MDPI

Article

QSAR Study of *N*-Myristoyltransferase Inhibitors of Antimalarial Agents

**Letícia Santos-Garcia [1], Marco Antônio de Mecenas Filho [1], Kamil Musilek [2,3],
Kamil Kuca [2,*], Teodorico Castro Ramalho [1,4] and Elaine Fontes Ferreira da Cunha [1,*]**

[1] Departamento de Química, Universidade Federal de Lavras, Lavras 37200-000, Brazil;
    leticiasantosg@hotmail.com (L.S.-G.); marcomecenasfilho@gmail.com (M.A.d.M.F.); teo@dqi.ufla.br (T.C.R.)
[2] Department of Chemistry, Faculty of Science, University of Hradec Kralove,
    Hradec Kralove 50005, Czech Republic; kamil.musilek@uhk.cz
[3] Biomedical Research Center, University Hospital Hradec Kralove, Hradec Kralove 50005, Czech Republic
[4] Center for Basic and Applied Research, University Hradec Kralove, Hradec Kralove 50005, Czech Republic
*   Correspondence: kamil.kuca@uhk.cz (K.K.); elaine_cunha@dqi.ufla.br (E.F.F.d.C.);
    Tel.: +420-603-289-166 (K.K.); +55-35-38291271 (E.F.F.d.C.)

check for updates

Abstract: Malaria is a disease caused by protozoan parasites of the genus *Plasmodium* that affects millions of people worldwide. In recent years there have been parasite resistances to several drugs, including the first-line antimalarial treatment. With the aim of proposing new drugs candidates for the treatment of disease, Quantitative Structure–Activity Relationship (QSAR) methodology was applied to 83 *N*-myristoyltransferase inhibitors, synthesized by Leatherbarrow et al. The QSAR models were developed using 63 compounds, the training set, and externally validated using 20 compounds, the test set. Ten different alignments for the two test sets were tested and the models were generated by the technique that combines genetic algorithms and partial least squares. The best model shows $r^2 = 0.757$, $q^2_{adjusted} = 0.634$, $R^2_{pred} = 0.746$, $R^2_m = 0.716$, $\Delta R^2_m = 0.133$, $R^2_p = 0.609$, and $R^2_r = 0.110$. This work suggested a good correlation with the experimental results and allows the design of new potent *N*-myristoyltransferase inhibitors.

Keywords: malaria; *N*-myristoyltransferase; drug development; QSAR; mosquito-borne protozoal infection

1. Introduction

Malaria is a mosquito-borne protozoal infection caused by five parasites of the genus *Plasmodium*: *P. falciparum*, *P. ovale*, *P. malariae*, *P. vivax*, and *P. knowlesi* [1]. Among these, *Plasmodium falciparum* is the most prevalent and lethal [2]. Over the past 50 years, the parasite resistance to chloroquine and sulphadoxine-pyrimethamine in endemic countries has been noted [3,4]. In addition, signs of resistance to artemisinin-based combination therapies (ACTs) have been detected. Actually, ACTs are the first-line treatment for malaria and thus, new drugs are constantly required [5–8]. Researchers have identified the *N*-myristoyltransferase (NMT) enzyme as an important target for a generation of drugs to be used for the treatment of malaria. NMT catalyzes the transfer of the myristoyl group from a myristoyl coenzyme A (CoA) to the N-terminal glycine residue after the targeted protein [9]. N-terminal myristoylation (MYR) by NMT occurs through the Bi-Bi mechanism, where MYR-CoA binds to the apo-enzyme, inducing a conformational change that allows the NMT substrate to bind [5,10]. *P. falciparum* has a single NMT isoform and mRNA is expressed in asexual blood-stage forms [11]. The first reported NMT inhibitors were obtained by mimicking the structure of peptide substrates (Figure 1A) [12] or by designing non hydrolysable, methylene-bridged analogue of myristoyl coenzyme

A (Figure 1B) [13]. After that, inhibitors based on a quinolone scaffold and furan core were reported (Figure 1C,D) [14]. In this work, we used dimensional quantitative structure-activity relationship analysis of 83 NMT inhibitors based on a phenyl scaffold [15] seeking to propose new candidates for NMT inhibitors. Furthermore, a physicochemical properties evaluation was performed in order to find the most appropriate compound predicted.
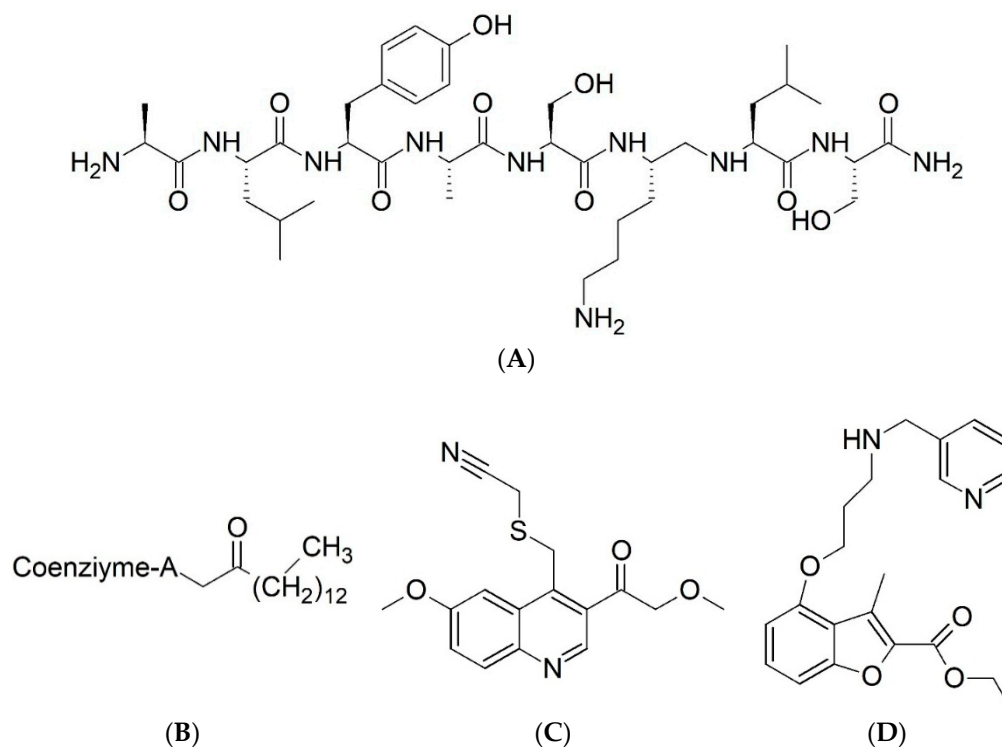

Figure 1. (**A**) Inhibitor mimicking the structure of substrates (*Ki* = 5.0, 8.0, and 35.0 μM for *S. cerevisiae*, *C. albicans*, and human NMT) [12]; (**B**) inhibitor methylene-bridged analogue of myristoyl coenzyme A (*Ki* = 24.0 nM) [13]; (**C**) inhibitor based on a quinolone scaffold (*Ki* = 4.7, and >100 μM for *Plasmodium vivax* and *Plasmodium falciparum NMT*, respectively); and (**D**) inhibitor based on a furan core RO-09-4609 (IC$_{50}$ = 0.1 and >540 μM for *C. albicans*, and human NMT) [14].

2. Results

The GA-PLS analysis using grid cells of 1.0 A generated several models or equations. The statistical parameters of ten alignments studied for Test Set I (compounds **1, 3, 5, 6, 12, 16, 20, 30, 33, 39, 40, 50, 56, 57, 61, 65, 66, 69, 76,** and **80**) and Test Set II (compounds **3, 6, 9, 13, 20, 21, 27, 28, 31, 32, 40, 56, 57, 58, 64, 70, 73, 76, 78,** and **82**) are shown in Tables 1 and 2, respectively. All tested alignments showed q^2 values higher than 0.5. This reveals that the model can be a useful tool for predicting affinities of new compounds based on these structures; r^2 greater than 0.7 indicates that the model is correlated and may be considered to represent the training set in the same manner [16]. Alignments 6B and 7B were eliminated from the analysis because it presented a low r^2 value (<0.7).

Table 1. Statistical parameters evaluated in the 4D-QSAR analysis for the ten performed alignments of Test I.

Alignment	r^2	$RMSE_C$	q^2_{adj}	$RMSE_{CV}$	R^2_{Pred}	$RMSE_P$	r^2_m	R^2_r	R^2_p
A1	0.746	0.481	0.607	0.549	0.532	0.65	0.71	0.312	0.82
A2	0.744	0.478	0.608	0.548	0.548	0.663	0.692	0.343	0.799
A3	0.761	0.469	0.609	0.546	0.508	0.702	0.735	0.182	0.994
A4	0.708	0.508	0.576	0.579	0.588	0.595	0.645	0.287	0.825
A5	0.736	0.511	0.589	0.566	0.477	0.698	0.766	0.245	0.895
A6	0.739	0.477	0.582	0.563	0.567	0.637	0.67	0.286	0.83
A7	0.722	0.503	0.584	0.571	0.555	0.656	0.683	0.291	0.831
A8	0.746	0.445	0.605	0.551	0.62	0.59	0.606	0.216	0.891
A9	0.734	0.491	0.578	0.566	0.547	0.684	0.693	0.25	0.861
A10	0.723	0.519	0.583	0.572	0.503	0.676	0.74	0.311	0.816

r^2: Coefficient of determination; $RMSE_c$: root mean square deviation of calibration; q^2_{adj}: adjusted cross-validated squared correlation coefficient; $RMSE_{cv}$: root mean square deviation of cross validation; R^2_{pred}: correlation coefficient of external validation; RMSEp: root mean square deviation of prediction; r^2_m(test): Equation (3) (Materials and Methods); R^2_r: Y-randomization; R^2_p: Equation (2) (Materials and Methods).

Table 2. Statistical parameters evaluated in the 4D-QSAR analysis for the ten performed alignments of Test II.

Alignment	r^2	$RMSE_C$	q^2_{adj}	$RMSE_{CV}$	R^2_{Pred}	$RMSE_P$	R^2_m	R^2_r	R^2_p
B1	0.728	0.504	0.617	0.544	0.728	0.532	0.688	0.301	0.476
B2	0.728	0.515	0.607	0.553	0.763	0.496	0.749	0.289	0.482
B3	0.757	0.472	0.634	0.527	0.746	0.515	0.716	0.11	0.609
B4	0.704	0.549	0.585	0.573	0.782	0.476	0.765	0.253	0.473
B5	0.725	0.5	0.601	0.55	0.706	0.553	0.692	0.198	0.526
B6	0.692	0.559	0.576	0.581	0.771	0.489	0.755	0.272	0.448
B7	0.69	0.556	0.581	0.577	0.751	0.509	0.735	0.289	0.437
B8	0.73	0.514	0.6	0.55	0.77	0.489	0.75	0.209	0.527
B9	0.723	0.528	0.605	0.555	0.786	0.472	0.773	0.229	0.508
B10	0.744	0.501	0.619	0.542	0.779	0.48	0.744	0.289	0.502

r^2: coefficient of determination; $RMSE_c$: root mean square deviation of calibration; q^2_{adj}: adjusted cross-validated squared correlation coefficient; $RMSE_{cv}$: root mean square deviation of cross validation; R^2_{pred}: correlation coefficient of external validation; RMSEp: root mean square deviation of prediction; r^2_m(test): Equation (3) (Materials and Methods); R^2_r: Y-randomization; R^2_p: Equation (2) (Materials and Methods).

Evaluating the predictive ability in terms of R^2_{pred} (means of an external validation), the alignment 5A was eliminated ($R^2_{pred} < 0.5$). All R^2_m values were greater than 0.60, and values over 0.5 are acceptable. Analyzing the R^2_p values, alignments B1, B2, B4, B6, and B7 were excluded because this parameter value was less than 0.5 [17].

Alignment 3 from Test Set II (B3) provides the best 4D-QSAR models as judged by the highest q^2_{adj}, in addition to presenting fewer descriptors. Among the alignments with only seven descriptors, this still has the highest values of r^2, q^2_{adj}, R^2_p, and the lowest value of R^2_r. According to these results, we selected Model B3 as the best alignment. We will only present the analysis of the best model derived from B3.

The statistical measures, including the values of r^2, q^2, q^2_{adj}, LSE, LOF, $RMSE_C$, $RMSE_{CV}$, $RMSE_P$, Y-Rand, R^2_{pred}, R^2_m, R^2_p, and R^2_r are presented below. Each GCOD (grid cell occupancy descriptors) is labeled as "x, y, z, IPE" which represent the cartesian coordinate positions of the selected grid cell (x, y, z) and the respective atom type (interaction pharmacophore elements, IPE): (i) any type (any); (ii) nonpolar (np); (iii) polar-positive charge density (p+); (iv) polar-negative charge density (p−); (v) hydrogen bond acceptor (hba); (vi) hydrogen bond donor (hbd); and (vii) aromatic systems (ar).

Model B3

pIC_{50} = 3.997 + 4.942(0,−3,−1, hba) + 2.345(0,−5,−1, any) + 2.100(0,−1,0, any) + 1.692(0,3,−3, any) + 1.191(−1,−4,−3, any) - 8.269(−1,−4,−4, np)

n = 63, GCODs = 7, r^2 = 0.757, q^2 = 0.702, q^2_{adj} = 0.634, LSE = 0.233, LOF = 0.418, $RMSE_C$ = 0.472, $RMSE_{CV}$ = 0.527, $RMSE_P$ = 0.515, $RMSE_{cy-rand}$ = 1.055, R^2_{pred} = 0.746, R^2_m = 0.716, R^2_p = 0.609, and R^2_r = 0.110.

Another different variant of R^2_m metrics was calculated from Model B3 to assess the predictive ability of the test set, ΔR^2_m. The value of ΔR^2_m found was 0.133. It has been suggested that to be considered a predictive model, this value should be less than 0.2 [18]. Model B3 generated seven descriptors, where GCODs (−1,−4,−3, any), (0−0, any), (0,6,2, any), (0,−5,−1, any), (0,3,−3, any), and (0,−3, −1, hba) present positive coefficients (Equation (3)) and correspond to favorable interactions between the molecule substituent and amino acid residues in the active site of NMT. Therefore, substituents in these positions increase the effectiveness of the compounds. The GCOD (−1,−4,−4, np) has negative coefficient and correspond to unfavorable interactions between the molecule substituent and amino acid residues in the active site of NMT. Therefore, the occupation of GCOD (−1,−4,−4, np) decreases the compound potency.

3. Discussion

GCODs are related to the coordinates of IPE mapped in a common grid. A graphic representation of the descriptors of Model B3 is shown in Figure 2 using Compound **81** as a reference. Light and dark spheres represent GCODs with positive and negative coefficients, respectively, in accordance with Model B3. GCOD-1 (0,−3,−1, hba) (Figure 3) is the descriptor that most contributes to the increased effectiveness of compounds and presents a coefficient of 4.942. This grid cell represents an acceptor hydrogen bond atom type (IPE) and shows high frequency of occupation for compounds **42**, **48**, **65**, **68**, and **69**. It is located close to the nitrogen atom of the oxadiazole ring and indicates an amino acid donor hydrogen bond in *N*-myristoyltransferase.
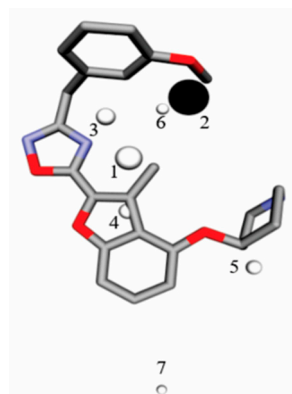
Figure 2. Graphic representation of Compound **81** according to the 4D-QSAR Model B3. GCODs occupancy represented by white spheres contributes to increasing the potency of compounds, and black spheres to decrease the potency of the compounds. The GCODs described are: (1) (0,−3,−1, hba), (2) (−1,−4,−4, np), (3) (0,−5,−1, any) (4) (0,−1,0, any), (5) (0,3,−3, any) (6) (−1,−4,−3, any), and (7) (0,6,2, any). The gray and red representations are carbons and oxygen atoms.

The oxadiazole ring in the ortho position allows the nitrogen atom to occupy this grid cell, exemplified by compound **42** (Figure 3). However, the most active molecules of the training set, **81** and **83**, do not have this descriptor that contributes most to the increase in the potency of the compounds. Once in this position there is a methyl group and the oxadiazole group that are displaced. In fact, the oxadiazole ring in these compounds does not occupy this grid cell.
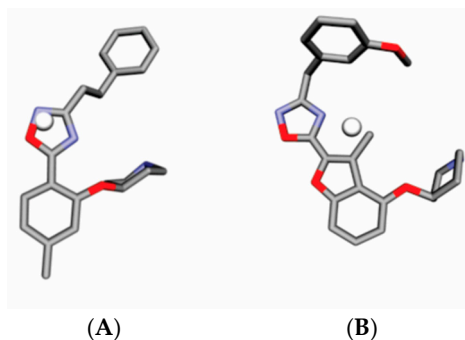
Figure 3. (**A**) Representation of compound **42** and GCOD-1 (0,−3,−1, hba) (white sphere), and (**B**) compound **81**. GCODs occupancy represented by white spheres contributes to increasing the potency of compounds. The gray and red representations are carbons and oxygen atoms.

GCOD-2 (−1,−4,−4, np) (Figure 4) contributes to decrease compound potency and presents a coefficient of −8.269. This grid cell corresponds to a nonpolar IPE and shows high occupation frequency for Compounds **42**, **48**, and **55**. These molecules present non-polar groups in this local, such as ethyl and methyl groups. Thus, the occupation of this cell is drastically reduced when this position is non-polar substituted, that decrease the activity of these compounds. Meanwhile, if the polar groups, such as OH or N in **81** or **83**, respectively, are localized in this grid cell, the GCOD-2 descriptor yields less negative value which does not decrease the predicted pIC$_{50}$. We can see that compound **42** occupies the descriptor that decreases the activity. However, it also presents occupancy for GCOD-1, which increases the activity. As expected, this compound exhibits an intermediate power.
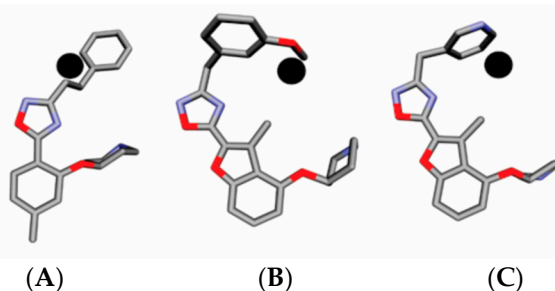
Figure 4. Representation of GCOD-2 (−1,−4,−4, np) (black sphere) and (**A**) compound **42**, (**B**) compound **81**, and (**C**) compound **83**. CODs occupancy represented by black spheres to decrease the potency of the compounds. The gray and red representations are carbons and oxygen atoms.

This suggests that the occupation of this cell by acceptor hydrogen bond should be favorable in contrast to nonpolar atoms that are unable to perform hydrogen bond interactions.

GCOD-3 (0,−5,−1, any) is present as a non-specific IPE. The atom type "any" is used when more than one specific atom [IPE] type across the training set is found to satisfy the interaction being captured by a particular GCOD [19]. GCOD (0,−5,−1, any) has a positive coefficient of 2.345 which increases the potency of the compounds (Figure 5). This grid cell is close to methyl and ethyl groups, toward the left side, and shows greater occupation frequency for compounds **26**, **30**, and **33**. On the other hand, compounds **42** and **45** showed no occupancy for this descriptor, because during the molecular dynamics simulation, they assumed a different conformation, facing right. On the other hand, GCOD (0,−5,−1, any) also provides occupation frequency in the benzene ring of molecules **59**, **60**, **61**, and **62**.
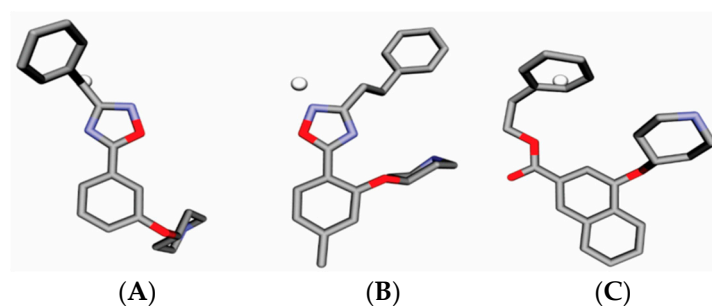
Figure 5. Representation of GCOD-3 (0,−5,−1, any) (white sphere) and (**A**) compound **30**, (**B**) compound **42**, and (**C**) compound **60**. CODs occupancy represented by white spheres contributes to increasing the potency of compounds. The gray and red representations are carbons and oxygen atoms.

GCOD-4 (0,−1,0, any) have a positive coefficient and, thus, also greatly influence the increase in inhibitor potency (Figure 6). It is located near the methyl group in benzofuran, 2,3-dihidro-3-methyl and represents a non-specific IPE. It shows the highest occupation frequency for the most active compounds in this series, compounds **1**, **81**, and **83**. The GCOD (0,−3,−1, hba) reflects the importance of occupation of this receptor region for the effectiveness of the inhibitors.
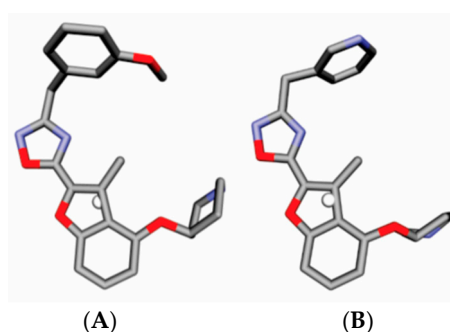
Figure 6. Representation of GCOD-4 (0,−1,0, any) (white sphere) and (**A**) compound **81**, and (**B**) compound **83**. CODs occupancy represented by white spheres contributes to increasing the potency of compounds, and the gray and red representations are carbons and oxygen atoms.

GCOD-5 (0,3,−3, any) have a positive coefficient, and so, improves the effectiveness of inhibitors (Figure 7). The GCOD (0,3,−3, any) is situated near the piperidine ring and has occupancy for a large majority of molecules, such as compounds **65** and **79**. The change of position of the piperidine ring in compounds **15**, **16**, and **30** is not favorable, so these do not have this GCOD.
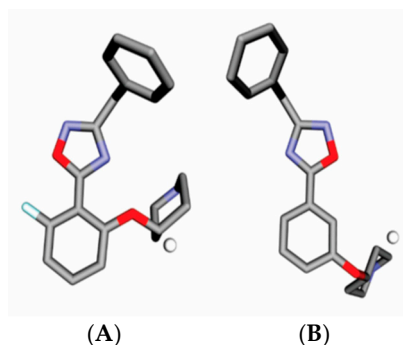
Figure 7. Representation of GCOD-5 (0,3,−3, any) (white sphere) and (**A**) compound **79**, and (**B**) compound **30**. CODs occupancy represented by white spheres contributes to increasing the potency of compounds, and the gray and red representations are carbons and oxygen atoms.

GCOD-6 (−1,−4,−3, any) represents a non-specific IPE and also has a positive coefficient, indicating an increase in potency of compounds which have high occupation frequency for this descriptor (Figure 8). The molecules **14**, **19**, and **22** have a high occupation frequency for this GCOD, located near the benzene group. This grid cell suggests a hydrophobic region in the receptor close to the benzene ring, which should be making a π–π staking interaction between the aromatic ring of an inhibitor and one aromatic amino acid residue.

Lastly, GCOD-7 (0,6,2, any) (Figure 9) has a positive coefficient and shows a non-specific class. Molecules **59–63** have a high occupation frequency for this descriptor. The presence of naphthalene group in this grid cell increase the activity. In fact, it shows that the aromatic substituents in this position should be preferred.

(**A**)   (**B**)   (**C**)
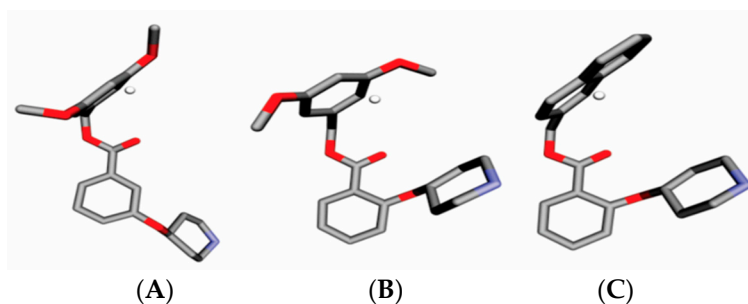
Figure 8. Representation of GCOD-6 (−1,−4,−3, any) (white sphere) and (**A**) compound **14**, (**B**) compound **19**, and (**C**) compound **22**. CODs occupancy represented by white spheres contributes to increasing the potency of compounds, and the gray and red representations are carbons and oxygen atoms.

(**A**)   (**B**)
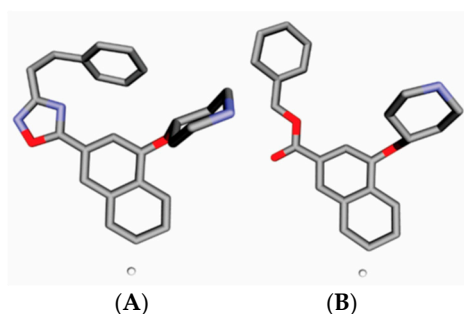
Figure 9. Representation of GCOD-7 (0,6,2, any) (white sphere) and (**A**) compound **62**, and (**B**) compound **59**, and the gray and red representations are carbons and oxygen atoms.

In order to find new active structures, the information of the descriptors obtained by the Model B3 was used. Modifications to the structure of compounds **60**, **65**, and **81** are suggested and compounds **A–E** were proposed.

The proposed compounds **C–E**, exhibited predicted pIC$_{50}$ higher than 81 (the best experimental compound). The structure of the five compounds and their predicted pIC$_{50}$ values are shown in Table 3. The ADME (absorption, distribution, metabolism, and excretion) of a drug is an important property that can determine the utilization of molecules proposed in the therapeutic usage. For the evaluation of pharmacokinetic parameters for molecules **A–E** we used the Lipinski's Rule of Five, where molecular properties are closely related to the oral bioavailability of a drug [20], wherein compounds should not violate more than one rule. In this rule, the compounds should present logP no more than 5, molecular weight of 500 Daltons, number of hydrogen bond acceptors (nON) of 10, number of hydrogen bond donors (nOHNH) of 5, and number of rotatable bonds (nrotb) no more than 10. The proposed molecules have been designed in the Molinspiration Online Property Calculation Software Toolkit [21] to evaluate the criteria discussed above.

Table 3. Structures of the compounds proposed and the predicted pIC$_{50}$ values based on Model B3.

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
A		7.014	B		7.171
C		7.622	D		8.161
E		7.894			

The Molinspiration Online Property Calculation Software Toolkit [21] was used to analyze drug likeness (Lipinski's Rule of Five) and the results are shown in Table 4. According to the data in Table 4, no compound violated the Lipinski's Rule of Five.

Table 4. Calculated parameters of the Lipinski rule of five for the proposed molecules.

Molecule	miLogP	MW	n_{ON}	n_{OHNH}	n	$n_{violations}$
A	3.13	514.97	10	1	8	1
B	2.74	430.89	8	1	7	0
C	3.13	415.88	7	1	6	0
D	2.41	356.81	6	1	4	0
E	3.52	415.88	7	1	6	0

4. Materials and Methods

4.1. Biological Data

In order to build QSAR models, 83 *Plasmodium falciparum* inhibitors were retrieved from Leatherbarrow et al. [15]. Twenty compounds (25%) were randomly selected to compose the test set (external validation). Two test groups were chosen. The first (Test Set I) has the following molecules: **1**, **3**, **5**, **6**, **12**, **16**, **20**, **30**, **33**, **39**, **40**, **50**, **56**, **57**, **61**, **65**, **66**, **69**, **76**, and **80**; Test Set II has the following molecules: **3**, **6**, **9**, **13**, **20**, **21**, **27**, **28**, **31**, **32**, **40**, **56**, **57**, **58**, **64**, **70**, **73**, **76**, **78**, and **82** (Table 5).

Table 5. Chemical structures and experimental pIC$_{50Exp}$ (M) values of *Plasmodium falciparum* inhibitors. Test Set I compound numbers are marked with an asterisk. Test Set II compound numbers are underlined.
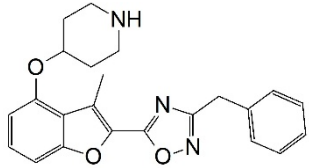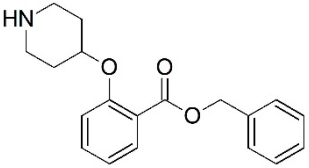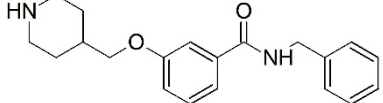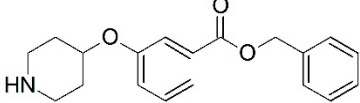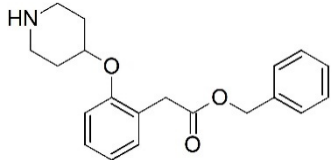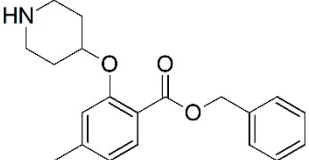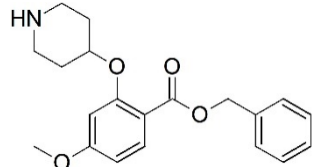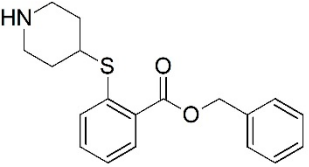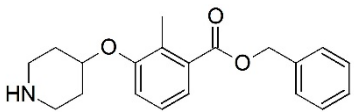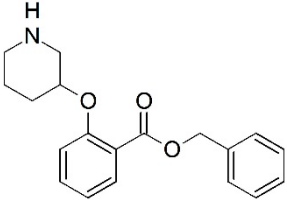
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
1 *		6.155	2		4.000
3 *		4.000	4		4.000
5 *		4.000	6 *		4.000
7		4.000	8		4.000
9		4.000	10		4.000

Table 5. *Cont.*
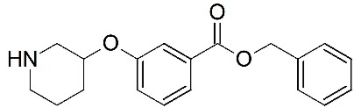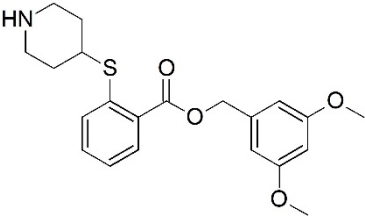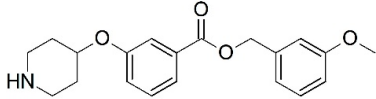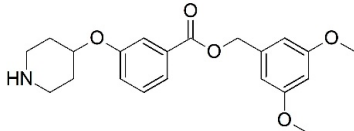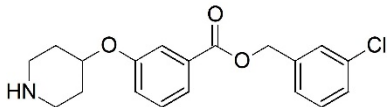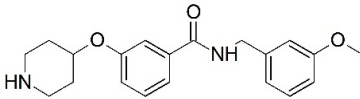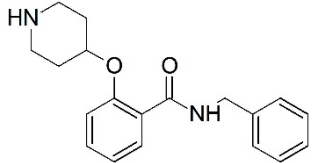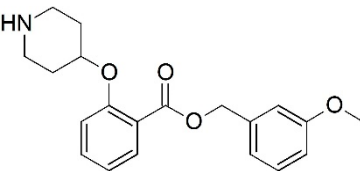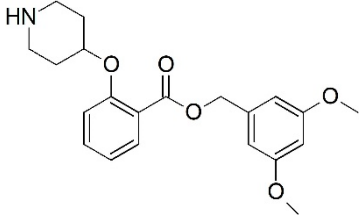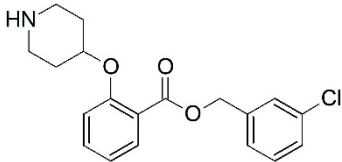
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
11		4.000	**12 ***		5.721
13		4.785	**14**		5.113
15		4.000	**16 ***		4.000
17		4.000	**18**		4.745
19		5.215	**20 ***		4.366

Table 5. *Cont.*

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
21		4.000	**22**		4.000
23		4.000	**24**		4.000
25		4.000	**26**		5.699
27		6.400	**28**		6.102
29		5.780	**30 ***		6.398

Table 5. *Cont.*
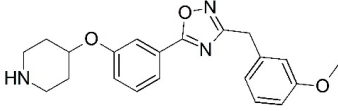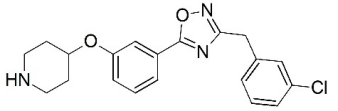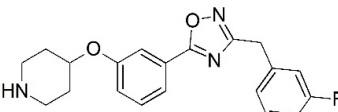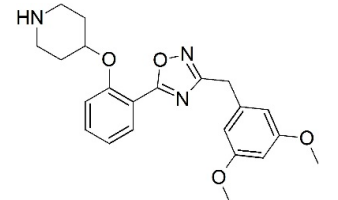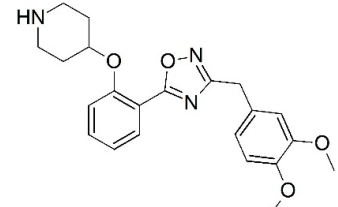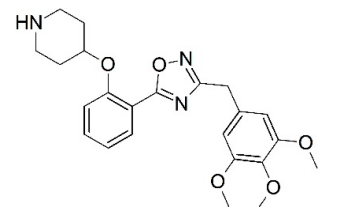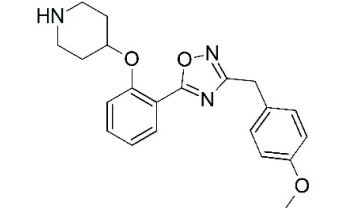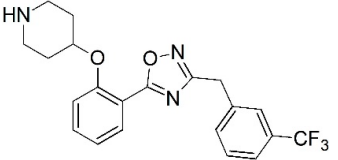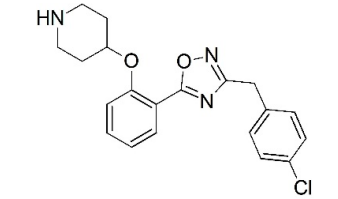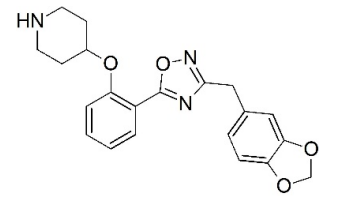
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
31		5.796	**32**		5.420
33 *		5.292	**34**		6.456
35		6.678	**36**		6.468
37		5.131	**38**		5.585
39 *		4.730	**40 ***		5.585

Table 5. *Cont.*
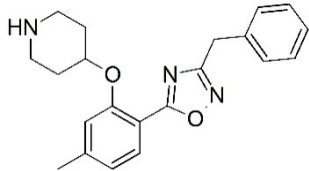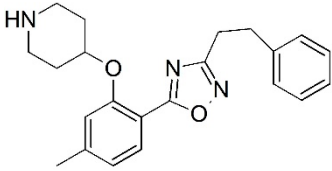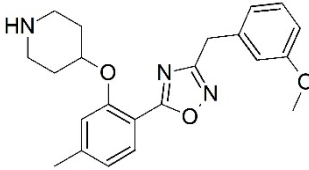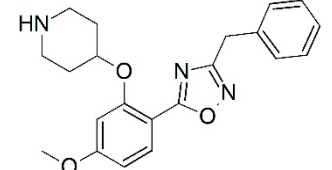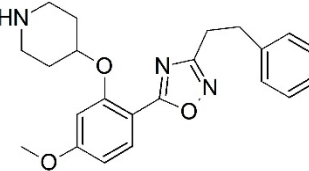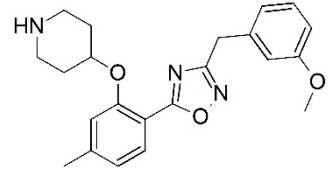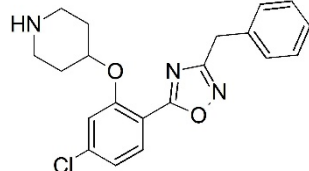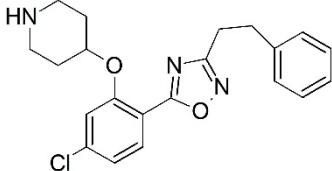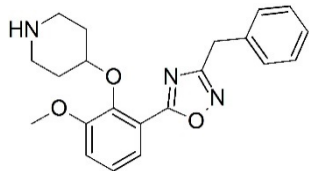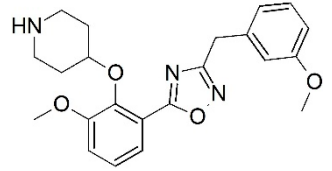
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
41		5.886	**42**		5.284
43		6.000	**44**		5.602
45		4.876	**46**		6.319
47		6.215	**48**		6.051
49		4.445	**50 ***		4.958

Table 5. *Cont.*
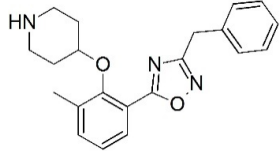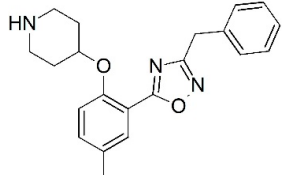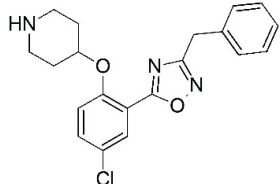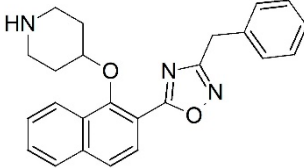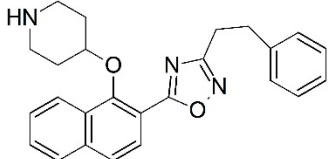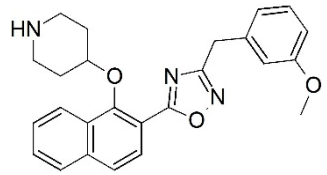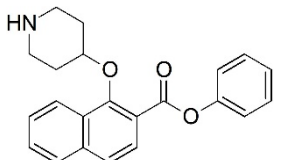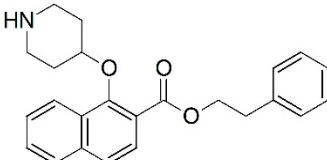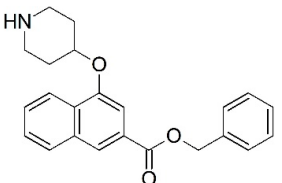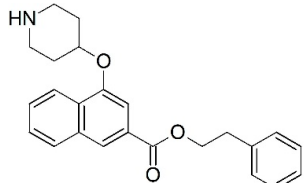
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
51		4.086	**52**		4.217
53		6.398	**54**		5.569
55		4.663	**56 ***		6.229
57 *		4.182	**58**		5.056
59		5.009	**60**		5.149

Table 5. *Cont.*
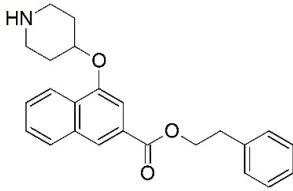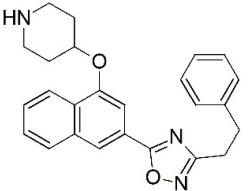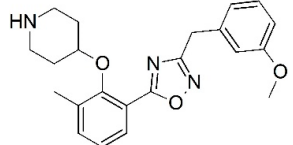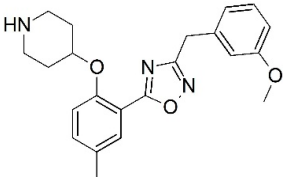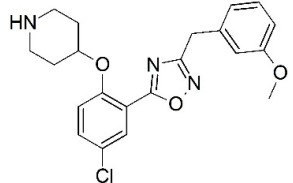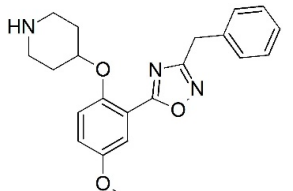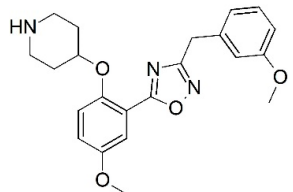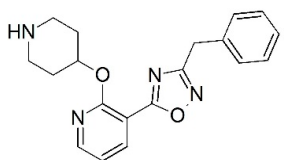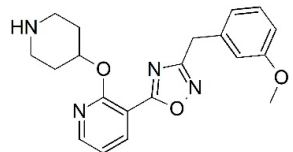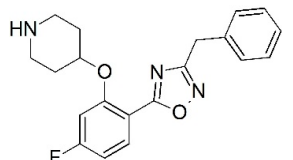
No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
61 *		5.886	**62**		5.886
63		4.801	**64**		5.201
65 *		6.959	**66 ***		5.538
67		6.482	**68**		5.921
69*		6.769	**70**		5.569

Table 5. *Cont.*
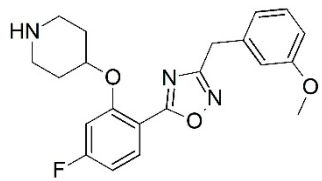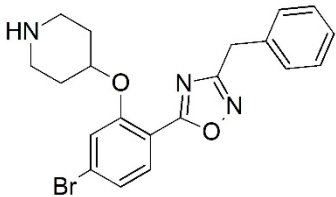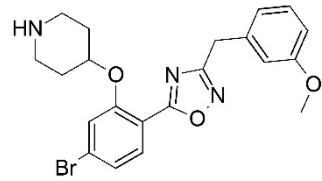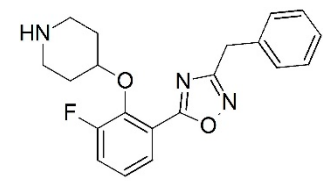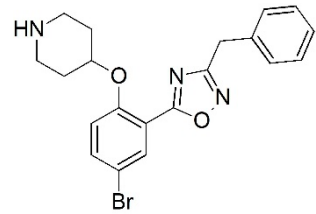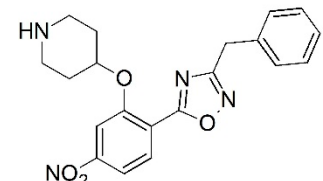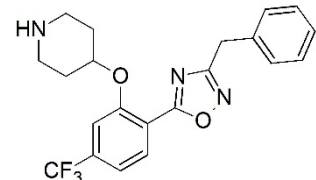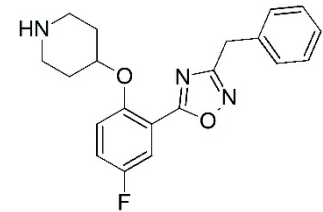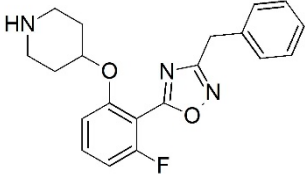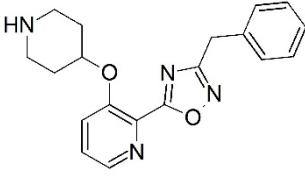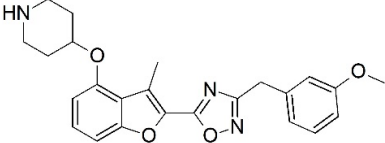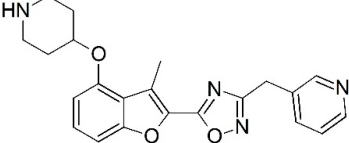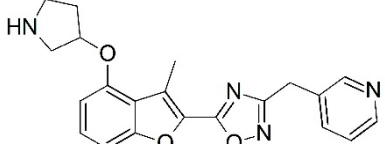
No.	Structure	pIC₅₀	No.	Structure	pIC₅₀
71		6.824	**72**		6.051
73		7.222	**74**		6.000
75		6.620	**76 ***		6.638
77		5.495	**78**		5.959

Table 5. *Cont*.

No.	Structure	pIC$_{50}$	No.	Structure	pIC$_{50}$
79		6.181	**80 ***		5.187
81		7.301	**82**		6.921
83		6.201			

* Test set group of compounds.

The biological activities of these compounds were reported as the negative logarithm of concentration capable of inhibiting 50% of the enzyme activity (IC_{50}), measured using an adapted version of the sensitive fluorescence-based assay based on detection of CoA by 7-diethylamino-3-(4-maleimido-phenyl)-4-methylcoumarin [15]. Furthermore, in an effort to eliminate the potential noise that might have been introduced by the pooling of data sets from different sources, all pharmacological data are obtained from the same laboratory. The IC_{50} (µM) values were converted into molar units and then expressed in negative logarithmic units (pIC_{50}), and are depicted in Table 4. The range of pIC_{50} values for the training and test set spans at least three orders of magnitude (4.00 to 7.30), and the biological activity values show a regular distribution over the whole range.

4.2. Molecular Dynamic Simulation (MDS)

The three-dimensional structure from 83 analogues (Table 5) were optimized in vacuum, without any restriction, and the partial atomic charges assigned using RM1 semiempirical Hamiltonian [22]. The MDS was carried out at 300 K, close to the temperature assays, with a simulation sampling time of 100 ps, and intervals of 0.001 ps. Thus, a total sample of 100,000 conformations of each compound was produced. MDS have been performed using the GROMACS 5.1 package [23].

4.3. Alignment Definition

As the compounds are structural analogs, we will assume that all molecules bind to the receptor in a similar mode. In general, the alignments are chosen to span the common framework of the molecules in the training and test sets [24–26]. In this work, ten alignments were performed using atoms of the (common) benzene ring. Three-ordered atom trial alignments were selected: (1) a-b-c, (2) a-b-d, (3) b-c-d, (4) c-d-f, (5) b-c-a, (6) b-a-c, (7) a-c-b, (8) c-b-a, (9) d-a-f, and (10) a-b-e (Figure 10). The order of the three ordered-atoms is important: the first atom specified for a molecule might be expected to occupy a similar location in space as the first atom specified for the second molecule. The conformational ensemble profile (CEP) for each compound obtained after the MDS step was overlaid onto a cubic lattice with grid cell size of 1A.
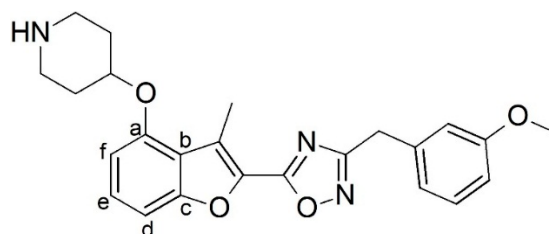
Figure 10. Ordered atom letter codes (a–f) used in the 4D-QSAR analysis defines the three trial alignments. Compound **81** (pIC_{50} = 7.301) is used to define the atom letter code.

4.4. Interaction Pharmacophore Elements

According to the 4D-QSAR methodology, atoms of each compound are defined by seven types of interaction pharmacophore elements (IPEs). IPEs correspond to the interactions that may occur between ligand and the active site: (i) any type (any); (ii) nonpolar (np); (iii) polar-positive charge density (p+); 9iv) polar-negative charge density (p−); (v) hydrogen bond acceptor (hba); (vi) hydrogen bond donor (hbd); and (vii) aromatic systems (ar). The occupancy of the grid cells by each IPE type are recorded over the conformational assembly profile, and forms the set of grid cell occupancy descriptors (GCOD) to be utilized as the pool of trial descriptors in the model building and optimization process [16]. The idea underlying a 4D-QSAR analysis is that variations in biological responses were related to differences in the Boltzmann average spatial distribution of molecular shape with respect to the IPE. Thus, the normalized grid cell absolute occupancy, defined as the number of times that a cell

was occupied by an atom type during the MDS, divided by the size of the CEP (1000 conformations), was used to define the GCODs.

4D-QSAR model calculation. In order to exclude data noise from databases generated by the alignments, partial least-squares (PLS) regression analysis was performed as a data reduction fit between the observed dependent variable measures and the corresponding set of GCOD values. Additionally, PLS identifies the most highly weighted GCODs from data set of local grid cells [16].

The two hundred GCODs with the highest weight from the data reduction were chosen to form the trial descriptor basis sets for model optimization by genetic function approximation (GFA) analysis [16]. Optimizations were initiated using 100 randomly generated models and 10,000–100,000 crossover operations. Mutation probability over the crossover optimization cycle was set at 10–30%. The smoothing factor, the variable that specifies the number of descriptors in the QSAR models, was varied between 1.0 and 3.0, in order to determine equations with no more than twelve terms. Each alignment was evaluated using the procedure described above.

The best models, resulting from the 4D-QSAR study were based on different criteria [18,26–28]:

(1) Coefficient of determination (r^2): is a measure of how well the regression line represents the data.
(2) Adjusted cross-validated squared correlation coefficient (q^2_{adj}): allows the comparison between models with different number of variables.
(3) Correlation coefficient of external validation set (R^2_{pred}): reflects the degree of correlation between the observed ($Y_{Exp(test)}$) and predicted ($Y_{Pred(test)}$) activity data of the test set:

$$R^2_{Pred} = 1 - \frac{\sum_1^n \left(Y_{Exp(test)} - Y_{Pred(test)} \right)^2}{\sum_1^n \left(Y_{Exp(test)} - \overline{Y}_{Training} \right)^2} \tag{1}$$

where $\overline{Y}_{Training}$ is average value for the dependent variable for the training set.
(4) Modified r^2 ($r^2_{m(test)}$) equation determining the proximity between the observed and predicted values with the zero axis intersection:

$$r^2_{m(test)} = r^2(1 - |\sqrt{r^2 - r_0^2}|) \tag{2}$$

(5) Y-randomization (R^2r) consists of the random exchange of the independent variable values. Thus, the R^2r value must be less than the correlation coefficient of the non-randomized models.
(6) R^2p penalizes the model R^2 for the difference between the squared mean correlation coefficient (R^2_r) of randomized models and the square correlation coefficient (r^2) of the non-randomized model:

$$R^2_p = r^2 * \sqrt{r^2 - R^2_r} \tag{3}$$

4.5. Conformational Selection

In the 4D-QSAR method, the conformation of each compound can be postulated as the lowest-energy conformer state from the set sampled for each compound, which predicted the maximum activity using the optimum 4D-QSAR model [16,29–32].

5. Conclusions

In summary, 4D-QSAR models for NMT inhibitors were built and evaluated. Two test groups were evaluated for the ten tested alignments. The best model was obtained from Alignment B3, and generated an equation with seven descriptors, six of which have positive coefficients and one a negative coefficient. Model B3 showed a satisfactory statistical quality and predictive abilities as shown by $r^2 = 0.757$, $q^2 = 0.702$, $q^2_{adjusted} = 0.634$, $R^2_{pred} = 0.746$, $R^2_m = 0.716$, and $R^2_p = 0.609$. Furthermore, it showed low values of $R^2_r = 0.110$, and $\Delta R^2_{m(test)} = 0.133$. 4D-QSAR analysis indicated an important

role of acceptor hydrogen bonding groups and aromatic groups, allowing to propose five structures. These have proved more active than compound **81**, in addition to being assessed by the Lipinski's Rule. Accordingly, these molecules may be considered promising prototypes against malaria.

References

1. Njoroge, M.; Njuguna, N.M.; Mutai, P.; Ongarora, D.S.B.; Smith, P.W.; Chibale, K. Recent Approaches to Chemical Discovery and Development against Malaria and the Neglected Tropical Diseases Human African Trypanosomiasis and Schistosomiasis. *Chem. Rev.* **2014**, *114*, 11138–11163. [CrossRef] [PubMed]

2. Liu, W.; Li, Y.; Learn, G.H.; Rudicell, R.S. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* **2010**, *467*, 420–467. [CrossRef] [PubMed]

3. Krungkrai, S.R.; Krungkrai, J. Insights into the pyrimidine biosynthetic pathway of human malaria parasite *Plasmodium falciparum* as chemotherapeutic target. *Asian Pac. J. Trop. Med.* **2016**, *9*, 525–534. [CrossRef] [PubMed]

4. Ridley, R.G. Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* **2002**, *415*, 686–693. [CrossRef] [PubMed]

5. Tate, E.W.; Bell, A.S.; Rackham, M.D.; Wright, A.H. *N*-yristoyltransferase as a potential drug target in malaria and leishmaniasis. *Parasitology* **2014**, *141*, 37–49. [CrossRef] [PubMed]

6. Fidock, D.A.; Nomura, T.; Talley, A.K.; Cooper, R.A.; Dzekunov, S.M.; Ferdig, M.T.; Ursos, L.M.B.; Sidhu, A.B.S.; Naude, B.; Deitsch, K.W. Mutations in the P-falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell.* **2000**, *6*, 861–871. [CrossRef]

7. Dondorp, A.M.; Nosten, F.; Yi, P.; Das, D.; Phyo, A.P.; Tarning, J.; Lwin, K.M.; Ariey, F.; Hanpithakpong, W.; Lee, S.J. Artemisinin Resistance in *Plasmodium falciparum* Malaria. *N. Engl. J. Med.* **2009**, *361*, 455–467. [CrossRef] [PubMed]

8. Okombo, J.; Chibale, K. Antiplasmodial drug targets: A patent review (2000–2013). *Expert.Opin. Ther. Pat.* **2016**, *26*, 107–130. [CrossRef] [PubMed]

9. Goldston, A.M.; Sharma, A.I.; Paul, K.S.; Engman, D.M. Acylation in trypanosomatids: An essential process and potential drug target. *Trends Parasitol.* **2014**, *30*, 350–360. [CrossRef] [PubMed]

10. Wright, M.H.; Heal, W.P.; Mann, D.J.; Tate, E.W. Protein myristoylation in health and disease. *J. Chem. Biol.* **2010**, *3*, 19–35. [CrossRef] [PubMed]

11. Gunaratne, R.S.; Sajid, M.; Ling, I.T.; Tripathi, R.; Pachebat, J.A.; Holder, A.A. Characterization of N-myristoyltransferase from *Plasmodium falciparum*. *Biochem. J.* **2000**, *348*, 459–463. [CrossRef] [PubMed]

12. Devadas, B.; ZUpec, M.E.; Freeman, S.K.; Brown, D.L.; Nagarajan, S.; Sikorski, J.A.; McWherter, C.A.; Getman, D.P.; Gordon, J.I. Design and Syntheses of Potent and Selective Dipeptide Inhibitors of Candida albicans Myristoyl-CoA: Protein *N*-Myristoyltransferase. *Med. Chem.* **1995**, *38*, 1837–1840. [CrossRef]

13. Paige, L.A.; Zheng, G.; DeFrees, S.A.; Cassady, J.M.; Geahlen, R.L. *S*-(2-Oxopentadecyl)-CoA, a Nonhydrolyzable Analogue of Myristoyl-CoA, Is a Potent Inhibitor of Myristoyl-CoA: Protein *N*-Myristoyltransferase. *J. Med. Chem.* **1989**, *32*, 1665–1667. [CrossRef] [PubMed]

14. Gonçalves, V.; Brannigan, J.A.; Laporte, A.; Bell, A.S.; Robert, S.M.; Wilkinson, A.J.; Leatherbarrow, R.J.; Tate, E.W. Structure-guided optimization of quinoline inhibitors of Plasmodium N-myristoyltransferase. *Med. Chem. Comm.* **2016**, *8*, 191–197. [CrossRef] [PubMed]

15. Leatherbarrow, R.; Tate, E.; Yu, Z.; Racklam, M. Novel Compounds and Their Use in Therapy. London Patent WO 2013/083991, 13 June 2013.

16. Hopfinger, A.J.; Wang, S.; Tokarski, J.S.; Jin, B.Q.; Albuquerque, M.G.; Madhav, P.J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524. [CrossRef]

17. Roy, P.P.; Paul, S.; Mitra, I.; Roy, K. On Two Novel Parameters for Validation of Predictive QSAR Models. *Molecules* **2009**, *14*, 1660–1701. [CrossRef]

18. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P.K.; Kar, S.; Das, R.N. Some case studies on application of "r_m^2" metrics for judging quality of quantitative structure-activity relationship predictions: Emphasis on scaling of response data. *J. Comput. Chem.* **2013**, *34*, 1071–1082. [CrossRef] [PubMed]

19. Pan, D.H.; Liu, J.Z.; Senese, C.; Hopfinger, A.J.; Tseng, Y. Characterization of a ligand-receptor binding event using receptor-dependent four-dimensional quantitative structure-activity relationship analysis. *J. Med. Chem.* **2004**, *47*, 3075–3088. [CrossRef] [PubMed]

20. Patel, H.M.; Sing, B.; Bhardwaj, V.; Palkar, M.; Shaikh, M.S.; Rane, R.; Alwan, W.S.; Gadad, A.K.; Noolvi, M.N.; Karpoormath, R. Design, synthesis and evaluation of small molecule imidazo [2,1-b][1,3,4] thiadiazoles as inhibitors of transforming growth factor-beta type-I receptor kinase (ALK5). *Eur. J. Med. Chem.* **2015**, *93*, 599–613. [CrossRef] [PubMed]

21. Molinspiration Cheminformatics (Nova ulica, Slovensky Grob, Slovak Republic). Available online: http://www.molinspiration.com/ (accessed on 13 September 2018).

22. Rocha, G.B.; Freire, R.O.; Simas, A.M.; Stewart, J.J.P. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101–1111. [CrossRef] [PubMed]

23. Páll, S.; Abraham, M.J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In *International Conference on Exascale Applications and Software*; Springer International Publishing: Cham, Switzerland, 2015; pp. 3–27.

24. Oliveira, A.A.; Ramalho, T.C.; da Cunha, E.F.F. QSAR Study of Androstenedione Analogs as Aromatase Inhibitors. *Lett. Drug Des. Discov.* **2009**, *6*, 554–562. [CrossRef]

25. da Cunha, E.F.F.; Albuquerque, M.G.; Antunes, O.A.C.; de Alencastro, R.B. 4D-QSAR models of HOE/BAY-793 analogues as HIV-1 protease inhibitors. *QSAR Comb. Sci.* **2015**, *24*, 240–253. [CrossRef]

26. Caldas, G.B.; Ramalho, T.C.; da Cunha, E.F.F. Application of 4D-QSAR studies to a series of benzothiophene analogs. *J. Mol. Mod.* **2014**, *20*, 2420. [CrossRef] [PubMed]

27. Kiralj, R.; Ferreira, M.M.C. Basic Validation Procedures for Regression Models in QSAR and QSPR Studies: Theory and Application. *J. Braz. Chem. Soc.* **2009**, *20*, 770–787. [CrossRef]

28. Roy, K.; Paul, S. Exploring 2D and 3D QSARs of 2,4-Diphenyl-1,3-oxazolines for Ovicidal Activity against Tetranychus urticae. *QSAR Comb. Sci.* **2009**, *28*, 406–425. [CrossRef]

29. Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C.P.; Agrawal, R.K. Validation of QSAR Models—Strategies and Importance. *Int. J. Drug Des. Discov.* **2011**, *3*, 511–519.

30. Cormanich, R.A.; Moreira, M.A.; Freitas, M.P.; Ramalho, T.C.; Anconi, C.P.A.; Rittner, R.; Contreras, R.H.; Tormena, C.F. $^{1h}J_{FH}$ coupling in 2-fluorophenol revisited: Is intramolecular hydrogen bond responsible for this long-range coupling? *Magn. Reson. Chem.* **2011**, *49*, 763–767. [CrossRef] [PubMed]

31. Caetano, M.S.; Ramalho, T.C.; Botrel, D.F.; da Cunha, E.F.F.; de Mello, W.C. Understanding the inactivation process of organophosphorus herbicides: A DFT study of glyphosate metallic complexes with Zn^{2+}, Ca^{2+}, Mg^{2+}, Cu^{2+}, Co^{3+}, Fe^{3+}, Cr^{3+}, and Al^{3+}. *Int. J. Quantum Chem.* **2012**, *112*, 2752–2762. [CrossRef]

32. De Castro, A.A.; Prandi, I.G.; Kuca, K.; Ramalho, T.C. Organophosphorus degrading enzymes: Molecular basis and perspectives for enzymatic bioremediation of agrochemicals. *Cienc. Agrotecnol.* **2017**, *41*, 471–482.

Sample Availability: Samples of all compounds are available from the authors.