RESEARCH ARTICLE

Molecular Modeling Studies of Benzimidazolyl-Chalcones as Antileishmanial Agents using Qsar, Docking, ADME and Molecular Dynamics Studies

N'Guessan DU Jean-Paul[1], Alzain A Abdulrahim[2], Adouko AM Eunice[1], Sékou Diomandé[3], Coulibaly Songuigama[1], Kablan AC Landry[4,5], Koné Soleymane[3], Bories Christian[6], Sissouma Drissa[3], Loiseau M Philippe[6], Ouattara Mahama[1*]

Abstract

Introduction: Present leishmaniasis treatment regimen has many limitations including severe adverse effects, toxicity, and *Leishmania* strains resistance. In the present study, the objective is to perform QSAR, molecular docking and ADME prediction studies on benzimidazolylchalcones in order to select an antileishmanial drug candidate.

Materials & methods: QSAR models were performed on 12 benzimidazolylchalcones with antileishmanial activities against promastigote strains of *L. donovani*. Binding free energy calculations were performed using MM-GBSA to assess the affinity of the ligands for the proteins. In addition, the three most active compounds (4a-c, IC_{50} <1-µM) were docked with the protein phosphodiesterase B1 (PDB ID: 2JK6).

Results and Discussion: The optimum model has squared correlation coefficient (R^2) of 0.983, and leave-one-out (LOO) cross-validation coefficient (Q^2_{CV}) value of 0.942. The number of descriptors involved in the model is acceptable (R^2 - Q^2_{CV} = 0.041), which confirms the model's stability and validates the developed model's predictive power. Docking studies revealed that the best compound 4c formed hydrogen bond with SER 464, pi-cation contact with LYS 61 and hydrophobic interactions with LEU 62, TYR 64 and LEU72 of the active site of *L. donovani* phosphodiesterase B1. ADME properties results showed that all three molecules have good pharmacokinetic properties.

Conclusion: Finally, molecular dynamics simulation studies at 30 ns revealed stable interactions with the 2JK6 protein. This study validates the choice of the ortho-chlorinated derivative of benzimidazolylchalcones as the lead compound for developing new derivatives with optimized antileishmanial properties.

Keywords: QSAR, Docking, molecular dynamic, benzimidazolyl-chalcones, antileishmanial, *L. donovani*.

Journal of Applied Pharmaceutical Sciences and Research, (2022); DOI: 10.31069/japsr.v4i3.4

Introduction

Leishmaniasis is an infectious disease transmitted by the female sandfly that has claimed more than 12 million victims worldwide in recent years.[1] Its co-infection with HIV-AIDS is also a major public health problem, as more than a thousand resistant strains have recently been isolated from patients living with HIV.[1] At present, no vaccine has been developed. Existing treatments such as pentamidine derivatives in the first line, and amphotericin B in the second line, are not only very toxic and costly, but above all have shown their ineffectiveness against chemo-resistant strains that are becoming increasingly widespread.[1] Hence, the researchers focused their interest on the design of new low-cost, low-toxicity potential anti-leishman drugs. Therefore, we have been interested in heteroaryl-chalcones which have been shown to possess various pharmacological properties including anti-infectious and anti-parasitic antileishmanial activities. [2] Through analyses carried out on several studies, we observed that the particularity of the pharmacological activities presented by the heteroaryl-chalcones was related to the nature of the heteroaryl associated with the chalcones. [2-5] Thus, in order to obtain targeted antileishmanial activities, benzimidazolyl-chalcones were designed as a result of the juxtaposition of the benzimidazole ring and the arylpropenone chain of the chalcones. These two chemical

[1]Department of Therapeutic Chemistry and Organic Chemistry, UFR Pharmaceutical and Biological Sciences, FHB University, 01 BP V34 Abidjan, Côte d'Ivoire.

[2]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Gezira, Gezira, Sudan

[3]Laboratory of Structural Organic Chemistry, UFR Structural Sciences of Matter and Technology, Félix Houphouët-Boigny University, 22 BP 582 Abidjan 22, Côte d'Ivoire

[4]Department of Mathematics, Physics and Chemistry, UPR of Organic Chemistry, UFR of Biological Sciences, Peleforo Gon Coulibaly University, BP 1328 Korhogo, Côte d'Ivoire

[5]Laboratory of Constitution and Reaction of Matter, UFR Structural Sciences of Matter and Technology, Félix Houphouët-Boigny University, 22 BP 582 Abidjan 22, Côte d'Ivoire

[6]Antiparasitic Chemotherapy Laboratory, UMR BioCIS, CNRS, Paris-Saclay University, 92296 Châtenay-Malabry, France

Corresponding Author: Ouattara Mahama, Antiparasitic Chemotherapy Laboratory, UMR BioCIS, CNRS, Paris-Saclay University, 92296 Châtenay-Malabry, France, Email: mahama.ouattara@univ-fhb.edu.ci

How to cite this article: Jean-Paul NDU, Abdulrahim AA, Eunice AAM, Diomandé S, Songuigama C, Landry KAC, Soleymane K, Christian B, Drissa S, Philippe LM, Mahama O. Molecular Modeling Studies of Benzimidazolyl-Chalcones as Antileishmanial Agents using Qsar, Docking, ADME and Molecular Dynamics Studies. Journal of Applied Pharmaceutical Sciences and Research. 2021; 4(3):18-28.

Source of support: Nil

Conflict of interest: None

entities have previously shown strong antileishmanial potentialities. Indeed, the benzimidazole ring, present in a large number of biologically active compounds, is known to possess remarkable antiprotozoal properties.[5,6] Also, in a specific way, recent studies have highlighted the anti-infectious potential of benzimidazole and its derivatives as powerful antileishmanial agents.[7,8] Moreover, the substitution of this heteroaryl by a functional group in its position 2 seems judicious because most of the biologically active compounds generally carry a functional group in this position.[6] As for the arylpropenones, known anti-infectives, antiprotozoals, they have been shown to be potent inhibitors of the development of several species of the genus *Leishmania* both *in vitro* and *in vivo* with low cytotoxicity.[1,8] Therefore, on this rational basis of the different antileishmanial results obtained for the benzimidazole ring and the arylpropenone chain, the association of these two chemical entities for the development of new antileishmanial drug candidates seemed to be judicious. In a previous study, our team showed that benzimidazolyl-chalcones were endowed with antileishmanial properties.[9] SAR studies, summarized in Figure 1, led to the characterization of a lead compound (ortho chloro derivative 4c) that showed very good *in vitro* activity against the promastigote stage of *L. donovani*.[9]

Subsequently in this study, ligand drug design approach such as quantitative structure-activity relationship studies (QSARs) of antileishmanial benzimidazolyl-chalcone compounds have been undertaken to establish a molecular model capable of predicting the biological activity of future analogues.[10] Indeed, QSARs are an innovative, efficient and rational tool in molecular modeling studies, particularly useful in the development of molecular leads with specific activity orientation.

The objective assigned to the present study is to highlight the expected chemical properties of future benzimidazolyl-chalcone analogues for antileishmanial purposes from the interpretation of the constructed QSAR model. Thus, DFT reactivity descriptors have been determined to optimize the compounds and predict their reactivity. The multiple linear regression (MLR) method was used to select the descriptors and to establish the correlation model that links the structural features of the compounds to their biological activities.

Moreover, QSARs can be coupled with structure-based drug design such as molecular docking, MM-GBSA and molecular dynamics, which highlight the best bonds and conformation of a ligand with its protein target and their stability. The combination of these tools allows the conceptualization of a pharmacophore whose interactions with the specific target are optimized. Therefore, we have performed the molecular docking, MM-GBSA and molecular dynamics of the three compounds with the best activities 4a, 4b and 4c, with Leishmania phosphodiesterase B1 (PDB ID: 2JK6).[11] Moreover, ADME properties were predicted using QikProp module of Schrödinger suite. [12] Furthermore, we investigated the interaction of the designed hybrids with the binding site of this enzyme, in order to gain structural insight for improved antileishmanial activity.

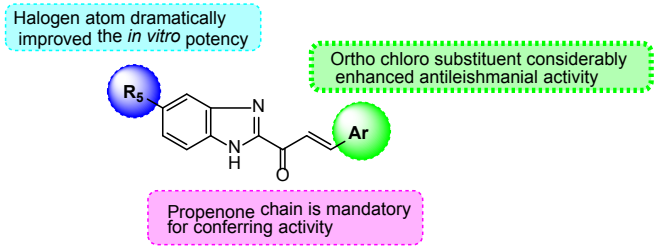


Figure 1: SAR of benzimidazolyl-chalcone derivatives as antileishmanial agents

Materials and Methods

Experimental Data Base

The present study was conducted on a pool of twelve benzimidazolyl-chalcones. The benzymidazolyl-chalcones (4a-l) result from the condensation of 2-acetylbenzimidazole (3a-c) with various benzaldehydes by the Claisen-Schmidt reaction.[13]

The obtained molecules were characterized by ^{1}H NMR (300MHz) and ^{13}C NMR (75 MHz). They were then evaluated for their anti-leishmanial activities, according to the *in vitro* method of live cell counting or quantitative colorimetric assay by methyltretazolium (MTT test) on LV9 strains of *Leishmania donovani* in the previous paper.[10]

The anti-*L. donovani* activities of the benzimidazolyl-chalcone derivatives studied are reported in the Table 1.

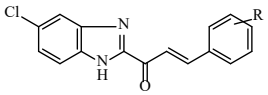
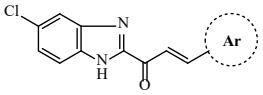
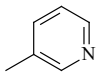
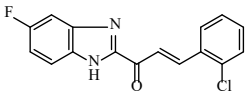
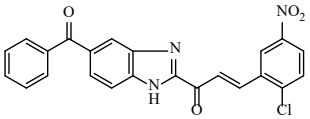
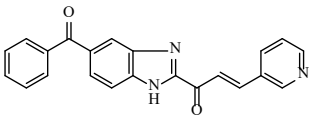
In addition, the strategy for developing a QSAR model requires the establishment of a structure-activity database from quantitative, reliable and normalized experimental measurements of the target activity expressed in molar concentration c, for each compound in the series studied. [14,15] For this purpose, the homogeneity and normality of the experimental data was ensured by converting the concentrations obtained into logarithm (log) according to the expression:

$$\text{Log}\frac{1}{C}$$

This is indeed a way to make the distribution of biological data normal without changing the information contained in the dataset. Representation of the structures and determination of the descriptors. The structures of our antileishmanial benzimidazolylchalcone molecules were represented in 3D and then optimized using Gaussian 5.0 software.

Subsequently, on these optimized structures, frequency calculations were performed by DFT at the B3LYP/6-31++G level through the Gaussian 5.0 software which integrates quantum mechanical algorithms. These frequency calculations provided the electronic quantum descriptors of electronic energy, boundary molecular orbital energies (E_{HOMO}, E_{LUMO}), dipole moment μ and polarizability. Global reactivity descriptors were subsequently calculated from these data. In fact, for this study, we chose to focus on

Table I: Benzylmidazolyl-chalcones derivatives and their anti-*L. donovani* activities

Structure	Compounds	R / Ar	*L. donovani*	
			IC_{50} (mg /L)	IC_{50} µM
	4a	H	0.15	0.53 ± 0.05
	4b	3-OH	0.15	0.50 ± 0.05
	4c	2-Cl	0.15	0.47 ± 0.04
	4d	3-NO_2	24.4	74.45 ± 7.44
	4e	4-CH_3	0.31	1.04 ± 0.10
	4f	4-Cl	7.80	24.59 ± 2.46
	4g	2,4-diCl	0.63	1.79 ± 0.18
	4h		0.31	1.14 ± 0.11
	4i		3.90	13.79 ± 1.38
	4j		0.40	1.33 ± 0.13
	4k		45.70	100.80 ± 10.08
	4l		18.60	43.07 ± 4.30

global reactivity descriptors of molecules because they are widely used to understand the overall chemical nature of the molecule and predict their chemical reactivity.[16] The determined descriptors are energy gap or HOMO-LUMO gap ΔE, Electron Affinity (EA), Ionic Energy (IE), electronegativity χ, chemical potential µ, hardness η, molarity S, and electrophilicity index ω. Besides these global descriptors of reactivity, we determined from the ACD/ Chemskech software the Log P or Octanol/Water partition coefficient of each of the studied molecules. This is an important physicochemical descriptor that can help predict the pharmacological activity of a compound because its transport, its passage through membranes and its pharmacological activity can be conditioned by its partition between a lipid phase and an aqueous phase.[15,16] The determination of molecular descriptors is an essential step in the establishment of the molecular model because they play a fundamental role in quantitative structure-activity relationship studies. They are the end result of a logical mathematical procedure that transforms the chemical information encoded in a representation of a molecule into a useful numerical value. [14] They are used as independent variables to predict a dependent variable (activity).

QSAR Model Development and Statistical Analysis

The development of the QSAR model from our pool of benzymidazolyl-chalcone molecules involved the use of the statistical analysis method of top-down MLR multiple linear regression which allows quantifying the relationship that exists between the biological activity under consideration and the structure (via descriptors) on the series of compounds in the training set. This method is implemented in the XLSTAT version 2017 software used for this purpose. Through this tool, the pool of studied molecules has been randomly subdivided into a training set (8 molecules) and a validation set (4 molecules). Subsequently, the MLR allowed the selection of descriptors that characterize the molecular structures of the compounds in the database in relation to the target activity in order to link them numerically to the experimental activity studied. The final choice of descriptors for the model is based on two fundamental criteria according to Vesserau.[17,18] The first criterion requires that there be a dependency between the activity being studied and the descriptors selected. This suggests that for each selected descriptor, $|R| \geq 0.5$, where R is the linear regression coefficient:[17-19]

$$R = \text{cov}(X,Y) / S_X.S_Y$$

$$a_{ij} = \text{cov}\ (X_i, Y_j)/\ \text{var}\ (X_i)$$

With cov (X,Y): Covariance of the two variables X and Y

S_X: Standard deviation of the variable X

S_Y: Standard deviation of the variable Y.

The second criterion indicates that the selected descriptors must be independent of each other. This is verified when aij< 0.70 where aij is the partial correlation coefficient between the pairs of descriptors i and j.[17,19]

$$a_{ij} = \text{cov}\ (X_i, Y_j)/\ \text{var}\ (X_i)$$

R and aij are computed by XLSTAT.

The QSAR model was established through the final equation taking into account the pre-selected descriptors which highlights the relationship between descriptors and data in the form of the following general equation.[15,20]

$$y = a_0 + \sum_{i=1}^{n} a_i x_i$$

With

y: biological activity studied

x_i: descriptors

a_0, a_i: corresponding regression coefficients of the statistical model.

Each regression coefficient must be significant with t < 0.05 calculated from a Student's t-test (t being the value calculated according to Student's t-test and 0.05 being the appropriate critical value in the Student's table).

Once developed, the quality of our obtained QSAR model was verified by a series of statistical tests grouped into two categories of criteria that are the internal validation criteria and the external validation criteria. The internal validation criteria reflect the degree of fit and robustness while the external validation criteria relate to the predictability of our model. Thus, the goodness of fit was evaluated by a set of quantitative statistical parameters which are the squared correlation coefficient R^2, the standard deviation s, the Fisher F coefficient.[21]

The R^2 squared correlation coefficient or coefficient of multiple determination is a quantitative measure of the precision of the adjustment for the values adjusted to those observed.

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

Where:

y_i is the experimentally obtained activity value for a compound in the training set.

$\hat{y}_i$ is the predicted (fitted) or calculated activity value for a compound in the training set.

$\bar{y}$ is the average of the experimental activities of the compounds in the data set to be examined.

R^2 informs about how well or poorly the model reproduces the experimental data. Also, R^2 should ideally be close to 1 (unity). Indeed, the closer it is to 1, the more similar the fitted values are to the experimental ones, which suggests that the model fits the data unerringly.[14,22,23]

The standard deviation s is given by the following relation:

$$s = \sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{n - p - 1}}$$

With

p representing the number of descriptors.

n is the number of molecules in the database to be examined.

s measures the dispersion of the observed values with respect to the regression line. The smaller s is, the better the correlation is.[15, 22]

Fisher's F coefficient is the statistical parameter that measures the level of statistical significance of the model at "x%".

$$F = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \hat{y}_i)^2} \times \frac{n - p - 1}{p}$$

The larger the F-value, the higher the probability that the equation is relevant. The equation is considered significant if the F-value is greater than the 95% tabulated value for a number of degrees of freedom (n-p-1).

As for the robustness of our constructed model, i.e., the influence of the training set compounds on the model, it was checked by determining the cross-validation squared correlation coefficient R^2_{cv} or Q^2:

$$Q^2 = 1 - \frac{\sum_i (y_i - \hat{y}_{i/i})^2}{\sum_i (y_i - \bar{y})^2}$$

Where

y_i is the experimentally obtained activity value for a compound in the training set from the cross-validation.

$\hat{y}_{i/i}$ is the predicted (fitted) or calculated activity value for a compound in the training set by excluding the i^{th} element in the model development.

$\bar{y}$ is the average of the experimental activities of the compounds in the training set. [15,22,24]

Thus, the QSAR model is considered "good" when $Q^2 \geq 0.5$ and "excellent" when $Q^2 \geq 0.9$.[23,25]

The ratio between R2 and Q2 was subsequently evaluated.

The difference $|R^2 - Q^2| < 0.3$ should be met because this standard indicates that the number of descriptors involved in the QSAR model is acceptable.[23,26]

As for the external validation of the constructed model, it consisted in predicting the activity of the series of molecules of the test set. We determined the parameter R^2 test, squared correlation coefficient determined for the validation set as well as other parameters known as "external validation criteria" or "Trophsa criteria", which meet the following standards:[27]

- $Q^2 > 0.5$
- $R^2_{test} > 0.6$
- or
- $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$

R_0^2 and R'^2_0represent the coefficients of determination when the regression line passes through zero for the graphs predicted values versus experimental values and experimental values versus predicted values respectively. As for k and k', they are respectively the slopes of these regression lines.[27]

Molecular Docking Methodology

The molecular docking procedure was used to investigate the binding interaction of the analogues in the binding pocket of the enzyme. The 3D crystal structure of the *Leishmania* phosphodiesterase B1 (PDB ID: 2JK6) was retrieved from the protein databank.[12] The protein preparation of Schrodinger suite was utilized to process the enzyme structure by assigning bond order, optimizing hydrogen bond k, creating zero bond order for metal and to form disulphide bond. The missing side chain and loops were filled using prime module while the water molecules beyond 5 Å were deleted. The structures of the top three bioactive benzimidazolyl-chalcones compounds were built in 2D sketcher of Schrodinger and prepared using ligprep. The grid was prepared using receptor grid generation as a centroid to search favorable binding between ligand and receptor molecule. The ligand in the active site was selected to generate the gird. The compounds were then docked into the active site of the target enzyme using standard precision (SP) mode of Glide.[28] The top-ranked conformation of each compound was used for further analysis.

MMGBSA Free Energy Calculation

Molecular mechanics generalized Born surface area (MM-GBSA) free binding energy of ligand-receptor complex has been calculated on the best SP docked poses using PRIME module of Schrödinger.

ADME Property Predictions

QikProp module of Schrodinger was used to predict the druggable property of three best hits by assessing the ADME profile. During this, the Lipinski rule of five and various descriptors like, QPPCaco, QPlogBB, MDCK, QlogS and % human oral absorption were calculated.

Molecular Dynamic Simulation Studies

A molecular dynamic (MD) simulation study was applied on the top hits for 30 ns obtained from the docking study and MM-GBSA calculation using Academic Desmond v6.5.[29] Three stages were done to perform this study: (a) System builder (b) Minimization (c) molecular dynamic simulation. For an SP docked complexes of 2JK6 and compound 4a-c, a predefined TIP3P solvent model was used to build a system model under orthorhombic boundary condition. A simulation study via NPT ensemble class at 300 K and 1 bar pressure was carried on for the minimized model.

Results and Discussion

Determined descriptors

The calculations of frequencies by the DFT B3LYP/6-31++G undertaken from these optimized structures, allowed to determine the descriptors reported in the Table 2.

The overall reactivity descriptors summarized in Table 3 were calculated from those in Table 2.

The results of the determination of the LogP of each of the studied molecules are reported in the Table 4.

Training set and validation set

The pool of molecules studied was subdivided into training and validation sets summarized in Table 5.

Construction of the QSAR model

To build the QSAR model, we first proceeded to select the descriptors involved in the final model based on Vessereau's criteria by calculating the linear regression and partial correlation coefficients between pairs of descriptors.

Thus, two descriptors meeting the stated criteria for our benzimidazolyl-chalcone series were selected. These are Log P and EA.

For these descriptors, R = 1.1664 and thus $|R| \geq 0.5$. Table 6 gives the values of the partial correlation coefficients.

The partial correlation coefficient between these two descriptors is less than 0.7. These descriptors are therefore independent.

Table 2: Descriptors obtained by the DFT B3LYP/6-31++G

COMPOUNDS	*ELECTRONIC ENERGY (kcal)*	*E_{HOMO} (ev)*	*E_{LUMO} (ev)*	*DIPOLE MOMENT μ (D)*	*POLARIZABILITY (u.a)*
4a	-791499.97	-6.64	-2.90	8.06	256.93
4b	-838706.02	-6.63	-2.89	9.27	262.94
4c	-1079906.56	-6.69	-3.00	8.50	267.15
4d	-919844.40	-6.78	-3.24	5.90	278.59
4e	-816185.98	-6.50	-2.79	2.96	281.45
4f	-1079908.73	-6.70	-3.03	6.38	277.44
4g	-1368314.58	-6.74	-3.13	6.93	287.64
4h	-790115.94	-6.39	-2.85	3.29	246.40
4i	-801572.24	-6.74	-3.07	0.59	254.26
4j	-853784.12	-6.71	-2.98	1.26	258.31
4k	-1135965.00	-6.93	-3.38	3.84	372.33
4l	-729275.33	-6.87	-3.16	7.93	329.57

Table 3: Descriptors calculated from the descriptors obtained by the DFT B3LYP/6-31++G

COMPOUNDS	η (ev)	S (ev)	μ	ω	χ (ev) according to Mulliken	ΔE	IE (ev)	EA (ev)
4a	1.86	0.53	-1.86	0.93	4.77	3.74	6.64	2.90
4b	1.87	0.53	-1.86	0.93	4.76	3.74	6.63	2.89
4c	1.84	0.54	-1.84	0.92	4.85	3.69	6.69	3.01
4d	1.77	0.56	-1.77	0.88	5.01	3.54	6.78	3.24
4e	1.85	0.53	-1.85	0.92	4.65	3.71	6.51	2.79
4f	1.83	0.54	-1.83	0.91	4.87	3.66	6.70	3.04
4g	1.81	0.55	-1.80	0.90	4.94	3.61	6.74	3.13
4h	1.77	0.56	-1.77	0.88	4.62	3.54	6.39	2.85
4i	1.83	0.54	-1.83	0.92	4.90	3.67	6.74	3.07
4j	1.86	0.53	-1.86	0.93	4.85	3.73	6.71	2.98
4k	1.77	0.56	-1.77	0.88	5.16	3.55	6.93	3.38
4l	1.85	0.54	-1.85	0.93	5.01	3.71	6.87	3.16

Table 4: Values of the partition coefficients of the studied molecules

Compounds	Log P
4a	4.52
4b	3.93
4c	5.11
4d	4.25
4e	4.98
4f	5.11
4g	5.72
4h	3.68
4i	3.28
4j	4.57
4k	5.27
4l	3.64

Table 5: Summary of the training and validation sets

Training Game	Validation Set
4a	4b
4c	4e
4d	4i
4g	4f
4h	
4j	
4l	
4k	

Table 6: Values of partial correlation coefficients between descriptors

	EA (ev)	Log P
EA (ev)	1	0.194
Log P	0.194	1

Table 7 summarizes the data of the selected descriptors and the activities of the molecules in the training and validation sets.

Finally, the QSAR model established from the anti-leishmanial descriptor-activity relationship of the training set of the studied benzimidazolyl-chalcones is given by the following equation:

$$\text{Log}\frac{1}{c} = 18.862 - 5.605 \times \text{AE}(ev) + 0.819 \times \text{Log}P$$

Validations of the QSAR model

Table 8 presents all the parameters calculated for the internal and external validation criteria for the developed QSAR model.

The parameters determined for the internal validation all meet the standards set for them. Thus, in view of these results, the equation of the QSAR model is significant (F being very high) and the constructed model presents a good agreement with the experimental data (R^2 close to 1). These observed values show a high correlation with the linear regression line (S small). As for the robustness of this model, it presents an excellent stability towards the molecules of the training set (Q^2_{CV} = 0.942) and the number of descriptors involved in the model is acceptable (R^2 - Q^2CV = 0.041). The QSAR model can therefore be validated with respect to internal validation.

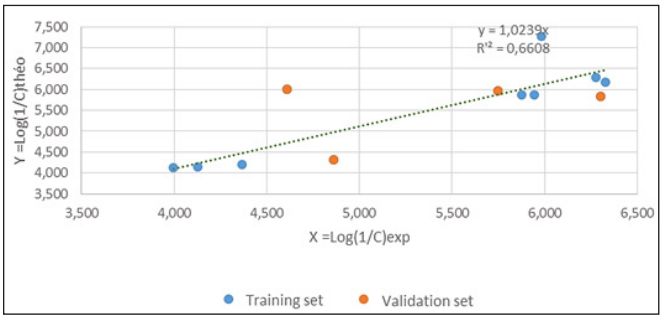


Figure 1: Linear regression curve of the theoretical data vs the experimental data of the benzimidazolyl-chalcone activities studied

Table 7: Molecular descriptors and activities of the training and test set

	Compounds	*IC_{50}(μM)*	*Log(1/C)*	*EA (ev)*	*Log P*
Training Set	4a	0.53	6.276	2.904	4.520
	4c	0.47	6.328	3.009	5.110
	4d	74.45	4.128	3.246	4.250
	4g	1.79	5.747	3.137	5.720
	4h	1.14	5.943	2.855	3.680
	4j	1.33	5.876	2.986	4.570
	4k	43.07	4.366	3.386	5.270
	4l	100.80	3.997	3.162	3.640
VALIDATION SET	4b	0.50	6.301	2.897	3.930
	4e	1.04	5.983	2.795	4.980
	4f	24.59	4.609	3.040	5.110
	4i	13.79	4.860	3.073	3.280

Table 8: Parameters determined for the internal and external validations of the developed QSAR model

QSAR model for the anti-leishmanial benzimidazolyl-chalcone series studied		
Validation criteria	*Calculated parameters*	*OECD standards*
Internal validation (training set N=8)	$R^2 = 0.983$	R^2 close to 1
	$S = 0.921$	S small
	$F = 146.668$	F high
	$Q^2_{CV} = 0.942$	$Q^2 \geq 0.9$
	$R^2 - Q^2_{CV} = 0.041$	$\lvert R^2 - Q^2 \rvert < 0.3$
External validation: Trospha criteria (test game N=4)	$Q^2 = 0.942$	$Q^2 > 0.5$
	$R^2_{test} = 2.57313632$	$R^2_{test} > 0.6$
	$\frac{R^2_{test} - R^2_0}{R^2_{test}} = 0.8$	$\frac{R^2_{test} - R^2_0}{R^2_{test}} < 0.1$
	$k' = 1.0239$	$0.85 \leq k' \leq 1.15$

Table 9: Ratio between theoretical and experimental values of biological activities in the validation set

Compounds	*$Log(1/C)_{theo}$*	*$Log(1/C)_{exp}$*	*$Log(1/C)_{theo}/Log(1/C)_{exp}$*
4b	5.843	6.301	0.9
4i	4.329	4.860	0.9
4e	7.276	5.983	1.2
4f	6.011	4.609	1.3

The values determined at the Trospha criteria level are supported by the data in Table IX and Figure 1.

This ratio tends to 1. This indicates a good correlation between the theoretical and experimental biological activities of the benzimidazolylchalcone series studied. The plot of predicted and experimental biological activities of the training and model validation sets is presented in Figure 1.

From the above regarding the Trospha criteria, the calculated parameters meet the set standards. According to Ouattara and Ziao, under these conditions, the QSAR model developed can be considered efficient in terms of activity prediction. [18]

Interpretation of the developed QSAR model

The developed QSAR model highlights the influence of two descriptors in the activities of the studied molecules: the partition coefficient and the electro affinity.

Thus, according to this QSAR model, the low values of the electronic affinity in absolute value coupled with high values of the partition coefficient should allow a better antileishmanial activity of these molecules.

However, it should be noted that these different descriptors are affected by positive and negative coefficients that impact the biological activity. Thus, a positive value of the coefficient indicates that an increase of the descriptor leads to an increase of the biological activity. Nevertheless, a negative value of the coefficient shows that an increase of the descriptor will rather lead to a decrease of the biological activity.

We note therefore, taking into account these different coefficients that the electronic affinity must be negative in this case to give rise to an increase in antileishmanial activity. This is possible for substituents which capture electrons such

as halogens, phenoxyl, nitro. [30,31] The partition coefficient must be positive for an increase in said activity. This suggests that the constituent groups of our antileishmanial molecules are lipophilic. The antileishmanial activity of the benzimidazolyl-chalcones submitted to our study is therefore related to their lipophilicity and electron-withdrawal capacity. This could be taken into account in the elucidation of the mechanism of action related to the antileishmanial activities of these molecules. Indeed, it could be assumed that the lipophilicity of these active constituents favors their penetration into the amastigote forms of leishmanias (pathogenic forms) whose plasma membrane is rich in phospholipids [32], in which they would exert their action. As for their electron affinity, it would confirm the importance of the basic heterocyclic nature of these molecules, which would be essential for the induction of the antileishmanial properties of benzimidazolylchalcones, as well as for the induction of their antiplasmodial properties. [13]

The Figure 2 reveals the contribution of each parameter of the developed QSAR model on the obtained biological activity.

It reveals that the contribution of electronic affinity is more important than that of lipophilicity. The electronic affinity is therefore the priority descriptor in the description of the antileishmanial activities of the benzimidazolylchalcones studied. This would confirm its importance in their induction.

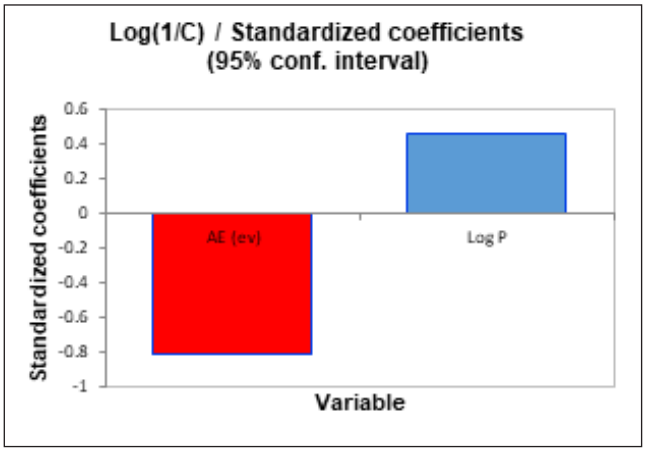


Figure 2: Contribution diagram of the different descriptors in the biological activity of the model

Thus, the benzimidazolylchalcones developed according to the QSAR model will have to present good basic and lipophilic properties to be active on *Leishmani donovani*.

Molecular Docking studies

The molecular docking was carried out to inspect the prospective interactions between the three most potent benzimidazolyl-chalcones and the active site of Leishmania phosphodiesterase B1. The binding energies of derivatives 4a-c were found to span from -6.50 Kcal/mol to -6.24 Kcal/mol at the binding site of 2JK6 indicate strong ligand receptor binding. Moreover, the docking results displayed that all the compounds were well accommodated in the active site of the enzyme as depicted in Figure 3. From the docking conformation of the most potent analogues, compound 4c (IC_{50} =0.47 µM) was observed that this compound formed one hydrogen bond with SER 464 and one pi-cation contact with LYS 61 of the binding pocket as displayed in Figure 3A. Moreover, 2-chlorophenyle atom moiety of compound 4c formed hydrophobic interactions with LEU 62, TYR 64 and LEU72. The strong bonding network of the compound with the residues of the active pocket might be one of the reasons to show excellent anti-leishmanial activity. The docking conformation of the second most active analogue 4b (IC_{50} = 0.50 µM) displayed one pi-cation contact with LYS61and two hydrogen bonds with SER and GLN 68 as shown in Figure 3 B. The docked conformation of the analogue 4a (IC_{50} =0.53 µM) exhibited one pi-cation contact with LYS 61 and one hydrogen bond with SER464 as presented in Figure 3C.

MMGBSA free energy calculation

Binding free energy was calculated using prime MMGBSA module of Schrodinger to assess the stability of the ligand-receptor complex. These MM-GBSA calculations consider the geometry of the ligand-receptor complex with the values of free ligand. More negative binding affinity value corresponds to stronger ligand-receptor complex. Compound 4c interacted with the receptor with binding energy -61.27 kcal/mol followed by 4b (-57.65 kcal/mol) and 4a (-56.22 kcal/mol) receptors (Table 5). These displayed good correlation with experimental activity values.

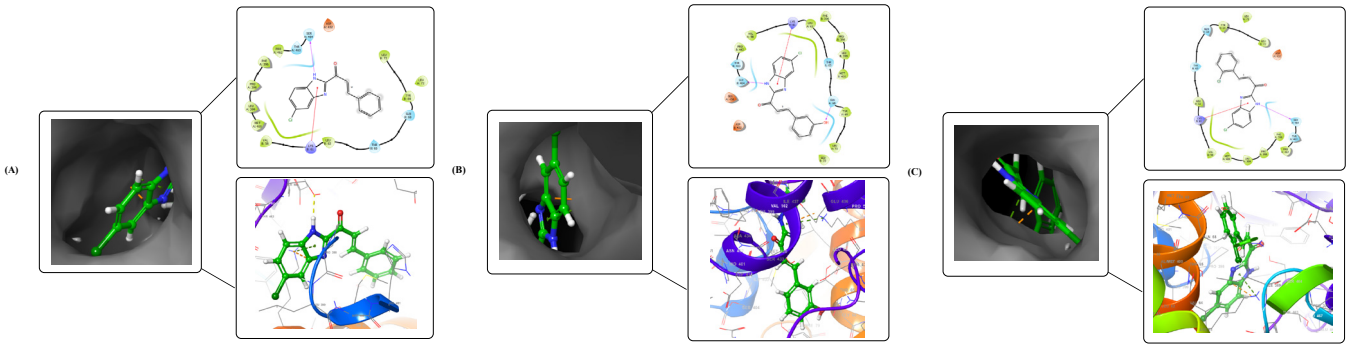


Figure 3: Molecular docking of receptor-binding domain of Leishmania phosphodiesterase B1 (PDB ID: 2JK6) with A) 4a, B) 4b and C) 4c

Figure 4: (A) protein-ligand root-mean-square deviation (RMSD) (B) protein root-mean-square fluctuation (RMSF) (C) protein-ligand contacts diagram of compound 4b seen during MD simulations

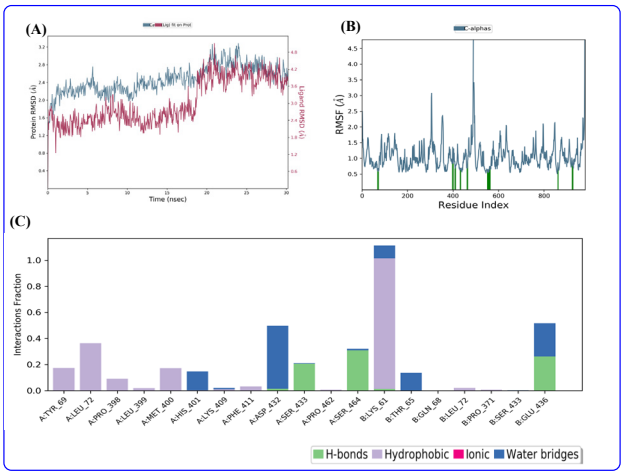


Figure 5: (A) protein-ligand root-mean-square deviation (RMSD) (B) protein root-mean-square fluctuation (RMSF) (C) protein-ligand contacts diagram of compound 4c seen during MD simulations

Table 10: Docking, MMGBSA score and ADME properties of the top 3 compounds.

Molecules	*Docking Score (kcal/mol)*	*MMGBSA dG bind (kcal/mol)*	*QPPCaco*[a]	*QPlogBB*[b]	*MDCK*[c]	*% Human oral absorption*[d]	*QPlog S*[e]	*Rule of five*
4a	-6.246	-56.22	1510.44	-0.334	1904.14	100	-4.61	0
4b	-6.505	-57.65	459.55	-0.939	526.39	91.27	-4.34	0
4c	-6.400	-61.27	1542.99	-0.183	4132.53	100	-5.17	0

[a] Predicted caco cell permeability in nm/s (acceptable range: <25 is poor and >500 is great). Caco-2 cells are a model for the gut-blood barrier.
[b] Predicted blood brain barrier permeability (acceptable range -3–1.2).
[c] Predicted apparent MDCK cell permeability in nm/s (acceptable range in nm/s (acceptable range: <25 is poor and >500 is great). MDCK cells are a good model for the blood–brain barrier
[d] Percentage of human oral absorption (acceptable range: <25 is poor and >80% is high.
[e] Predicted aqueous solubility in mol/L (acceptable range -6.5 –0.5).

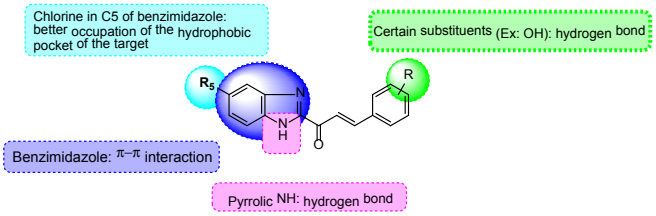


Figure 6: Summary of docking studies on the target: Leishmania phosphodiesterase B1 (PDB ID: 2JK6)

ADME property predictions

Some important Computer-Aided Prediction of Pharmacokinetic (ADME) Properties for the three compounds were evaluated using QikProp module of Schrodinger. Many basic physicochemical properties of these compounds were predicted. The values of those compounds were found in the recommended range as depicted in Table X. None of the three molecules violated any rule of five.

In addition, these molecules have acceptable aqueous solubility and high gastrointestinal absorption (≥ 90%). This predicts good oral bioavailability for these compounds.

Molecular dynamic simulation

The dynamic behavior of compound at the binding cavity of the enzyme is crucial to evaluate the stability of that particular compound inside the binding site. A 30 ns molecular dynamics simulations of 2JK6 and compounds **4b** and **4c** complexes gives further insights into molecular interaction of these compounds in motion.

RMSD analysis of the 2JK6 protein and top two screened leads complexes revealed comparable deviation throughout the 30 ns long molecular dynamic simulations. RMSDs for 4b-2JK6 stabilized around 2Å for 18 ns and increase thereafter to 2.8Å, however, it remained within the acceptable range (≤ 3Å) as shown in Figure 4A. The RMSD plot for 4c-2JK6 (Figure 5 A) showed that the protein Cα stabilized during simulation after 18 ns at 2.8Å. During initial 18 ns, ligand RMSD value was 1.8 Å indicating its high stability compared to protein RMSD during this period. The overall RMSD of the two compounds in the protein in the binding region is stable. The per residue RMSF analysis of the complexes exhibited comparatively similar fluctuation patterns and did not display major fluctuations as depicted in Figures 4B and 5B. The detailed interaction fraction of the hydrogen

bonds, hydrophobic, water bridge interactions of 4b and 4c with 2JK6 is elaborated in Figure 4C and 5C. Hydrogen bonds have a significant contribution on the stability of 4b-2JK6 than 4c-2JK6.

The structure-activity relationship studies undertaken in previous work established that the presence of halogen at C5 on benzimidazole improves potency *in vitro* antileishmanial activities.[9] Furthermore, the C2 propenone chain of benzimidazole would be necessary for the appearance of activities. In addition, the presence of an ortho chloro substituent on the C_3 phenyl of propenone, improves the antileishmanial activity.

The Qsar studies revealed that the presence of electrophilic substituents such as halogens (Cl), hydroxyl or nitro on the benzimidazolyl-chalcone would enhance the negativity of the electron affinity of the molecules which would contribute to increase their antileishmanial activities. In addition, the increase in overall lipophilicity would increase the partition coefficient of the molecules which would be beneficial to the antileishmanial activities in this series.

Molecular docking studies performed on the three best molecules (4a, 4b and 4c) suggest that the excellent antileishmanial activities would be promoted by different interactions involving the benzimidazole ring. Indeed, the best compound 4c formed 2 key interactions at the active site namely a π-π interaction established through the benzimidazole ring at LYS B 61 and a hydrogen bonding through the pyrrolic nitrogen of this heterocycle at SER A: 464. Moreover, its chlorine atom in C5 of the benzimidazole allows a better occupation of the hydrophobic pocket. Another interaction by hydrogen bonding thanks to the phenolic OH of the phenylpropenone chain at the level of GLN A 68 probably increases the residence time of certain derivatives such as 4b in the active site. However, this last interaction would have only a marginal role on the global activity of the molecules of this series (Figure 6). In addition, compound 4c exhibited the most favorable binding free energy (-61.27 kcal/mol) in the *Leishmania phosphodiesterase* B1 binding site compared to the other derivatives (Table 10).

Thus, the QSAR, docking and other molecular dynamics studies performed in the present study partially corroborated the results of the SAR studies and provided valuable information for the development of benzimidazolyl-chalcone derivatives against leishmaniasis.

Conclusion

In this study, the aim was to build a QSAR model around a series of benzimidazolyl-chalcone compounds with anti-leishmanial targets. In addition, the molecular docking, MMGBSA, and molecular dynamics studies were carried out on compounds with the best activities allowed us to highlight the strength of the interactions of the substituents with the molecular target as well as their stability. Pharmacokinetic properties analysis of the selected compounds revealed the drug-like property of the lead compounds. Our results suggested lipophilicity and electron affinity as predictors of the activity of these benzimidazolyl-chalcone derivatives against *Leishmania donovani*. Moreover, the control of the electronic affinity, the basic properties of this series, appears to be essential in the exaltation of the antileishmanial activities. This work is consistent with the SAR made in our previous work and gives us an orientation for the design of new analogues more active on *Leishmania donovani*. These will have to present in their respective structures key groups such as chlorine, phenyl or phenoxyl to have an optimal activity. These findings play an important role in understanding the relationship between the physicochemical parameters of the structure and the biological activity. The analysis and use of the QSAR model coupled to the information given by the ligands-protein interactions and the stability of the interactions through molecular docking, MM-GBSA, molecular dynamics, enabled to select the appropriate substituent and design new compounds with improved biological activity. Finally, molecular dynamics simulation studies at 30 ns display stable interactions with the protein.

References

1. Aponte JC, Castillo D, Estevez Y, Gonzalez G, Arevalo J, Hammond GB, et al. *In vitro* and *in vivo* anti-Leishmania activity of polysubstituted synthetic chalcones. *Bioorg Med Chem Lett.* 2010;20(1):100-103.
2. Elmagd ARAA. Synthesis and Medicinal Significance of Chalcones-A Review. *Asian J Biomed Pharm Sci.* 2016;6(52):1-7.
3. Mahama O, Aboudramane K, Soleymane K, Sylvain C, Sekou D, Drissa S. Anticancer Activities and QSAR Study of Novel Agents with a Chemical Profile of Benzimidazolyl-Retrochalcone. *Open J Med Chem.* 2020;10(03):113-127.
4. Athanasios V and Thomais V. Studies in Natural Products Chemistry: Chapter 8. Plant Polyphenols: Recent Advances in Epidemiological Research and Other Studies on Cancer Prevention. Edited by Atta-ur-Rahman. 2013;39: 269-295.
5. Ouattara M, Sissouma D, Yavo W, Kone M. Synthèse et criblage antiplasmodial de quelques benzimidazolyl-chalcones. I*nt J Biol Chem Sci.* 2015;9(3):1697-1710.
6. N'Ta CA. Synthèse d'analogues de la pentamidine porteurs de plateformes hétérocycliques (rhodanine, benzimidazole, pyrazole et imidazole) et leurs évaluations biologiques. Université Rennes 1; Université Nangui Abrogoua, 2015. (NNT: 2015REN1S137). (tel-01665809)
7. Nieto-Meneses R, Castillo R, Hernández-Campos A, Maldonado-Rangel A, Matius-Ruiz JB, Trejo-Soto PJ, et al. *In vitro* activity of new *N*-benzyl-1*H*-benzimidazol-2-amine derivatives against cutaneous, mucocutaneous and visceral Leishmania species. *Exp Parasitol.* 2018;184:82-89.
8. Nagle AS, Khare S, Kumar AB, Supek F, Buchynskyy A, Mathison CJN, et al. Recent Developments in Drug Discovery for Leishmaniasis and Human African Trypanosomiasis. *Chem Rev.* 2014;114(22):11305-11347.

9. N'Guessan DUJP, Kablan LAC, Kacou A, Bories C, Coulibaly S, Sissouma D, Loiseau PM, Ouattara M. Synthesis and biological profiles of some benzimidazolyl-chalcones as anti-leishmanial and trypanocidal agents. *Chemical science international journal.* 2021; 30(8): 47-56.
10. Luan F, Cordeiro MNDS. Overview of QSAR Modeling in Rational Drug Design. in Recent Trends on QSAR in the Pharmaceutical Perceptions. *Bentham science publishers*; 2012;(48):194-241.
11. Ochoa R, García E, Robledo SM, Cardona W G. Virtual and experimental screening of phenylfuranchalcones as potential anti-Leishmania candidates, *Journal of Molecular Graphics and Modeling*. 2019;(91):164-171
12. QikProp, version 3. 6, 2013. Schrödinger, LLC, New York.
13. Ouattara M, Sissouma D, Yavo W, Kone M. Synthèse et criblage antiplasmodial de quelques benzimidazolyl-chalcones. *Int J Biol Chem Sci.* 2015;9(3): 1697-1710.
14. Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design--a review. *Curr Top Med Chem.* 2010;10(1):95-115.
15. Chtita S, Bouachrine M, Tahar L. Modélisation de molécules organiques hétérocycliques biologiquement actives par des méthodes QSAR/QSPR. Recherche de nouveaux médicaments. Chimie théorique et/ou physique. Faculté des sciences,Université Moulay Ismaïl, Meknès, 2017. (tel-01568788)
16. Babu NS, Jayaprakash D. Global and Reactivity Descriptors Studies of Cyanuric Acid Tautomers in Different Solvents by using of Density Functional Theory (DFT). *Int J Sci Res IJSR*. 2015;4(6):615-620.
17. Vessereau A. Méthodes statistiques en biologie et en agronomie. Paris: Lavoisier; 1992.
18. Ouattara O, Ziao N. Quantum Chemistry Prediction of Molecular Lipophilicity Using Semi-Empirical AM1 and Ab Initio HF/6-311++G Levels. *Computational Chemistry*. 2017, 05(1):38-50
19. Weisberg S. Applied Linear Regression, Third Edition. Published by John Wiley & Sons, Inc., Hoboken, New Jersey. 2005; 1- 292.
20. Roy K, Kar S, Das RN. Statistical Methods in QSAR/QSPR. In: A Primer on QSAR/QSPR Modeling. *Cham: Springer International Publishing*. 2015;37-59.
21. Tareq HKM. Overview of QSAR Modeling in Rational Drug Design. In: Recent Trends on QSAR in the Pharmaceutical Perceptions. *Bentham science publishers*. 2012; 194-241.
22. Organization for Economic Cooperation and Development (OECD), Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment, n° 69, Éditions OCDE, Paris, Environment Health and Safety Publications, 2014. https://doi.org/10.1787/9789264085442-en.
23. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs. *Environ Health Perspect.* 2003;111(10):1361-75.
24. Prana V. Approches structure-propriété pour la prédiction des propriétés physico-chimiques des substances chimiques. Thèse de doctorat en Chimie théorique et informatique, Pierre et Marie Curie; 2013, 1- 208.
25. Tichý M, Rucki M. Validation of QSAR models for legislative purposes. Validation of QSAR models for legislative purposes. *Interdisciplinary Toxicology.* 2009;2(3):184-186
26. Ravichandran V, Harish R, Abhishek J, Shalini S, Christapher PV, Agrawal RK. Validation of QSAR Models - Strategies and Importance. *Int J Drug Des Discov*. 2011;2(3): 511-519.
27. Golbraikh A, Tropsha A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J Comput Aided Mol Des*. 2002;16(5-6):357-369.
28. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shaw DE, Shelley M, Perry JK, Francis P, Shenkin PS. "Glide: A New Approach for Rapid, Accurate Docking and Scoring. Method and Assessment of Docking Accuracy," *J. Med. Chem.*, 2004; 47: 1739–1749.
29. Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021.
30. Lide DR. CRC handbook of chemistry and physics. CRC Press; 1995.
31. Kotz JC, Treichel P, Townsend JR. Chemistry & chemical reactivity. 7th ed. Belmont, CA: Thomson Brooks/Cole. 2009. 1095 p.
32. Filippi C, Malherbe L, Julia V, Glaichenhaus N. L'immunité contre les leishmanies. *Médecine Sci.* 2001;17(11):1120-1128.