , and the LUMO energy (E_{LUMO}) values more negative than -2.5 eV. In addition, the LogP values range from 1.6 to 3.5, and the HOMO energy (E_{HOMO}) values are more negative than -5.9 eV.

Unfortunately, the fragment-based drug-likeness values predicted by the OSIRIS server for these compounds are negative like most of the Fluka chemicals that have negative values, whereas 80% of the commercial drugs have a positive drug-likeness value. Toxicity was also predicted by the OSIRIS, and compounds **3g** and **4a–d** showed alerts of mutagenic risks. On the other hand, **3d** showed the highest drug-score value (0.63). The drug-score index combines drug-likeness, cLogP (lipophilicity), LogS (water solubility), MW, and toxicity risks in one value used to predict the compound's overall potential as a drug.

Lipinski's rule-of-five¹⁸ proposes that poor absorption or cell permeability of a drug occurs when its chemical structure fulfils more than one of the following criteria: the molecular weight (MW) is greater than 500 Daltons;

the calculated LogP is greater than 5; the number of hydrogen bond donors (NH + OH) are more than 5; and the number of hydrogen bond acceptors (N + O) are more than 10. According to the Veber's rule¹⁹, for good oral availability, the PSA value must be less than or equal to 140 Å². The physicochemical properties calculated for the studied compounds fit these parameters, except the LogP values for **4a–d** (Table 2).

In order to improve these parameters, structural modifications on the studied compounds were proposed to design an antileishmanial agent with higher chances to become a drug.

Cogo and co-workers¹¹ noticed that hydrogen replacement at R1 position (Table 1) by halogen elements (Cl or Br) increases the activity, and substitution at R2 position (Table 1) did not show great interference on the activity. The methylsulfonyl group is present in all the most active compounds studied in this work (Table 1) and literature data also indicates that it is one of the main groups at 3-position of quinoxaline derivatives, which are responsible for the observed activity against *Trypanosoma cruzi* and *Leishmania amazonensis*¹¹.

Based on this SAR analysis, several structural modifications were proposed and their synthetic viability as well as the OSIRIS Property Explorer's risk alerts were evaluated. After that, some of the designed compounds were selected for structural optimization and calculation of the corresponding physicochemical properties. Considering the properties related to the biological activity, compound **5** was proposed as a potential antileishmanial agent (Fig. 3).

It fulfilled all the requirements, presenting the physicochemical descriptors according to the most active compounds of the studied series: LUMO energy of -2.79 eV, five H-bond acceptors, polar surface area of 53.18 Å², LogP equal to 1.74, and HOMO energy of -6.52 eV.

Besides, according to the OSIRIS Property Explorer server, compound **5** (Fig. 3) seems to have low toxicity risks (green color) and the drug-likeness and drug-score indexes were improved to 0.88 and 0.82, respectively, when compared to the other compounds of the series. It is also important to mention that compound **5** follows Lipinski's rule-of-five and Veber's rule related to PSA range of drug candidates.

QSAR model construction and validation

A QSAR model was built to predict the activity value of compound **5**. Firstly, the degree of correlation between all pairs among the thirteen descriptors (Table 1) was verified by constructing a cross-correlation matrix. After removing multicollinear descriptors, seven of them were selected (E_T , E_{HOMO} , E_{LUMO} , dipole moment, LogP, MW, and PSA), and equations that describe the dependency relationship between the independent (X, properties or descriptors) and dependent (Y, biological activity) variables were obtained based on Hansh and Unger's work²⁰, who suggest that, in a selection of independent variables, for each independent variable included in the QSAR model, there must be no more than five observations (i.e., compounds), thus avoiding chance correlation²¹.

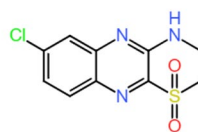
Therefore, the calculated values of those seven descriptors (Table 1) were set as the independent (X) variables used to calculate the QSAR equations along with the values of the dependent (Y) variable (i.e., biological activities), which were converted from IC₅₀ (μM) (Table 1) to the corresponding pIC₅₀ (M) values, before the QSAR equations generation.

Among the main methods used in the selection of the independent variables in QSAR, we applied the systematic search method, which consists in combining the available independent variables to build and analyze

OSIRIS Property Explorer

Predicted toxicity risks

- mutagenic
- tumorigenic
- irritant
- reproductive effective



Predicted properties

cLogP	■	0.80
Solubility	■	-2.92
Molweight	■	271.73
TPSA	■	80.33
Druglikeness	■	0.88
H bond acceptor	■	5
H bond donor	■	1
Nb stereocenters	■	0
Nb rotatable bonds	■	2
Drug-Score	■	0.82

Figure 3. Toxicity risks and physicochemical properties predicted by the OSIRIS Property Explorer server for compound (**5**), 7-chloro-N-methyl-3-(methylsulfonyl)quinoxalin-2-amine, proposed as a potential antileishmanial agent.

all possible linear regression equations. In the QSAR method, compounds are generally divided into training set and test set, compounds from the training set are used in the construction of QSAR equations and compounds from the test set are used in validation. Since there are 22 compounds (Table 1) and that part of them (~20% from the total number of compounds) should be removed from the model as a test group, we used a maximum of three independent variables to be included in each equation, considering $N = 18$ for the training set and $N = 4$ for the test set (namely, compounds **2i**, **3g**, **3h**, and **4d**).

The systematic search generated 63 regression equations: seven equations with one independent variable, 20 equations with two independent variables, and 34 equations with three independent variables. Tables 3, 4 and 5 list the previously selected independent variables included in the linear equations and the following statistical parameters of each equation calculated by the Microsoft Office Excel® program (Microsoft Inc.): correlation coefficient (R), coefficient of determination (R^2), adjusted coefficient of determination (R^2_{Adj}), standard error (s) and F-test.

Comparing the best equations (highlighted in bold on Tables 3, 4 and 5) of the three groups containing one (Eqs. 1–7), two (Eqs. 8–28) and three (Eqs. 29–63) theoretical physicochemical descriptors (independent variables or terms), i.e., Eq. 2 ($pIC_{50} = -0.84 (E_{HOMO})$, $R^2_{Adj} = 0.936$), Eq. 18 ($pIC_{50} = -1.58 - 0.97 (E_{HOMO}) + 0.02 (PSA)$, $R^2_{Adj} = 0.967$), and Eq. 33 ($pIC_{50} = -1.85 + 4.31 \times 10^{-5} (E_T) - 1.02 (E_{HOMO}) + 0.02 (PSA)$, $R^2_{Adj} = 0.967$), respectively, the E_{HOMO} term is present in all of them. However, Eq. 2 should be excluded because it has the lowest R^2_{Adj} value (a normalized R^2 value, used to compare equations containing a different number of terms).

Therefore, considering only Eqs. 18 and 33, we can observe that the inclusion of the E_T term in Eq. 33 does not alter the R^2_{Adj} value, which makes these two equations to be equivalent. Nevertheless, since the parsimony

Eq.	E_T	E_{HOMO}	E_{LUMO}	μ	LogP	MW	PSA	R	R^2	R^2_{Adj}	s	F
1	X							0.541	0.293	0.249	0.579	6.624
2		X						0.997	0.995	0.936	0.390	3119.914
3			X					0.947	0.897	0.890	0.222	138.659
4				X				0.687	0.472	0.439	0.501	14.315
5					X			0.266	0.071	0.013	0.664	1.220
6						X		0.561	0.315	0.272	0.570	7.359
7							X	0.702	0.493	0.461	0.490	15.567

Table 3. Statistical data for the seven QSAR equations with one term ($N = 18$ and $p = 0.05$), generated by systematic combination of the seven theoretical physicochemical descriptors.

Eq.	E_T	E_{HOMO}	E_{LUMO}	μ	LogP	MW	PSA	R	R^2	R^2_{Adj}	s	F
8	X	X						0.843	0.711	0.672	0.383	18.441
9	X		X					0.948	0.898	0.885	0.227	66.260
10	X			X				0.710	0.504	0.438	0.501	7.624
11	X				X			0.590	0.348	0.261	0.575	4.003
12	X					X		0.635	0.404	0.324	0.549	5.079
13	X						X	0.808	0.653	0.606	0.419	14.100
14		X	X					0.947	0.897	0.883	0.229	65.065
15		X		X				0.882	0.777	0.747	0.336	26.142
16		X			X			0.928	0.861	0.843	0.265	46.504
17		X				X		0.852	0.726	0.689	0.373	19.826
18		X					X	0.985	0.971	0.967	0.122	247.151
19			X	X				0.951	0.904	0.892	0.220	70.911
20			X		X			0.964	0.929	0.920	0.189	98.758
21			X			X		0.954	0.910	0.898	0.213	76.049
22			X				X	0.959	0.920	0.909	0.201	86.098
23				X	X			0.688	0.474	0.404	0.516	6.753
24				X		X		0.753	0.566	0.509	0.469	9.794
25				X			X	0.785	0.616	0.565	0.441	12.053
26					X	X		0.725	0.526	0.463	0.490	8.326
27					X		X	0.769	0.592	0.538	0.454	10.887
28						X	X	0.771	0.594	0.540	0.453	10.992

Table 4. Statistical data for the 20 QSAR equations with two terms ($N = 18$ and $p = 0.05$) generated by systematic combination of seven theoretical physicochemical descriptors.

Eq.	E _T	E _{HOMO}	E _{LUMO}	μ	LogP	MW	PSA	R	R ²	R ² _{Adj}	s	F
29	X	X	X					0.948	0.898	0.877	0.235	41.278
30	X	X		X				0.887	0.787	0.742	0.340	17.288
31	X	X			X			0.932	0.869	0.841	0.266	30.989
32	X	X				X		0.853	0.727	0.669	0.385	12.440
33	X	X					X	0.986	0.973	0.967	0.121	167.267
34	X		X	X				0.951	0.904	0.884	0.228	44.193
35	X		X		X			0.965	0.931	0.916	0.193	62.979
36	X		X			X		0.957	0.915	0.897	0.214	50.399
37	X		X				X	0.962	0.925	0.908	0.202	57.251
38	X			X	X			0.714	0.509	0.404	0.516	4.842
39	X			X		X		0.756	0.572	0.480	0.482	6.225
40	X			X			X	0.822	0.676	0.607	0.419	9.741
41	X				X	X		0.749	0.560	0.466	0.488	5.948
42	X				X		X	0.850	0.723	0.663	0.388	12.157
43	X						X	0.821	0.675	0.605	0.420	9.675
44		X	X	X				0.951	0.905	0.884	0.228	44.228
45		X	X		X			0.970	0.941	0.929	0.178	74.858
46		X	X			X		0.954	0.910	0.891	0.221	47.389
47		X	X				X	0.986	0.972	0.966	0.124	160.655
48		X		X	X			0.931	0.866	0.838	0.269	30.254
49		X		X		X		0.886	0.784	0.738	0.342	16.954
50		X		X			X	0.985	0.971	0.965	0.126	155.383
51		X			X	X		0.958	0.918	0.900	0.211	52.002
52		X			X		X	0.985	0.971	0.964	0.126	153.837
53		X					X	0.986	0.972	0.966	0.124	160.256
54			X	X	X			0.964	0.930	0.915	0.195	61.996
55			X	X		X		0.957	0.917	0.899	0.213	51.360
56			X	X			X	0.960	0.922	0.906	0.205	55.310
57			X		X	X		0.965	0.930	0.915	0.194	62.278
58			X		X		X	0.964	0.929	0.914	0.196	61.500
59			X			X	X	0.965	0.931	0.917	0.193	63.229
60				X	X	X		0.787	0.619	0.537	0.455	7.585
61				X	X		X	0.833	0.694	0.628	0.407	10.581
62				X		X	X	0.820	0.672	0.602	0.422	9.579
63					X	X	X	0.775	0.601	0.515	0.465	7.019

Table 5. Statistical data for the 34 QSAR equations with three terms (N = 18 and $p = 0.05$) generated by systematic combination of seven theoretical physicochemical descriptors.

principle advises the choice of the simplest model, Eq. 18 was used to calculate the antileishmanial activity value for the 18 quinoxaline derivatives (Table 6), using as descriptors the E_{HOMO} and PSA independent variables (Table 4).

The HOMO and LUMO energies are important properties in chemical and pharmacological processes because these properties give information on the electron-donating and electron-accepting character of a compound. It is possible to notice that the E_{HOMO} for the most active studied compounds are the more negative ones (Table 2). This means that the more active compounds are not so good electron-donor molecules when compared to the less active ones.

The PSA is a molecular descriptor extensively used to characterizing the transport properties of drugs, related to its intestinal absorption and the penetration of the blood–brain barrier. According to the model (Eq. 18), together with E_{HOMO}, it is a key descriptor to explain the biological activity of the quinoxaline derivatives.

These descriptors shows that not only steric but also electronic properties are important to understand the interaction between quinoxaline derivatives that present antileishmanial activity and the biological receptor. The steric properties are related to the positioning of the molecule when interacting with the receptor, while the electronic properties are related to the intensity of the molecular association due to electronic interaction.

Among the 18 compounds, only one (4b) presents a deviation greater than 5% from the experimental activity value, characterizing it as an outlier. Compound 4b has bulky substituents, altering its physicochemical properties (such as a much larger area and volume values—see Table 2) when compared to the other compounds of the series, and consequently making discrepant the relationship between structure and biological activity through the proposed equation.

#	pIC ₅₀ (observed)	pIC ₅₀ (calculated)	Residue	Deviation (%)
2a	4.37	4.49	-0.12	2.75
2b	4.45	4.45	0.00	0.00
2c	4.53	4.41	0.12	2.65
2d	4.57	4.43	0.14	3.06
2e	4.60	4.72	-0.12	2.61
2f	4.56	4.41	0.15	3.29
2g	4.61	4.70	-0.09	1.95
2h	4.57	4.44	0.13	2.84
3a	5.60	5.46	0.14	2.50
3b	5.54	5.39	0.15	2.71
3c	5.80	5.78	0.02	0.34
3d	5.59	5.47	0.12	2.15
3e	5.85	5.77	0.08	1.37
3f	5.66	5.69	-0.03	0.53
3i	6.70	6.61	0.09	1.34
4a	5.28	5.16	0.12	2.27
4b	4.52	4.76	-0.24	5.31
4c	5.05	4.89	0.16	3.17

Table 6. Observed (experimental) and calculated (Eq. 18) pIC₅₀ (M) values, residuals (pIC₅₀ (observed)–pIC₅₀ (calculated)), and percent deviation.

Excluding the outlier 4b, coefficients were recalculated for Eq. 18 providing Eq. 1 (N = 17, R² = 0.980, R²_{Adj} = 0.977, s = 0.103, and R² from the leave-one-out-cross-validation (Q²) = 0.971), in which analysis of residues (Table 7) and plot of pIC₅₀ (calculated) versus pIC₅₀ (observed) (Fig. 4) did not show any outlier.

$$\text{pIC}_{50} = -1.51 - 0.96(\text{E}_{\text{HOMO}}) + 0.02(\text{PSA}) \quad (1)$$

Since literature data indicates that there is evidence that only models validated externally, after internal validation, can be considered reliable and applicable for external prediction and regulatory purposes^{22,23}, the model was applied for external molecules.

#	pIC ₅₀ (observed)	pIC ₅₀ (calculated)	Residue	Deviation (%)
2a	4.37	4.49	-0.12	2.75
2b	4.45	4.45	0.00	0.00
2c	4.53	4.41	0.12	2.65
2d	4.57	4.43	0.14	3.06
2e	4.60	4.72	-0.12	2.61
2f	4.56	4.41	0.15	3.29
2g	4.61	4.70	-0.09	1.95
2h	4.57	4.44	0.13	2.84
3a	5.60	5.46	0.14	2.50
3b	5.54	5.39	0.15	2.71
3c	5.80	5.78	0.02	0.34
3d	5.59	5.47	0.12	2.15
3e	5.85	5.77	0.08	1.37
3f	5.66	5.69	-0.03	0.53
3i	6.70	6.61	0.09	1.34
4a	5.28	5.16	0.12	2.27
4c	5.05	4.89	0.16	3.17

Table 7. Observed (experimental) and calculated (Eq. 1) pIC₅₀ (M) values, residuals (pIC₅₀ (observed)–pIC₅₀ (calculated)), and percent deviation after removing outlier 4b.

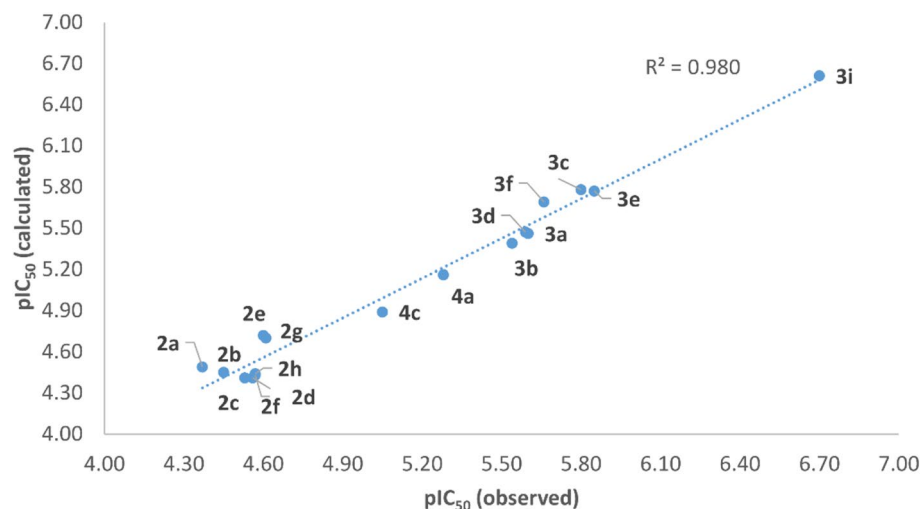


Figure 4. Experimentally observed antileishmanial activity values (pIC_{50} (observed)) versus calculated activity values (pIC_{50} (calculated)), using Eq. 1, for the 17 quinoxaline derivatives in the training set (after removing outlier 4b).

Carrying out an external validation, it was possible to confirm the robustness of the proposed model (Eq. 1). A set of four compounds (2i, 3g, 3h, and 4d) was used as external test, representing about 20% of the quantity of observations ($N = 22$). The test group with its values of observed and calculated pIC_{50} , residuals, and percentual deviation are shown in Table 8, where is possible to verify that all of them present a deviation smaller than or equal to 5% of the biological activity value observed experimentally.

Synthesis of the new derivative and activity prediction by the QSAR model

Unpublished compound 5 was synthesized, characterized by NMR, and its biological activity in the promastigote form of *Leishmania amazonensis* was evaluated. The built and validated QSAR model, corresponding to Eq. 1, was used to predict the activity of this new derivative.

Therefore, the descriptors present in Eq. 1 were calculated for the new compound 5 ($E_{HOMO} = -6.52$ eV and $PSA = 53.19 \text{ \AA}^2$) and a value of 5.81 was predicted for biological activity (pIC_{50}) against *Leishmania amazonensis*. Comparison with the experimental result ($IC_{50} = 2.0 \pm 1.2 \text{ \mu M}$ and $pIC_{50} = 5.70$) shows that the QSAR model (Eq. 1) proposed here, presented a good predictive capacity with a deviation of 1.93%, being useful to drive the synthesis of new quinoxaline derivatives, saving time and resources that would be spent on synthesis and testing of biological activity.

NMR data

7-chloro-*N*-methyl-3-(methylsulfonyl)quinoxalin-2-amine (5).

61% yield. 1H NMR (300 MHz, $CDCl_3$) δ : 7.77 (d, $J = 8.9$ Hz, 1H), 7.72 (d, $J = 2.3$ Hz, 1H), 7.35 (dd, $J = 8.9$, 2.3 Hz, 1H), 6.93 (m, 1H), 3.41 (s, 3H), 3.12 (d, $J = 4.8$ Hz, 3H). ^{13}C NMR (300 MHz, $CDCl_3$) δ : 148.68, 144.21, 141.19, 138.69, 132.96, 130.47, 126.38, 125.43, 40.45, 27.92.

Conclusions

SAR studies of a series of quinoxaline derivatives were carried out and a new quinoxaline derivative was proposed as a potential antileishmanial agent. The unpublished compound was synthesized and tested against *Leishmania amazonensis* promastigotes. A new QSAR model was built, and it was capable to predict the activity of the new compound being useful to drive the synthesis of other ones.

#	pIC_{50} (observed)	pIC_{50} (calculated)	Residue	Deviation (%)
2i	4.52	4.47	0.05	1.11
3g	5.54	5.63	-0.09	1.62
3h	6.10	6.32	-0.22	3.61
4d	4.89	4.79	0.10	2.04

Table 8. Observed (experimental) and calculated (Eq. 1) pIC_{50} values, residues (pIC_{50} (observed) - pIC_{50} (calculated)), and percent deviation for the test set compounds.

Data availability

The authors confirm all data generated and analyzed during this study are available in the article.

Received: 14 August 2023; Accepted: 19 October 2023

Published online: 24 October 2023

References

- World Health Organization. Global report on neglected tropical diseases 2023; Retrieved Aug 2023 from WHO/CED/PHE/18.10; World Health Organization: Geneva, Switzerland, 2023.
- Fonseca, B. P., Albuquerque, P. C. & Zicker, F. Neglected tropical diseases in Brazil: Lack of correlation between disease burden, research funding and output. *Trop. Med. Int. Health* **25**, 1373–1384. <https://doi.org/10.1111/tmi.13478> (2020).
- World Health Organization. Ending the neglect to attain the Sustainable Development Goals: A road map for neglected tropical diseases 2021–2030; World Health Organization, 2021. Retrieved Feb 2023.
- Varikuti, S. *et al.* Host-directed drug therapies for neglected tropical diseases caused by protozoan parasites. *Front. Microbiol.* **9**, 2655. <https://doi.org/10.3389/fmicb.2018.02655> (2018).
- World Health Organization. Leishmaniasis; World Health Organization, 2023. Retrieved Feb 2023.
- Tuon, F. F., Amato, V. S., Zequinao, T. & Cruz, J. A. W. Emerging computational technologies in human leishmaniasis: Where are we?. *Trans. R. Soc. Trop. Med. Hyg.* **116**, 981–985. <https://doi.org/10.1093/trstmh/trac047> (2022).
- Kerru, N., Gummidi, L., Maddila, S., Gangu, K. K. & Jonnalagadda, S. B. A review on recent advances in nitrogen-containing molecules and their biological applications. *Molecules* **25**, 1909. <https://doi.org/10.3390/molecules25081909> (2020).
- Rivera, G. Quinoxaline 1,4-di-*N*-oxide derivatives: Are they unselective or selective inhibitors?. *Mini-Rev. Med. Chem.* **22**, 15–25. <https://doi.org/10.2174/1389557521666210126142541> (2022).
- Montana, M., Montero, V., Khoumeri, O. & Vanelle, P. Quinoxaline derivatives as antiviral agents: A systematic review. *Molecules* **25**, 2784. <https://doi.org/10.3390/molecules25122784> (2020).
- Aakasha, V. B., Ramalakshmia, N., Bhuvaneshwaria, S., Sankaria, E. & Arunkumar, S. Comprehensive review on versatile pharmacology of quinoxaline derivative. *Russ. J. Bioorg. Chem.* **48**, 657–677. <https://doi.org/10.1134/S1068162022040069> (2022).
- Cogo, J. *et al.* Synthesis and biological evaluation of novel 2,3-disubstituted quinoxaline derivatives as antileishmanial and antitrypanosomal agents. *Eur. J. Med. Chem.* **90**, 107–123. <https://doi.org/10.1016/j.ejmech.2014.11.018> (2015).
- Cogo, J. *et al.* Quinoxaline derivatives as potential antitrypanosomal and antileishmanial agents. *Bioorg. Med. Chem.* **26**, 4065–4072. <https://doi.org/10.1016/j.bmc.2018.06.033> (2018).
- Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* **29**, 476–488. <https://doi.org/10.1002/minf.201000061> (2010).
- Cherkasov, A. *et al.* QSAR modeling: Where have you been? Where are you going to?. *J. Med. Chem.* **57**, 4977–5010. <https://doi.org/10.1021/jm4004285> (2014).
- SPARTAN'10, Wavefunction Inc. Irvine, CA 92612, U.S.A., 2011.
- Sander, T. OSIRIS Property Explorer. Organic Chemistry Portal (2001). <https://www.organic-chemistry.org/prog/peo>
- Karelson, M., Lobanov, V. S. & Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **96**, 1027–1044. <https://doi.org/10.1021/cr950202r> (1996).
- Lipinski, C. A. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov. Today: Technol.* **1**, 337–341. <https://doi.org/10.1016/j.ddtec.2004.11.007> (2004).
- Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623. <https://doi.org/10.1021/jm020017n> (2002).
- Hansch, C. & Unger, S. H. Strategy in drug design. Cluster analysis as an aid in the selection of substituents. *J. Med. Chem.* **26**, 1217–1221 (1973).
- Gaudio, A. C. & Zonade, E. Proposição, validação e análise de modelos que correlacionam estrutura química e atividade biológica. *Quím. Nova.* **24**, 658–671 (2001).
- Gramatica, P. Principles of QSAR models validation: Internal and external. *QSAR Comb. Sci.* **26**(5), 694–701. <https://doi.org/10.1002/qsar.200610151> (2007).
- Golbraikh, A. & Tropsha, A. Beware of q²!. *J. Mol. Graph. Model.* **20**, 269–276. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1) (2002).

Acknowledgements

To the Brazilian governmental agencies: FAPERJ (“Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro”), CAPES (“Coordenação de Aperfeiçoamento de Pessoal de Nível Superior”), and CNPq (“Conselho Nacional de Desenvolvimento Científico e Tecnológico”).

Author contributions

Project design, students supervision, writing the manuscript and approval of the final version of the manuscript were the responsibility of J.Y. Suggestion of research idea was performed by D.P.S. He and J.A.D were responsible for the organic synthesis. J.A.D. also proposed the structural modifications on the studied compounds. C.V.N. realized the biological tests. F.L.L. worked on theoretical calculations for the SAR study. A.C.S.J.P. developed the QSAR study together with M.G.A., which reviewed the manuscript and also approved the final version. The authors confirm that the individuals named are the authors of this work, and are satisfied to publish their research work in Journal of Scientific Reports.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023