Molecular factors governing inhibition of arylimidamides against *Leishmania*: Conservative computational modeling to improve chemotherapies

Catharine J. Collar [a], Xiaohua Zhu [b], Karl Werbovetz [b], David W. Boykin [a], W. David Wilson [a,*]

[a] *Department of Chemistry, Georgia State University, Atlanta, GA 30303, United States*
[b] *Division of Medicinal Chemistry and Pharmacognosy, The Ohio State University, Columbus, OH 43210, United States*

ABSTRACT

A dataset of 55 compounds with inhibitory activity against *Leishmania donovani* axenic amastigotes and *Leishmania amazonensis* intracellular parasites was examined through three-dimensional quantitative structure–activity relationship modeling employing molecular descriptors from both rigid and flexible compound alignments. For training and testing purposes, the compounds were divided into two datasets of 45 and 10 compounds, respectively. Statistically significant models were constructed and validated via the internal and external predictions. For all models employing steric, electrostatic, hydrophobic, H-donor and H-acceptor molecular descriptors, the R^2 values were greater than 0.90 and the SEE values were less than 0.22. The models obtained from rigid and flexible compounds were employed together to obtain a conservative method for predictions. This method minimized under predictions. Molecular descriptors from the models were then extrapolated, for the overall predictive devices and the individual compounds, and examined with regard to inhibitory activity. Information gained from the molecular descriptors is useful in the design of novel compounds. The models obtained can be employed to predict activities of the compounds designed and/or form predictions for compounds that exist and have not yet been examined with biological inhibitory assays.

1. Introduction

Leishmania species cause leishmaniasis, an endemic disease found in tropic and subtropic regions riddled with poverty and neglect.[1–3] This infection is most often in the form of cutaneous leishmaniasis where the result is visible skin sores, or visceral leishmaniasis which affects the liver and spleen. Leishmaniasis is primarily transported through the bite of a female phlebotomine sandfly and millions of new cases are reported annually; although, the number of reported cases is probably much lower that the number of actual cases (http://www.who.int/leishmaniasis/burden/magnitude/burden_magnitude/en/print.html).

An estimated half million cases of visceral leishmaniasis occur worldwide annually, and symptomatic infection usually ends in death in the absence of treatment (http://www.who.int/leishmaniasis/visceral_leishmaniasis/en/index.html). Similar to most parasitic diseases plaguing developing countries, there is a lack of affordable and effective therapeutics for leishmaniasis. Traditional treatment of this disease is with pentavalent antimonials such as

sodium stibogluconate and meglumine antimoniate, but resistance to antimonials has become a serious problem in India.[4] The current preferred therapeutics include liposomal amphotericin B (AmBisome), paromomycin, and miltefosine. Single dose AmBisome has shown outstanding efficacy in treating visceral leishmaniasis in India but must be given by injection and is relatively costly.[5] Paromomycin is effective against visceral leishmaniasis in India and is relatively inexpensive, but optimal treatment requires twenty-one daily injections of this drug.[6] Miltefosine is effective when given orally for the treatment of Indian visceral leishmaniasis, but the drug is teratogenic, is more expensive than paromomycin, and is prone to the development of resistance.[7,8] In East Africa, antimonial regimens remain the first-line treatment for visceral disease.[9] Combination therapy has also been implemented in the last few years for the treatment of visceral leishmaniasis to reduce treatment duration, total drug doses, and toxic effects.[9,10] This treatment method also reduces the development of resistance against the drugs and is cost effective.[7]

While not lethal, cutaneous leishmaniasis is a serious problem in developing countries; this infection can lead to severe disfiguring skin lesions when untreated.[11] The infection is also most often treated with pentavalent antimonials although liposomal amphotericin B has recently been shown to be effective in the treatment of cutaneous leishmaniasis.[12] Pentamidine is used to treat *Leishmania guyanensis* infections[13] but not cutaneous infections caused

by other *Leishmania* species. Several other drugs for cutaneous leishmaniasis have been proposed, including allopurinol,[14] rifampicin,[15] dapsone,[16] chloroquine,[17] and nifurtimox.[18]

Through several research endeavors, activities and toxicities of series of compounds have been gathered and such data have been implemented in rational drug design.[19–22] These studies employ biological data of natural and/or synthetic compounds and computational tools to examine compounds with activity against *Leishmania* species. Examination of such compounds has led to the formation of predictive models and from these models the importance of some molecular structures has been ascertained. Although specific receptor interaction studies are important, especially when studying mechanisms, intact parasite studies of inhibition and toxicity are crucial for identifying compounds that will eradicate the parasite from hosts.[23] Such studies of synthetic chalcones and phospholipids display effective antileishmanial activity for compounds with: (1) a long alkyl chain, (2) bulky groups at the ends of the alkyl chains, and (3) an electron deficient group.[20,24]

Our studies examine a biological dataset of synthetic arylimidamides which possess activities against *Leishmania donovani* axenic amastigotes and *Leishmania amazonensis* intracellular parasites. Inhibitory data, in the form of IC_{50} values, and Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) molecular descriptors were employed with partial least squares (PLS) regression to correlate the biological data with molecular structures and properties. Predictive models and molecular descriptor potentials derived from the models contribute to the identification and understanding of important molecular features that govern the inhibitory activities of arylimidamides against species of *Leishmania*.

2. Materials and methods

Inhibitory data were gathered for arylimidamides with activity against *L. donovani* axenic amastigotes and *L. amazonensis* intracellular parasites. This data was implemented for three-dimensional QSAR modeling employing rigid and pharmacophore alignments. The compounds acquired for this study are from the David Boykin compound library at Georgia State University, and representative synthetic procedures are found in the literature.[25–30] The syntheses of many of the compounds used in this study have been reported[29,30] and the remainder were prepared using the same methodology. All compounds were characterized by ^1H and ^{13}C NMR and by giving satisfactory elemental analysis (C,H,N within ±0.4% of theory). The toxicities of some of the most active arylimidamides have been examined through in vitro and in vivo assays, and the data demonstrate that arylimidamides are promising preclinical candidates for the treatment of visceral leishmaniasis.[30]

2.1. Inhibition data

Briefly, IC_{50} (µM) values were determined for compounds of interest using two assays. The first assay screened against axenic amastigote-like *L. donovani*, while the second screened against *L. amazonensis* intracellular parasites. Screening against *L. donovani* was conducted by: (1) culturing Ld1s parasites in potassium-based medium at pH 5.5, 37 °C, (2) incubating for three days with compounds in a 96-well plate, and (3) adding tetrazolium dye and quantifying the assay spectrophotometrically as outlined previously.[29] Evaluation of compounds against *L. amazonensis* intracellular parasites was conducted by: (1) plating macrophages and allowing the host cells to adhere overnight, (2) adding *L. amazonensis* promastigotes transfected with the β-lactamase gene (MOI: 5:1) and incubating overnight, (3) adding compounds of interest and incubating for 72 h at 34 °C, (4) adding nitrocefin in lysis buffer

and incubating an additional 3–5 h, and (5) reading the plate at 490 nm as described earlier.[31,32] Experimental IC_{50} values for *L. donovani* axenic amastigotes and *L. amazonensis* intracellular parasites were obtained for 55 compounds.

2.2. Preparation of compounds for computational studies

SYBYL 8.1[33] software was employed to construct all compounds in three-dimensional space. Compounds were then divided into training and testing datasets. These datasets consisted of 45 and 10 compounds, respectively. The compounds of the training dataset then underwent a short molecular dynamics simulation of 1 ns. This system employed SYBYL 8.1 default settings at a constant temperature and volume (NTV). Briefly, (1) the system temperature was set at 300 K with a coupling constant of 100 fs, (2) Maxwell–Boltzmann distribution was employed for initial atom velocities, (3) the non-bonded pair list was updated every 25 fs, and (4) the duration of the molecular dynamics simulations in vacuo was 1 ns with a time step of 100 fs and a snapshot every 1000 fs. The dynamics snapshots displayed several low energy structures. Torsional angles of all training dataset compounds were modified to explore the low energy conformations and modified structures were minimized to convergence using the Tripos force field, conjugate gradient algorithm, and Gasteiger–Hückel charges. The termination gradient was 0.01 kcal/(mol Å) and the maximum iterations were 10^4.

2.3. Rigid alignment of compounds and resulting models

Each training dataset of compounds with modified torsional angles was aligned using the 'Align Database' option of the QSAR module in SYBYL. CoMFA (steric and electrostatic) and CoMSIA (steric, electrostatic, hydrophobic, H-donors and H-acceptors) molecular descriptors were calculated for the aligned structures and PLS regression was employed to correlate the molecular descriptors of compounds to experimental average IC_{50} values. The number of components was determined by the smallest predicted error sum of squares. Optimum models employing CoMFA molecular descriptors consisted of three components, whereas the ones with CoMSIA molecular descriptors employed six.

2.4. Flexible alignment of compounds and resulting models

Five compounds with low IC_{50} values in the *L. amazonensis* intracellular parasite assay were employed for flexible compound alignment using the 'Align Pharmacophore' option of the GALAHAD module. Parameters were acquired through the 'Suggest from Data' option and the best 20 models were collected. The highest scoring model with respect to maximized pharmacophore consensus, maximized steric consensus, and minimized energy was employed as a template for individual compound alignment of the entire training dataset. The 'Align Molecules to Template Individually' option was selected and parameters were acquired once more through the 'Suggest from Data' option; the 'Keep Best N Models' option was reset to 20. Molecular descriptors were calculated for the highest scoring model and PLS regression was implemented in the same manner as for the rigid compounds. The optimum numbers of components were determined as previously described; models with CoMFA molecular descriptors consisted of three components, whereas models with CoMSIA molecular descriptors employed six.

2.5. Statistical analyzes

The statistics calculated from PLS regression included: a leave-one-out cross-validated correlation coefficient (Q^2), the coefficient of determination (R^2), the standard error of estimate (SEE), the F

statistic, a bootstrap R^2 (R_{bs}^2), and a bootstrap SEE (SEE$_{bs}$). The bootstrap analysis was used to check the stability of the models through cross-validation into two, five, and ten groups. The average values of the bootstrap analysis are displayed with the rest of the statistics. A scramble test was also preformed to address chance correlation; statistics and predictions are reported in Supplementary Tables S2 and S3.

2.6. Testing datasets and the conservative model method

The models constructed from the rigid and flexible alignments were employed to examine testing datasets that were aligned via rigid and flexible methods. Of the pIC$_{50}$ values predicted, for both training and testing datasets, the more negative pIC$_{50}$ prediction was considered the most viable. This method favors over prediction, larger IC$_{50}$ values, rather than under prediction.

2.7. Molecular descriptor potentials

Molecular descriptor potentials acquired through the mapping of the product standard deviation with respect to molecular descriptor values and coefficients at each lattice point were extrapolated from the models. Default levels of contour by contribution were employed to gather favored and disfavored potentials for overall models. The individual compounds of the models were analyzed via the contour by actual analysis method. Software output was used to determine the proper ranges of assigned favored and disfavored contour regions for individual compounds.

3. Results

The entirety of the dataset, 45 training and 10 testing compounds (Supplementary Table S1 and Table 3), can be represented via the scaffold structure displayed in Figure 1. At each of the five positions labeled in this figure there are differing atoms or groups: positions one and four display single atom changes in the form of carbon, oxygen, sulfur, and nitrogen, whereas positions two, three and five display larger group substituent modifications as shown in Supplementary Table S1 and Table 3.

Biological IC$_{50}$ values were determined for each compound of the training and testing datasets through two assays targeting *L. donovani* axenic amastigotes and *L. amazonensis* intracellular parasites. These inhibitory values were averaged and standard errors ($n \geqslant 3$) were calculated (Table 3). For modeling purposes, the IC$_{50}$ values were log transformed into pIC$_{50}$ values (pIC$_{50}$ = −log(IC$_{50}$)). Figure 2 displays the pIC$_{50}$ data; experimental values against the *L. donovani* axenic amastigotes are shown in green, whereas the values against the *L. amazonensis* intracellular parasites are displayed in blue. The standard error of the data is represented in general by trend lines and the averaged values are shown as triangles and squares, respectively. Notice that the slopes are
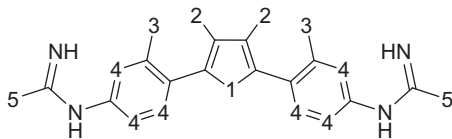

Figure 2. Biological pIC$_{50}$ data of synthetic arylimidamides active against *L. donovani* axenic amastigotes (green) and *L. amazonensis* intracellular parasites (blue). The negative log values of average experimentally obtained IC$_{50}$ data, displayed in Supplementary Table S1 and Table 3, and these values plus and minus respective standard deviations are all plotted against the negative log value of average experimentally obtained IC$_{50}$ data.

very similar with values between 0.96 and 1.0, and R^2 values are 0.95 or higher. This displays the relative range of inhibitory activity and the standard error from the average inhibitory values associated with each synthetic compound in the training and testing datasets. The pIC$_{50}$ distribution of data is also shown in this figure; the inhibitory activity of arylimidamides against *L. donovani* axenic amastigotes ranges between approximately −2.5 and 0.5, whereas those active against *L. amazonensis* intracellular parasites range between about −1.5 and 1.5.

Compounds examined through biological assays were aligned in three-dimensional conformations using two methods: (1) rigid alignments of compounds were obtained through the implementation of the SYBYL 'Align Database' option of the QSAR module, and (2) flexible alignments of compounds were acquired through the use of the 'Pharmacophore Alignment' option of the GALAHAD module. Rigid alignments were performed on low energy conformations of compounds. Molecular descriptors were then calculated and PLS regression was employed to construct predictive models from the descriptors and respective biological inhibitory data. The best computational models formed consisted of compounds in their most linear conformation with an overall plus one charge. The plus one value was used, since the arylimidamides have a pK_a near 7.

Flexible alignment of compounds employed the five most active compounds against the *L. amazonensis* intracellular parasites from the training dataset (Fig. 3). Figure 4 displays the outcome of pharmacophore simulations that lead to PLS regression models employing flexible compounds. The rotation of the compounds allows for visualization of alignment and positioning of identified feature potentials. The observed features governing structure alignment are: (1) four aromatic rings (cyan); (2) N=C−N groups, two positive nitrogens (red) and a H-donor (magenta); (3) atoms at the one and two position of Figure 1, two H-acceptor (green); (4) atoms at a five position, a H-donor or H-acceptor (overlaid magenta and green equates to dark green). In the next step all of the training and testing compounds were flexibly aligned to the pharmacophore. These alignments can be viewed in relation to rigid alignments (Fig. 5). Rigid compounds were aligned by N=C−N groups. Notice that there is a difference in the spatial relationships of the compounds.

Figure 1. Scaffold structure for compounds with biological inhibitory data for *L. donovani* axenic amastigotes and *L. amazonensis* intracellular parasites. The numbered positions on the scaffold identify the locations where compounds differ, and these serve as a guide for explanation of model findings. All training dataset structures and respective inhibitory data can be viewed in Supplementary Table S1 and Table 3. There is an overall plus one charge on these compounds; protonation takes place on the N–C=N groups between the aromatic rings with labeled position 4 and the groups at position 5.
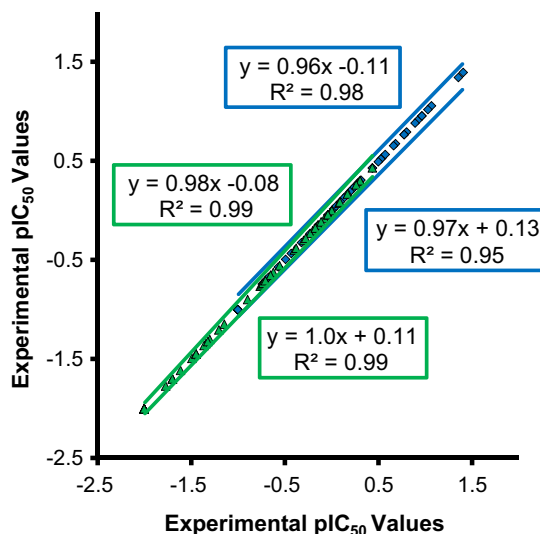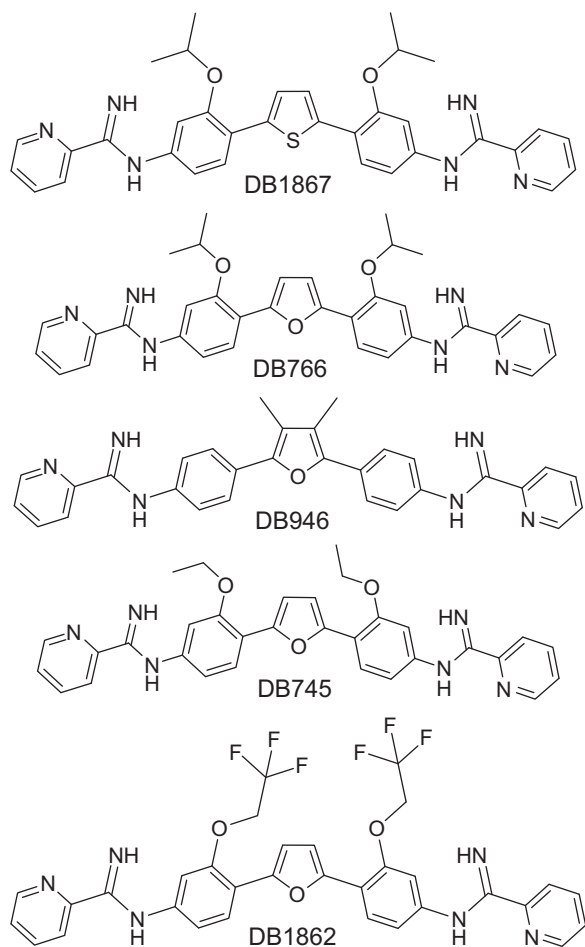
Figure 3. Five of the most active compounds against *L. amazonensis* intracellular parasites and *L. donovani* axenic amastigotes.

Inhibitory activities with structures and properties of the aligned compounds were employed to construct predictive models through PLS regression methods. The statistics for these models indicate that CoMSIA molecular descriptors outperform those of CoMFA molecular descriptors (Table 1). This is shown in higher Q^2, R^2, and F statistics and lower SEE statistics for the models constructed with rigid compounds. Models from flexible compounds displayed higher R^2 and F statistics and lower SEE statistics but lower Q^2 values. The lower Q^2 are attributed to: (1) torsional variability, (2) differences in optimal low energy structural conformations, and (3) contributions of compound inhibitory activities.[34]

The internal (training dataset) and external (testing dataset) predictions of the models are displayed in Figure 6 where predicted pIC_{50} values are plotted versus the experimental pIC_{50} values. The training dataset of this figure is colored in accordance to Figure 2, whereas all testing dataset predictions are in red. Although internal predictions were linear, some testing dataset compounds were more difficult to predict for than others. The variance in compound prediction differed between the models for compounds of rigid and flexible alignments; hence, by taking the most negative prediction of each compound regardless of rigid or flexible alignment and plotting these values against respective experimental data a conservative method for prediction can be obtained. The combination of the models reduces under prediction. Table 3 displays the testing dataset along with experimental average IC_{50} values, plus or minus respective standard error, along with the conservative model IC_{50} prediction. The R^2_{Test} and SEE_{Test} values for the experimental and predicted conservative data shown in Table 3 are, respectively, 0.36 and 0.23 for *L. amazonensis* intracellular parasites, and 0.72 and 0.20 for *L. donovani* axenic amastigotes; the statistics are calculated on the pIC_{50} data. When the R^2_{Test} statistics are calculated on the IC_{50} data of Table 3 the R^2_{Test} statistics are 0.59 and 0.93. The statistics give a rough idea of model predictability, whereas the actual prediction in relation to experimental data shows the model's true ability to predict with relation to standard error.

From the models employing CoMSIA, three-dimensional molecular descriptor potentials were determined and viewed as surfaces
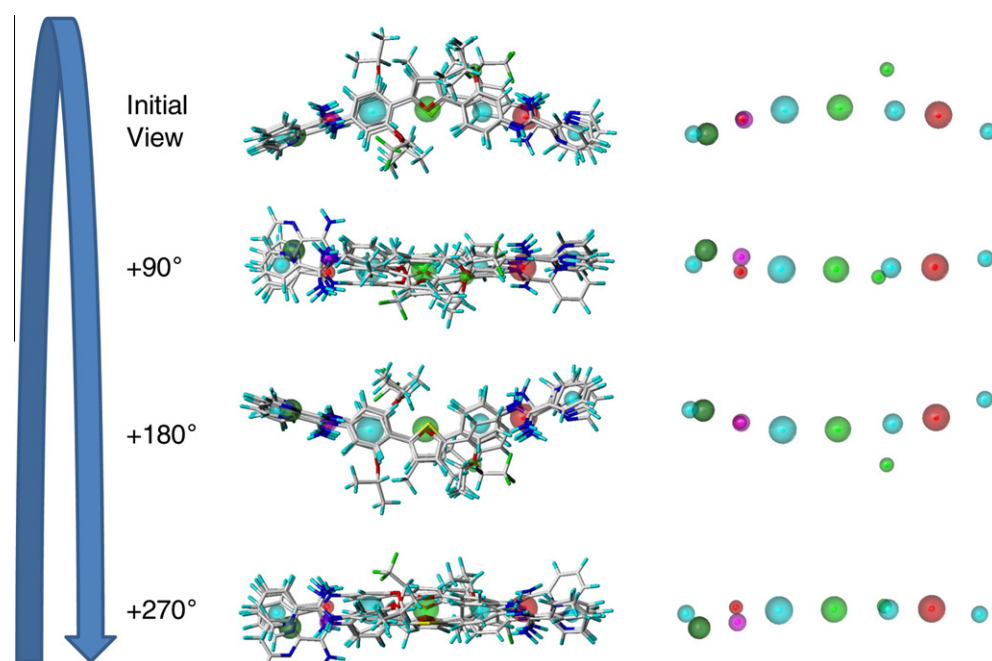

Figure 4. GALAHAD potentials as identified by simulations employing the compounds of Figure 3. The identified features are color coded: cyan, hydrophobes; magenta, donor atoms; green, acceptor atoms; red, positive nitrogens.
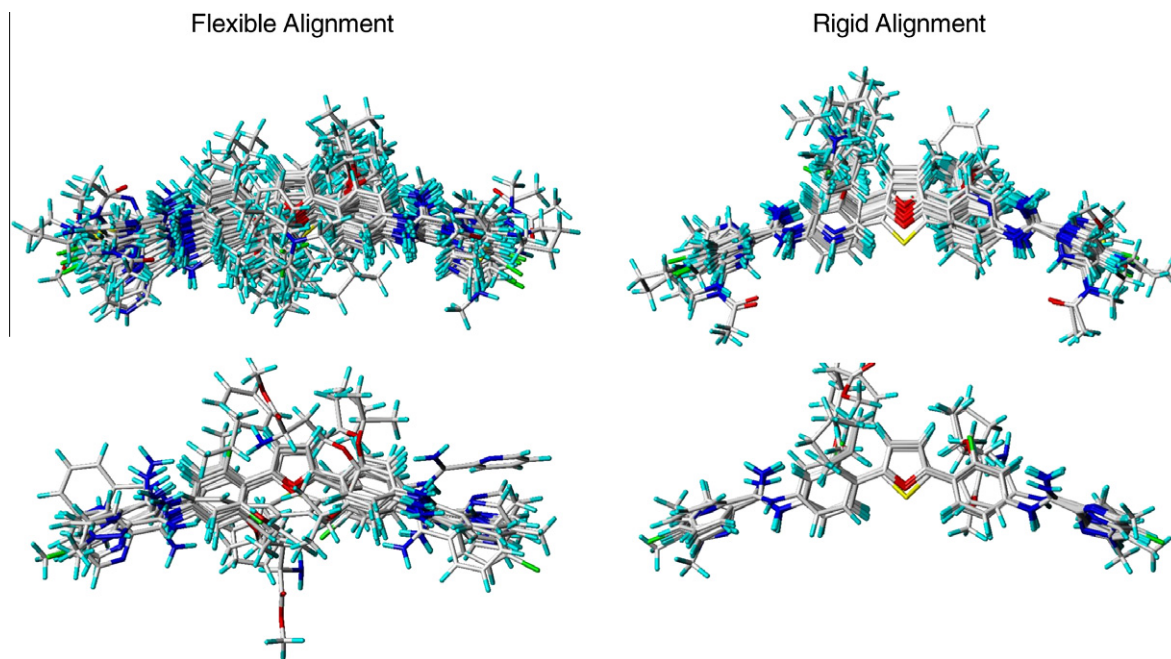
Flexible Alignment   Rigid Alignment

Figure 5. Final training (top) and testing (bottom) datasets: flexible alignments (left) and rigid alignments (right).

Table 1
Statistics of partial least squares predictive models for a biological dataset of synthetic arylimidamides with activities against *L. donovani* axenic amastigotes (LD) and *L. amazonensis* intracellular parasites (LA)

| | Rigid alignment | | | | Flexible alignment | | | |
| | CoMFA | | CoMSIA | | CoMFA | | CoMSIA | |
	LA	LD	LA	LD	LA	LD	LA	LD
Q^2	0.23	0.25	0.47	0.59	0.16	0.60	0.07	0.22
SEE	0.45	0.41	0.25	0.18	0.24	0.33	0.16	0.14
R^2	0.68	0.77	0.91	0.96	0.91	0.85	0.96	0.97
F	29.1	44.5	61.4	137	131	78.0	159	229
SEE_{bs}	0.35	0.37	0.21	0.16	0.21	0.26	0.12	0.12
R^2_{bs}	0.80	0.82	0.94	0.97	0.93	0.90	0.98	0.98

in relation to the overall models (Fig. 7 and Supplementary Fig. S1) as well as individual compounds (Fig. 8 and Supplementary Fig. S2). Figure 7 displays the overall CoMSIA molecular descriptors for the rigid and flexible models. It is evident that each overall model displays different molecular descriptor potential contributions for the steric, electrostatic, hydrophobic, H-donor and H-acceptor potentials. This indicates that each model is constructed somewhat differently; although there are similarities between the results obtained. Using the positions of Figure 1 as a reference: (1) steric bulk is favored (green) at positions three and five and perhaps not symmetrically, whereas disfavored steric bulk (yellow) regions are just outside those favored, (2) positive electrostatic charge is favored (blue) at one, if not both, of the N═C–N groups near position five, whereas negative charge is favored (red) predominantly at or near position one and outside one of the N═C–N groups, (3) hydrophobic interactions are favored (yellow) at positions two and five, whereas disfavored hydrophobic interactions (gray) are near positions three and outside positions five, (4) H-donor atoms are favored (cyan) predominantly at or below the five position and disfavored (purple) in areas beyond favored regions, and (5) H-acceptors are favored (magenta) near the terminal N═C–N groups, and disfavored (red) below the four position(s) and outside favored N═C–N groups of the comparison molecule DB766.
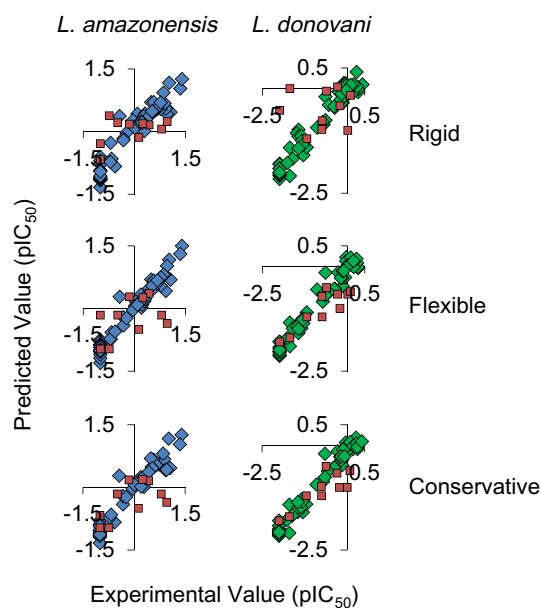

Figure 6. Internal (blue and green) and external (red) predictions. The internal predictions are those for the training compounds and external predictions are for the testing dataset of compounds. The *L. amazonensis* experimental versus predicted results are shown in blue (left) and those for *L. donovani* in green (right). The experimental versus predicted results from top to bottom are predictions from implementing rigid (top) and flexible (center) compounds. The conservative predictions (bottom) are essentially the more negative of the two pIC_{50} predictions resulting from the models with rigid and flexible compounds. Since the scale observed is the negative log of the IC_{50}, this method reduces under prediction.

With regard to the scaffold structure of Figure 1, Figure 8 displays the molecular descriptor potentials of individual compounds based on CoMSIA molecular descriptors. The molecular descriptor potential regions of individual molecules appear to be more consistent within their respective rigid and flexible models than they were in the overall models of Figure 7 and Supplementary Figure S1. The molecular potentials that resulted were, however, also
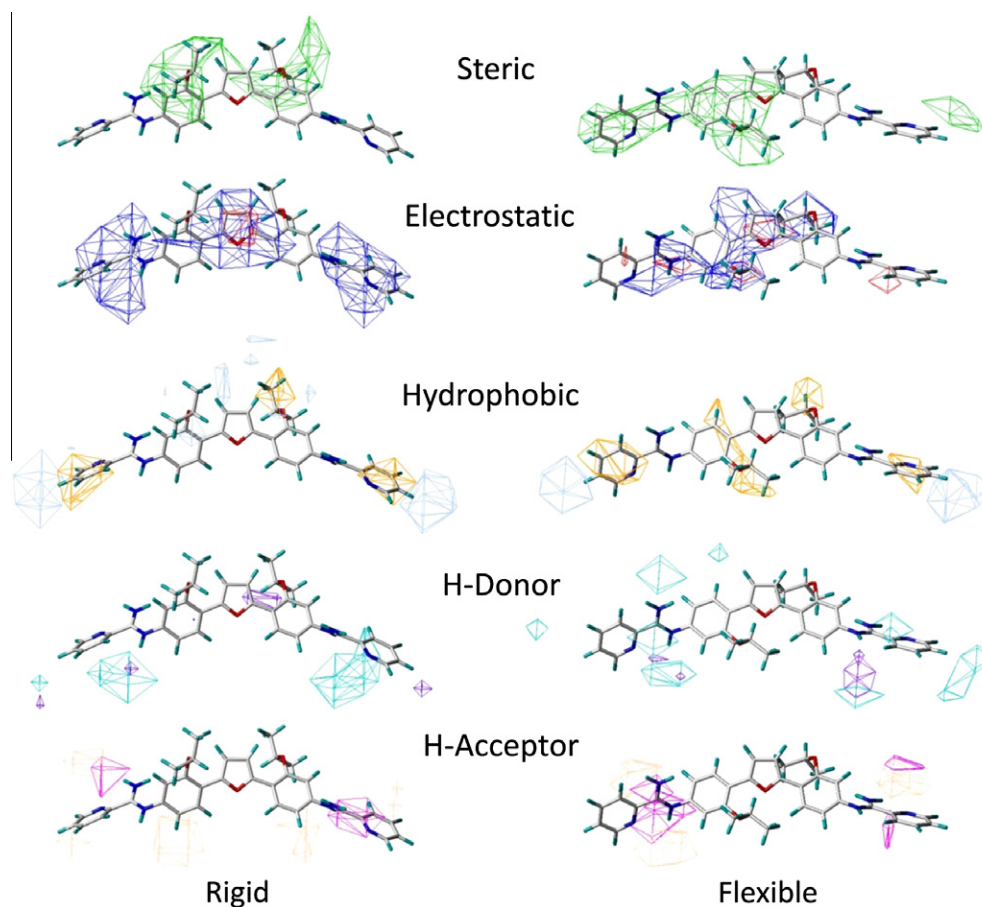
Figure 7. Overall models with CoMSIA molecular descriptor surfaces for both rigid and flexible structure alignments of compounds active against *L. amazonensis*; DB766 is displayed as a reference compound. Favored potentials from steric to H-acceptor molecular descriptors are green, blue, yellow, cyan, and magenta, whereas disfavored potentials from steric to H-acceptor molecular descriptors are yellow, red, gray, purple, and red. Results for *L. donovani* can be found in Supplementary Figure S1.

fewer, and included favored and disfavored hydrophobic, favored H-donor and favored H-acceptor potentials.

With the models in Figure 8 and biological data in Supplementary Table S1, it is possible to examine not only the molecular descriptor potentials with regard to model contribution but also the contribution of substructures to biological inhibitory activity. To most effectively describe these findings, it is important that comparisons are made to a compound that is active in both datasets. DB766, a potential drug candidate against visceral leishmaniasis,[30] was selected for analyzes. With respect to Figure 1, the molecular descriptor potentials for DB766 include: (1) favored hydrophobic potentials near the aromatic groups, (2) favored H-donor potentials are displayed below the left side five position (N=C−N group), and (3) favored H-acceptor potentials are on the N=C−N group opposite the side of the favored H-donor and extended to the terminal aromatic ring. The IC$_{50}$ values for this compound against *L. amazonensis* intracellular parasites and *L. donovani* axenic amastigotes are 0.09 and 0.50 μM, respectively. The general structure of DB1867, compared to DB766, differs only by a sulfur atom at position one and with this change the compound becomes more linear and favored hydrophobic interactions are spread to positions one and two (Fig. 8). Potentials for favored H-acceptors are near N=C−N groups and the IC$_{50}$ values are 0.05 and 0.68 μM, respectively. DB946 is the only compound in the training dataset to differ from DB766 at position two; this compound also differs at position three. The methyl groups at position two fill similar special areas as substituents in position three. Favored hydrophobic potentials reside in position two, three, and five locations.

This compound's IC$_{50}$ values are 0.11 and 0.37 μM, respectively. DB667 and DB1876 differ from DB766 at position three. DB667 possesses hydrogen atoms at position three and molecular descriptor potentials similar to those of DB946 (hydrophobic) and DB766 (H-donor and H-acceptor), although the favored hydrophobic potentials span a greater length for DB667. The IC$_{50}$ values for this compound are 0.53 and 1.6 μM, respectively. DB1876 displays large disfavored hydrophobic molecular descriptor potentials at the three positions. The remaining potentials are favored hydrophobic potentials near the aromatic rings and favored H-donor and H-acceptor potentials near the N=C−N group(s). This compound has IC$_{50}$ values of 2.1 and 28 μM, respectively. DB1851 differs from DB766 at position four and this change resulted in favored hydrophobic interactions that span more of the molecule than previous compounds discussed with H-donor and H-acceptor potentials similar to those of DB1876. The IC$_{50}$ values for this compound are poor, greater than 10 and 50 μM, respectively. DB1921, DB1942, and DB1906 all differ from DB766 at the five positions. DB1921 is flanked at the five positions and has different substituents at the three positions. This compound consists of potentials similar to DB1876; however, it is also missing most of the favored hydrophobic and H-acceptor potentials. The IC$_{50}$ values for this compound are 4.7 and 41 μM, respectively. DB1942 consists of a longer, more flexible ring structure than DB766 and consists of disfavored hydrophobic molecular descriptor potentials primarily at the five positions. Positive hydrophobic potentials are on the inner aromatic rings or the outer rings near the five position, whereas H-donors are favored on one side of an N=C−N group
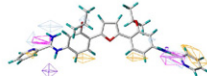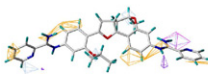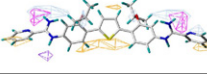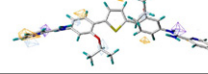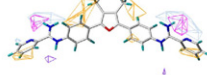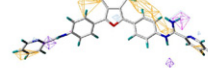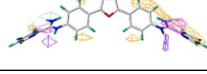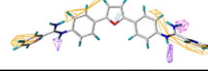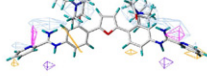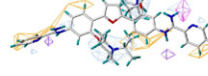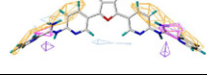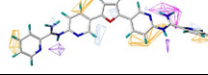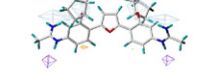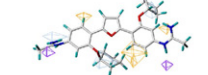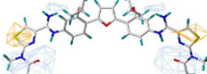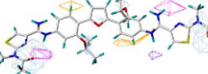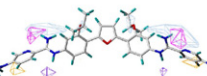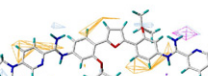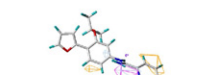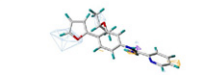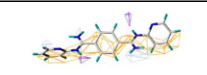
X	Compound	Rigid	Flexible
0	DB766		
1	DB1867		
2	DB946		
3a	DB667		
3b	DB1876		
4	DB1851		
5a	DB1921		
5b	DB1942		
5c	DB1906		
6,7	DB1882		
8	DB1911		

Figure 8. CoMSIA findings with respect to Figure 1 and molecular descriptor potentials of Figure 7. The favored hydrophobic potentials have been changed to orange to improve visualization and insure that steric potentials were not displayed. The left most column consists of numbers correlated to positions of Figure 1. The column to the right consists of the compounds name. This is followed by the compounds and their respective molecular descriptor potentials for each the final models. To improve compound visibility, the results for only *L. amazonensis* are shown above. Results for *L. donovani* can be found in Supplementary Figure S2.

and H-acceptors are favored at both N=C–N group(s). This compound has IC_{50} values of 0.81 and 3.6 μM, respectively. The five positions of DB1906 consist of more rings than DB766. The molecular descriptor potentials for this compound were similar to those of DB766 (hydrophobic and H-donor) and DB946 (H-acceptor), IC_{50} values are 0.27 and 1.9 μM, respectively.

4. Discussion

QSAR studies have previously been employed to determine the importance of chalcone and phospholipid molecular structures, and predict activities and toxicities for series of compounds with activity against *Leishmania* species.[19–22] These studies found that potent antileishmanial activity occurred when compounds possessed a long alkyl chain, bulky groups at the end of the alkyl chain, and an electron deficient group.[20,24] The structures of chalcones and phospholipids are quite different from each other, and these compounds differ substantially from the arylimidamides examined in this study (Supplementary Table S1).

Table 2
Contribution of CoMSIA molecular descriptors for rigid and flexible models employing structures of training dataset compounds and respective biological activities

	Rigid alignment		Flexible alignment	
	L. amazonensis	*L. donovani*	*L. amazonensis*	*L. donovani*
Steric	0.15	0.14	0.13	0.13
Electrostatic	0.11	0.08	0.14	0.15
Hydrophobic	0.47	0.43	0.33	0.34
H-donor	0.15	0.21	0.20	0.21
H-acceptor	0.12	0.14	0.20	0.17

4.1. Pharmacophore selection

The numbered locations of Figure 1 aid in the explanation of inhibition results displayed in Figure 2 through the interpretation of pharmacophore consensus potentials (Fig. 4) and molecular descriptor contribution potentials (Figs. 7 and 8). The pharmacophore alignment of Figure 4 is calculated using the compounds of Figure 3. By only employing the most active compounds, the

Table 3
Predictions in terms of IC$_{50}$. Experimental values for compound inhibitory activity against *L. donovani* axenic amastigotes (LD) and *L. amazonensis* intracellular parasites (LA) are displayed in columns LD Exp and LA Exp. The predicted values are those from the conservative predictions, LD Calcd and LA Calcd. These are also displayed in Figure 6

Name	Structure	LA Exp (μM)	LA Calcd (μM)	LD Exp (μM)	LD Calcd (μM)
DB710		0.16 ± 0.04	1.4	0.84 ± 0.2	4.0
DB712		0.56 ± 0.08	0.65	2.0 ± 0.6	4.6
DB749		3.1 ± 0.7	1.4	>50	50
DB874		1.4 ± 0.3	0.66	4.2 ± 1.3	3.2
DB889		0.11 ± 0.02	2.3	1.7 ± 0.5	10
DB1856		0.74 ± 0.3	3.2	5.8 ± 0.7	10
DB1857		0.37 ± 0.2	0.69	1.0 ± 0.2	10
DB1864		>10	9.3	5.6 ± 1.8	16
DB1908		>10	4.7	14 ± 1	16
DB1930		5.5 ± 0.2	9.3	>100	63

pharmacophore is strictly for compounds of similar structure and activity. The pharmacophore results suggest the importance of aromatic and positively charged N=C−N groups, and these results can be examined with respect to the chalcone and phospholipid results reported previously.[19–22] The aromatic rings of the arylimidamides can be related to the bulky groups, the positively charged N=C−N groups can be related to the electron deficient groups, and the linker between the central aromatic rings can be related to the long alkyl chain. The pharmacophore results show that the linker between the center aromatic rings does not need to be a ring structure.

4.2. Regression analyzes

PLS regression of calculated molecular descriptor potentials and respective biological inhibitory values for both the rigid and flexible alignments of compounds presented in Figure 5 produced statistically significant models. Models employing the CoMSIA molecular descriptors from rigid structure alignment were the only ones with Q^2 values greater than 0.5 (Table 1). Q^2 values greater than 0.5 indicate models with predictability better than chance.[33] What we realize from our models, especially those aligned by flexible conformations, is that each compound contributes to the entirety
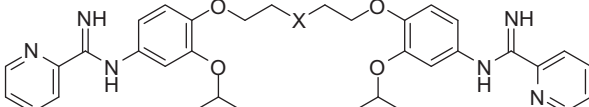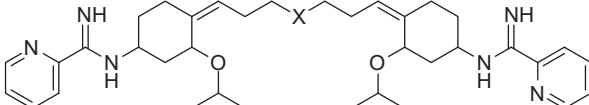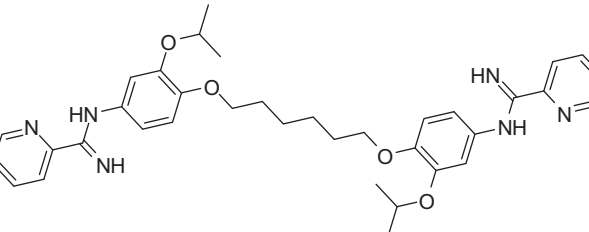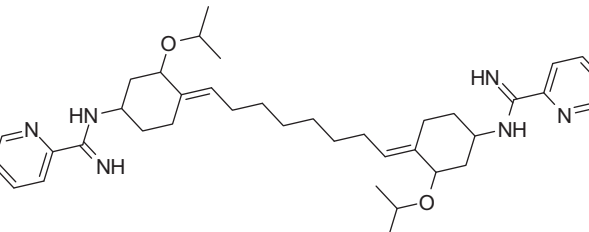
4560

C. J. Collar et al./Bioorg. Med. Chem. 19 (2011) 4552–4561

Compound	*L. donovani*		*L. amazonensis*	
[structure: bis-amidine aryl ether with isopropoxy groups, central X linker]	X	Predict.	X	Predict.
	C	(0.39, 0.96)	C	(0.08, 0.41)
	O	(0.72, 1.8)	O	(0.12, 0.70)
	S	(0.50, 1.3)	S	(0.10, 0.53)
[structure: bis-amidine cyclohexylidene with isopropoxy groups, central X linker]	X	Predict.	X	Predict.
	C	(0.40, 2.0)	C	(0.14, 1.7)
	O	(1.6, 4.5)	O	(0.21, 3.0)
	S	(1.2, 6.1)	S	(0.14, 4.5)
[structure: bis-amidine aryl ether with isopropoxy groups, long O-alkyl-O chain]	(0.81, 2.0)		(0.08, 0.28)	
[structure: bis-amidine cyclohexylidene with isopropoxy groups, long alkenyl chain]	(1.0, 7.3)		(0.16, 3.5)	

Figure 9. Compounds designed to illustrate use of the pharmacophore data of Figure 4 and the CoMSIA molecular descriptor fields of Figures 7 and 8, as well as respective descriptor fields of Supplementary Figures S1 and S2. The ranges are the smallest and largest predicted values obtained from the models constructed of rigid and flexible compound structures.

of the model and that the models constructed from molecular descriptors of flexible compounds may be predicting just as well, if not better, than those constructed from molecular descriptors of rigid compounds (Fig. 6). The rigid models of Figure 6 produce a greater amount of under prediction than the flexible models. For example, for the rigid model, one of the compounds active against *L. donovani* has an experimentally determined pIC_{50} value of -1.7 and a predicted value of 0, these IC_{50} values are 50 and 1, respectively, whereas for the flexible model the same compound has a predicted pIC_{50} value of -1.7, the same as the experimental value.

Under prediction is a problem that needs to be addressed since predictive models such as the ones constructed in this study can be employed to scan potential candidates for synthetic drug design. Often synthetic measures are costly and time consuming; hence, it is better to synthesize only compounds expected to have better inhibitory activity. To minimize under prediction, the minimum pIC_{50} predictions from the rigid or flexible models were plotted against the average experimental values. These data are shown as the conservative predictions of Figure 6. In this column we see that under predictions are no longer occurring for these models, yet there are still over predictions. Over predictions, as long as they are few, are not as problematic since these values are larger and synthesis will most likely result in experimentally determined inhibitory activity better than calculated.

4.3. Molecular contributions

A previous study focused on synthetic phospholipids that employed CoMSIA molecular descriptors found that steric and hydro-

phobic interactions governed the model.[20] Similarly, our model was governed by hydrophobic interactions; yet, this contribution of molecular descriptor interactions was followed by H-donor, H-acceptor, steric and then electrostatic (Table 2). When the overall CoMSIA potentials for each molecular descriptor are visualized, it is easier to design modifications to the compounds. Figure 7 allows for comparison between the models and overall analyzes. It is important to realize that molecular descriptor potentials are unique to each model; hence, no two models are the same. Positive electrostatic potentials indicate the importance of the N=C–N groups, whereas the steric and hydrophobic potentials show the importance of the rings and substituents. To fully understand molecular descriptor contribution in relation to biological inhibitory data, it is important to analyze the potentials of individual compounds, a selected set of which are shown in Figure 8. These potentials display much more consistency than those for the overall models of Figure 7.

4.4. Employing results

An exciting feature of this type of study is that new compounds can be designed by employing the data acquired from the pharmacophore (Fig. 4) and extrapolated molecular descriptor potentials (Figs. 7 and 8). To do this, the basic pharmacophore must remain intact and the potentials of the overall models and those of individual compounds should be used for guidance. Since the importance of hydrophobic, H-donor and H-acceptor atoms are clearly displayed as essential potentials for the individual compounds in Figure 8, this is a good place to begin. The favored hydrophobic

potentials of the four aromatic rings exhibit significance (Figs. 7 and 8); and these are also seen as important in the pharmacophore (Fig. 4). Hence, it appears imperative that the four rings remain a constant in our initial modeling efforts. H-bond donors appear to be important to regions near the N=C−N groups (Figs. 7 and 8). One of the N=C−N groups is shown as essential in the pharmacophore (Fig. 4). Likewise, H-bond acceptors appear to be significant to the region including and between the N=C−N groups and the N of the outer most aromatic rings (Figs. 7 and 8). One such region was identified in the pharmacophore (Fig. 4). Based on these observations, new compounds have been designed and activity predictions have been obtained (Fig. 9). The ranges include the smallest and largest predictions obtained via the models constructed of rigid and flexible compound structures. These are interesting new structure types will be synthesized in due course.

5. Conclusion

For this study, a biological dataset of synthetic arylimidamides and their activities against *L. donovani* axenic amastigotes and *L. amazonensis* intracellular parasites were employed for pharmacophore and three-dimensional QSAR modeling. The pharmacophore alignment displayed important features, and the three-dimensional QSAR was performed with and without these findings. The data from the models developed are being used to design novel compounds; whereas the models themselves are being implemented to estimate activities for known and designed compounds of interest. In summary, by employing three-dimensional QSAR it is possible to scan for potentially active compounds both efficiently and conservatively through the use of predictive models. The models are used to design compounds and respective IC_{50} activities can be predicted.

Acknowledgments

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2011.06.026.

References and notes

1. Alvar, J.; Yactayo, S.; Bern, C. *Trends Parasitol.* **2006**, *22*, 552.
2. Boelaert, M.; Meheus, F.; Sanchez, A.; Singh, S.; Vanlerberghe, V.; Picado, A.; Meessen, B.; Sundar, S. *Tropical Med. Int. Health* **2009**, 14.
3. Adhikari, S.; Maskay, N.; Sharma, B. *Health Policy Plan.* **2009**, *24*, 129.
4. Sundar, S.; More, D.; Singh, M.; Singh, V.; Sharma, S.; Makharia, A.; Kumar, P.; Murray, H. *Clin. Infect. Dis.* **2000**, *31*, 1104.
5. Sundar, S.; Chakravarty, J.; Agarwal, D.; Rai, M.; Murray, H. *N. Eng. J. Med.* **2010**, *362*, 504.
6. Sundar, S.; Agrawal, N.; Arora, R.; Agarwal, D.; Rai, M.; Chakravarty, J. *Clin. Infect. Dis.* **2009**, *49*, 914.
7. Meheus, F.; Balasegaram, M.; Olliaro, P.; Sundar, S.; Rijal, S.; Faiz, M.; Boelaert, M. *PLoS Negl. Trop. Dis.* **2010**, *4*, e818.
8. Sundar, S.; Murray, H. *Bull. World Health Organ.* **2005**, *83*, 394.
9. Moore, E.; Lockwood, D. *J. Glob. Infect. Dis.* **2010**, *2*, 151.
10. van Griensven, J.; Balasegaram, M.; Meheus, F.; Alvar, J.; Lynen, L.; Boelaert, M. *Lancet Infect. Dis.* **2010**, *10*, 184.
11. González, U.; Pinart, M.; Reveiz, L.; Rengifo-Pardo, M.; Tweed, J.; Macaya, A.; Alvar, J. *Clin. Infect. Dis.* **2010**, *51*, 409.
12. Wortmann, G.; Zapor, M.; Ressner, R.; Fraser, S.; Hartzell, J.; Pierson, J.; Weintrob, A.; Magill, A. *Am. J. Trop. Med. Hyg.* **2010**, *83*, 1028.
13. Blum, J.; Desjeux, P.; Schwartz, E.; Beck, B.; Hatz, C. *J. Antimicrob. Chemother.* **2004**, *53*, 158.
14. Mishra, J.; Saxena, A.; Singh, S. *Curr. Med. Chem.* **2007**, *14*, 1153.
15. Mahajan, V. K.; Sharma, N. L. *J. Dermatolog. Treat.* **2007**, *18*, 97.
16. Al-Mutairi, N.; Alshiltawy, M.; El Khalawany, M.; Joshi, A.; Eassa, B. I.; Manchanda, Y.; Gomaa, S.; Darwish, I.; Rijhwani, M. *Int. J. Dermatol.* **2009**, *48*, 862.
17. Rosenblatt, J. E. *Mayo Clin. Proc.* **1999**, *74*, 1161.
18. Nussbaum, K.; Honek, J.; Cadmus, C. M.; Efferth, T. *Curr. Med. Chem.* **2010**, *17*, 1594.
19. Avery, M. A.; Muraleedharan, K. M.; Desai, P. V.; Bandyopadhyaya, A. K.; Furtado, M. M.; Tekwani, B. L. *J. Med. Chem.* **2003**, *46*, 4244.
20. Kapou, A.; Benetis, N. P.; Avlonitis, N.; Calogeropoulou, T.; Koufaki, M.; Scoulica, E.; Nikolaropoulos, S. S.; Mavromoustakos, T. *Bioorg. Med. Chem.* **2007**, *15*, 1252.
21. Ryu, C. K.; Lee, Y.; Park, S. G.; You, H. J.; Lee, R. Y.; Lee, S. Y.; Choi, S. *Bioorg. Med. Chem.* **2008**, *16*, 9772.
22. Sanders, J. M.; Gomez, A. O.; Mao, J.; Meints, G. A.; Van Brussel, E. M.; Burzynska, A.; Kafarski, P.; Gonzalez-Pacanowska, D.; Oldfield, E. *J. Med. Chem.* **2003**, *46*, 5171.
23. Santos, D. O.; Coutinho, C. E.; Madeira, M. F.; Bottino, C. G.; Vieira, R. T.; Nascimento, S. B.; Bernardino, A.; Bourguignon, S. C.; Corte-Real, S.; Pinho, R. T.; Rodrigues, C. R.; Castro, H. C. *Parasitol. Res.* **2008**, *103*, 1.
24. Liu, M.; Wilairat, P.; Croft, S. L.; Tan, A. L.; Go, M. L. *Bioorg. Med. Chem.* **2003**, *11*, 2729.
25. Arafa, R. K.; Brun, R.; Wenzler, T.; Tanious, F. A.; Wilson, W. D.; Stephens, C. E.; Boykin, D. W. *J. Med. Chem.* **2005**, *48*, 5480.
26. Munde, M.; Lee, M.; Neidle, S.; Arafa, R.; Boykin, D. W.; Liu, Y.; Bailly, C.; Wilson, W. D. *J. Am. Chem. Soc.* **2007**, *129*, 5688.
27. Stephens, C. E.; Tanious, F.; Kim, S.; Wilson, W. D.; Schell, W. A.; Perfect, J. R.; Franzblau, S. G.; Boykin, D. W. *J. Med. Chem.* **2001**, *44*, 1741.
28. Wilson, W. D.; Tanious, F. A.; Mathis, A.; Tevis, D.; Hall, J. E.; Boykin, D. W. *Biochimie* **2008**, *90*, 999.
29. Werbovetz, K. A.; Sackett, D. L.; Delfin, D.; Bhattacharya, G.; Salem, M.; Obrzut, T.; Rattendi, D.; Bacchi, C. *Mol. Pharmacol.* **2003**, *64*, 1325.
30. Wang, M. Z.; Zhu, X.; Srivastava, A.; Liu, Q.; Sweat, J. M.; Pandharkar, T.; Stephens, C. E.; Riccio, E.; Parman, T.; Munde, M.; Mandal, S.; Madhubala, R.; Tidwell, R. R.; Wilson, W. D.; Boykin, D.; Hall, J. E.; Kyle, D. E.; Werbovetz, K. A. *Antimicrob. Agents Chemother.* **2010**, *54*, 2507.
31. Delfin, D. A.; Morgan, R. E.; Zhu, X.; Werbovetz, K. A. *Bioorg. Med. Chem.* **2009**, *17*, 820.
32. Buckner, F. S.; Wilson, A. J. *Am. J. Trop. Med. Hyg.* **2005**, *72*, 600.
33. SYBYL Molecular Modeling Software, 8.1 ed., Tripos Inc.: St. Louis, MO, 2008.
34. Golbraikh, A.; Tropsha, A. *J. Mol. Graphics Modell.* **2002**, *20*, 269.