Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry
ELSEVIER
Bioorganic & Medicinal Chemistry 14 (2006) 3144–3152

3D QSAR on a library of heterocyclic diamidine derivatives with antiparasitic activity

Prashanth Athri,[a] Tanja Wenzler,[b] Patricia Ruiz,[a] Reto Brun,[b] David W. Boykin,[a] Richard Tidwell[c] and W. David Wilson[a,*]

[a]*Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA*
[b]*Swiss Tropical Institute, Department of Medical Parasitology and Infection Biology, CH-4002 Basel, Switzerland*
[c]*Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA*

Received 20 September 2005; revised 24 November 2005; accepted 16 December 2005
Available online 25 January 2006

Abstract—African trypanosomes, *Trypanosoma brucei rhodesiense* (TBR) and *Trypanosoma brucei gambiense* (TBG), affect hundreds of thousands of lives in tropical regions of the world. The toxicity of the diamidine pentamidine, an effective drug against TBG, necessitates the design of better drugs. An orally effective prodrug of the diamidine, furamidine (**DB75**), presently scheduled for phase III clinical trials, has excellent activity against TBG with toxicity lower than that of pentamidine. As part of an effort to develop additional and improved diamidines against African trypanosomes, CoMFA and CoMSIA 3D QSAR analyses have been conducted with furamidine and a set of 25 other structurally related compounds. Two different alignment strategies, based on a putative kinetoplast DNA minor groove target, were used. Due to conserved electrostatic properties across the compounds, models that used only steric and electronic properties did not perform well in predicting biological results. An extended CoMSIA model with additional descriptors for hydrophobic, donor, and acceptor properties had good predictive ability with a $q^2 = 0.699$, $r^2 = 0.974$, SEE, standard error of estimate = 0.1, and $F = 120.04$. The results have been used as a guide to design compounds that, potentially, have better activity against African trypanosomes.

1. Introduction

Infectious diseases such as trypanosomiasis, malaria, and leishmaniasis, which are spread by protozoan parasites, infect millions of people throughout most of the world.[2,3] There is now a serious epidemic phase of several of these diseases due to factors that include vector spread, travel of infected individuals, civil wars and, in particular, lack of available and effective drugs for treatment. These diseases strike all age groups and severely limit the health and economic outlook in infected regions. Since the synthesis of the aromatic diamidine, pentamidine (Fig. 1), and the discovery of its broad antiparasitic activities, amidines have been of interest for development of antiparasitic compounds.[1–5] Although pentamidine has had significant clinical success, its toxicity, lack of oral availability, and the appearance of pentamidine-resistant organisms underscore the essential need to develop additional drugs for treatment.[1,4] The recent synthetic preparation and clinical success of orally effective diamidine prodrugs make new synthetic diamidine compounds an important group for discovery of additional and improved drugs against diseases due to protozoan parasites.[1,5–7]

Although the mechanism of action of pentamidine is not clearly established and more than one biological target may be involved, current evidence supports an essential

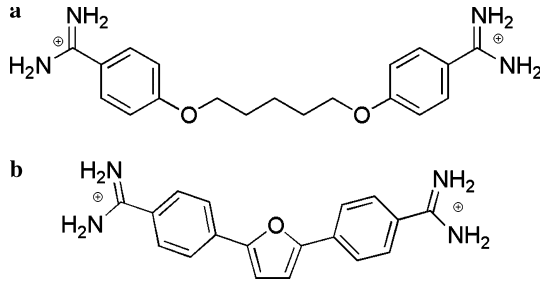


Figure 1. (a) Pentamidine; (b) **DB75**.

Keywords: Heterocyclic diamidines; 3D QSAR; Anti-parasitic; CoMFA/CoMSIA.
* Corresponding author. Tel.: +1 404 651 3903; fax: +1 404 651 2751; e-mail: wdw@gsu.edu

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.12.029

DNA binding step in the biological activity of diamidine derivatives that target infectious disease organisms.[1,8] In this model, heterocyclic dications form a complex in the minor groove of AT-rich DNA sequences and selectively perturb the action of one or more microbial enzymes and/or transcription factors that must act on DNA in the target organisms.[8–13] Analogues that bind poorly to DNA generally display poor biological activity.[1,8] Amidines have selective uptake systems in target organisms, such as trypanosomes, and fluorescence microscopy studies show that the amidines are localized to DNA-rich regions.[8,14,15] In organisms such as leishmania and trypanosomes, the mitochondrial kinetoplast DNA (kDNA) is the initial target of diamidines.[8,16,17] The DNA in most parasitic microorganisms is AT rich and in some organelles, such as the trypanosomal kinetoplast, repeated AT tract sequences form a selective and susceptible target for heterocyclic diamidines.[17] Compounds that bind selectively to AT sequences, thus, have an advantage in targeting the kDNA of these organisms.

Given these intriguing links to DNA targeting in the action of diamidines, the long and successful history of use of pentamidine in humans, and the recent ability to synthesize prodrugs of diamidines, compounds of this type are very attractive for rational drug development. Structurally, pentamidine is a highly flexible molecule that can assume an array of linked conformations related through torsional rotation (Fig. 1). A strategy to remove much of the torsional freedom of pentamidine and preorganize the molecular structure for binding to the DNA minor groove is to replace the alkyldiether linking group of pentamidine with a five-membered heterocycle (Fig. 1). The prototype for this type of structure is furamidine, **DB75**, and an orally active prodrug of this compound, **DB289**, has successfully completed phase II clinical trials in first stage sleeping sickness patients and revealed low host toxicity.[1,5–8,18]

The details of the biological action of diamidines are far from understood and we are seeking additional methods to probe their structure–activity relationships to assist in rational drug design. It seems certain that activity will depend strongly on the molecular structure and chemical properties of the compounds, and how these match the DNA minor groove receptor site. An X-ray structure of furamidine bound to a model system for kDNA AT sequences is available and supports the DNA minor groove as a key component in the cellular therapeutic target of antiparasitic diamidines.[19] A library of diphenyldiamidines with central five-membered rings has been prepared and antitrypanosomal biological testing conducted. Several of these derivatives have also been crystallized with the same AT DNA sequence as with furamidine.[20] This wealth of pertinent structural information and the availability of accurate biological testing data present an attractive opportunity for the use of 3D QSAR methods. Efforts have been concentrated on trying to understand the underlying correlations between various chemical descriptors and biological activity to help predict new synthesis directions for preparation of improved derivatives. Other examples of such a protocol are.[21,22] To initiate this study, we have carried out CoMFA and CoMSIA-based 3D QSAR studies on all available diphenyldiamidines in our library that have a central five-membered ring system and for which antitrypanosomal testing results were available. The goals of this study are: (1) to correlate the structural features of this class of compounds with their biological activity by using our extensive knowledge of the DNA minor groove bioreceptor in AT sequences and (2) to use this information to predict new compounds for synthesis that have a high probability of enhanced activity.

1.1. Application of 3D QSAR to diamidines

CoMFA and CoMSIA techniques[23,24] use field-based descriptors to generate contour maps that provide a visual rendering of the molecular properties that are important to biological activity. Starting with the training set, organized in a specific alignment with each molecule assumed to be in its 'bio-active'[25] conformation, probe atoms are used to define a field, as represented by points of a 3D grid of user-chosen density around each atom of the molecule. The fields calculated for each molecule at each grid point in the alignment are correlated with the biological activity. The output is represented by 3D contour surfaces that represent relative spatial contributions of the fields around the molecular alignment. The surfaces are obtained by correlating the calculated fields to experimentally measured biological activity using PLS, partial least squares.[26]

One of the major differences between CoMFA and CoMSIA is the way the fields are calculated to describe the environment around an aligned set of molecules. In CoMFA, the steric fields are calculated using Lennard-Jones potential and the electrostatic fields are calculated using Coulomb potentials.[21] CoMSIA uses a smoother Gaussian function to calculate the same fields. Three additional similarity fields as implemented by Tripos' default version of CoMSIA namely, hydrophobic, donor, and acceptor fields,[27] were also used in our work.

It has been observed in CoMFA calculations that steep curves near van der Waals surface can result in unrealistically large changes in calculated values,[28,29] and to avoid these large values arbitrary cut-offs are needed during potential calculations. The potentials calculated using the two different functions result in these cut-offs being at different distances for different terms[29] and can result in disjointed contour maps. Finally, it has been observed that CoMFA is quite sensitive to changes in alignment with respect to the grid. These problems are reduced in the case of CoMSIA by using a flatter Gaussian type function that is defined by Eq. 1:

$$A_{F,k}^{q}(j) = \sum_{i} w_{\text{probe},k} w_{ik} e^{-\alpha r_{kq}^{2}}, \qquad (1)$$

where A denotes the similarity index, used to calculate each of the types of fields, at each grid point q. The summation is over all points, i, of the molecule under investigation, j. $w_{\text{probe},k}$ is the probe atom with a user selected radius. w_{ik} is the actual value of the physicochemical property k of atom i and r_{ik} is the mutual distance between the probe atom at grid point q and atom i in

the test molecule. α is the attenuation factor with a default value of 0.3. Previous studies[30] suggest that the optimal value lies between 0.2 and 0.4 with larger values resulting in steeper Gaussians, and hence larger values result in functions that resemble those used in CoMFA.

2. Materials and methods

2.1. Dataset selection

The molecular set used in this study includes all compounds with a 6–5–6 unfused heterocyclic aromatic system such as furamidine (Fig. 1) for which antitrypanosomal in vitro and in vivo results were available. The list of compounds is presented in Table 1, along with their associated biological activities represented by IC_{50} (concentrations in micromolars) in the form of pIC_{50} ($-\log IC_{50}$). Twenty two of the 26 compounds were synthesized earlier on in the study. Initial biological testing of each of these 22 compounds was performed soon after its synthesis and testing was done over an extended period as compounds were prepared after each other. The promising results encouraged us to perform additional biological tests in quadruplet under carefully controlled conditions to enhance data quality and comparison reliability. A plot of initial testing results against the average of the four recent results, however, shows good agreement with Pearson's correlation coefficient, *R*, being 0.951 (Supplementary material, Figure A). We were able to test the four compounds synthesized later, **DB240**, **DB484**, **DB690**, and **1RJL164** once. Nevertheless, as indicated by the average standard deviation between trials (0.003 μM, see Supplementary material) we are certain of the integrity of these values.

Five of the compounds in this dataset have solved X-ray crystal structures, all from the Neidle laboratory,[19,20] for their DNA complexes. Coordinates are available at the RCSB Protein Data Bank[31] (**DB193**: PDB ID-298D, **DB244**: PDB ID-1EEL, **DB249**: PDB ID-1FMS, **DB313**: PDB ID-1FMQ, and **DB75**: PDB ID-227D). All five compounds preserve the 6–5–6 ring core (note that Fig. 2 is a pictorial representation of the common sub-structure). This preserved ring system formed the basis of our selection of compounds. Internal substitutions in the rings by various hetero atoms were allowed but exocyclic substituents were not permitted since the extrapolation of X-ray structures, for the purpose of molecular minimization constraints, would be uncertain. These restrictions ensured that the crystal structures could be used as reasonable approximations to the 'bio-active'[25] conformation of the respective molecules. Also, these restrictions provide an optimum method to produce results that are interpretable in terms of structure and specific molecular features for compound design.

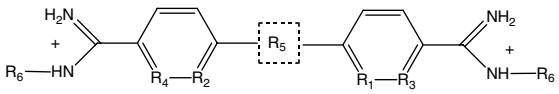
2.2. Determination of in vitro activity against *Trypanosoma brucei rhodesiense* (TBR)

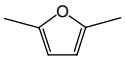
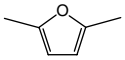
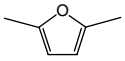
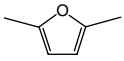
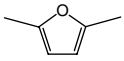
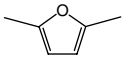
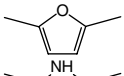
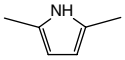
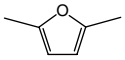
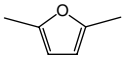
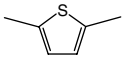
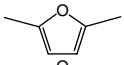
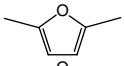
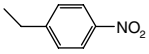
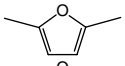
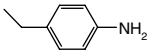
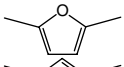
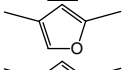
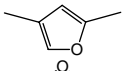
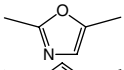
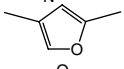
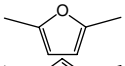
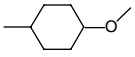
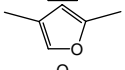
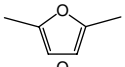
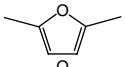
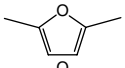
Minimum essential medium (50 μl) supplemented according to Baltz et al.[32] with 2-mercaptoethanol and 15% heat-inactivated horse serum was added to each well of a 96-well microtiter plate. Serial drug dilutions were prepared covering a range from 1 to 0.0014 μg/ml. Then 2×10^3 bloodstream forms of TBR STIB 900 in 50 μl were added to each well and the plate was incubated at 37 °C under a 5% CO_2 atmosphere for 70 h. Ten microliters of Alamar Blue (12.5 mg resazurin dissolved in 100 mL phosphate-buffered saline, PBS) was then added to each well and incubation was continued for a further 2–4 h. The plate was then read in a Spectramax Gemini XS microplate fluorometer (Molecular Devices Cooperation, Sunnyvale, CA, USA) using an excitation wavelength of 536 nm and an emission wavelength of 588 nm.[33] Fluorescence development was measured and expressed as percentage of the control. Data were transferred into the graphic program Softmax Pro (Molecular Devices) which calculated the IC_{50} values. Cytotoxicity was assessed using the same assay procedure but with rat skeletal myoblasts (L-6 cells).

2.3. Molecular modeling and geometric optimization

QSAR analysis and molecular modeling were performed using the SYBYL 6. 9. 2[34] software package from Tripos on an SGI O_2 machine. Crystal structures are available for five molecules, **DB75**, **DB193**, **DB244**, **DB249**, and **DB313**, in a DNA complex with an AATT DNA sequence. Although the crystal structures are similar, there are differences due to the fit of each compound into the minor groove of DNA. The conformations of each of the 21 molecules for which no crystal structure is available were constrained to adopt the binding conformation of one of the crystal templates. Two factors were considered when deciding which crystal structure would act as a template for each molecule: (1) structural similarity; (2) correlation with activity. For example, **DB867** would use torsional restraints of **DB75** due to its similarity in structure (see Table 1). Some molecules were structurally similar to more than one template. For example, **DB181** is similar to **DB193** through **DB313** in the list of crystal structures mentioned above. In such cases, the correlation of activity between the compound and the template molecules was used to select the most appropriate template. With reference to Figure 2, harmonic torsional constraints from the X-ray templates were placed on torsional angles *T1* ($1 \rightarrow 3 \rightarrow 4 \rightarrow 5$), *T2* ($6 \rightarrow 7 \rightarrow 10 \rightarrow 11$), *T3* ($11 \rightarrow 12 \rightarrow 15 \rightarrow 16$), and *T4* ($17 \rightarrow 18 \rightarrow 21 \rightarrow 22$). Table 1 lists all the compounds used in this study. The template column indicates which template was used in that case and 'X' indicates that the X-ray structure was available for the molecule in question. The cationic terminal substituents were allowed to keep the low energy conformations they adopted upon energy minimization. In cases where the cationic substituents assumed conformations that would hinder interactions with the minor groove, rotatable bonds were used to manually change the conformation. The resulting structure was then re-minimized to obtain a new low energy conformation. The rings were forced to be rigid to avoid out-of-plane bending within the ring. Once the constraints were applied, the molecules were re-minimized to a conformation that should be close to their respective DNA binding geometries. All mole-

Table 1. List of compounds[a]



Compd	Templt	R1	R2	R3	R4	R5	R6	Activity (pIC_{50})	Activity (μM)
DB75	X	–CH	–CH	–CH	–CH		–H	2.29	0.0051
DB181	DB249	–CH	–CH	–CH	–CH			1.16	0.0691
DB193	X	–CH	–CH	–CH	–CH			1.63	0.0234
DB235	DB244	–CH	–CH	–CH	–CH			1.45	0.0354
DB240	DB249	–CH	–CH	–CH	–CH			1.35	0.0446
DB244	X	–CH	–CH	–CH	–CH			1.45	0.0354
DB249	X	–CH	–CH	–CH	–CH			1.17	0.0676
DB262	DB75	–CH	–CH	–CH	–CH		–H	1.85	0.0141
DB312	DB193	–CH	–CH	–CH	–CH			1.63	0.0234
DB313	X	–CH	–CH	–CH	–CH			1.97	0.0107
DB351	DB75	–CH	–CH	–CH	–CH		–H	2.52	0.00302
DB417	DB75	–CH	–CH	–CH	–CH		—	2.18	0.0066
DB421	DB249	–CH	–CH	–CH	–CH			0.99	0.1023
DB422	DB249	–CH	–CH	–CH	–CH			1.01	0.0977
DB427	DB313	–CH	–CH	–CH	–CH			2	0.0100
DB480	DB193	–CH	–CH	–CH	–CH			1.41	0.0389
DB481	DB244	–CH	–CH	–CH	–CH			1.72	0.0190
DB484	DB75	–CH	–CH	–CH	–CH		–H	2.15	0.0070
DB518	DB244	–CH	–CH	–CH	–CH			1.86	0.0138
DB568	DB249	–CH	–CH	–CH	–CH			0.77	0.1698
DB690	DB75	–CH	–CH	–CH	–CH		–H	2.77	0.0016
DB820	DB75	N	–CH	–CH	–CH		–H	2.31	0.0048
DB829	DB75	N	N	–CH	–CH		–H	1.77	0.0169
DB867	DB75	–CH	–CH	N	–CH		–H	2.68	0.0020
DB994	DB75	–CH	–CH	N	N		–H	2.39	0.0040
1RJL164	DB75	–CH	–CH	–CH	–CH		–H	2.28	0.0052

[a] Compd: list of compounds used. Templt: Conformations of compounds that have 'X' served as templates to other molecules, since they had solved X-ray structures available; R1 through R6: respective substituents with respect to template shown above the table; Activity (pIC_{50})/activity (μM): biological activity, measured in micromolars, and represented in pIC_{50} ($-\log IC_{50}$).

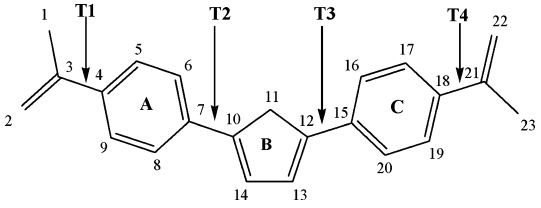


Figure 2. Indexed core to reference atom positions and rings.

cules were built using standard Tripos-SYBYL force field parameters.[35] Optimizations were performed to completion, using a distance-dependent dielectric and the BFGS algorithm.[36] The convergence limit was set to 0.001 kcal/mol. Each torsional constraint was given a 2 kcal penalty. The Geisteiger–Huckel method was used for charge calculations. All CoMFA and CoMSIA analyses were done by using SYBYL default parameters. Column filtering was set at 2 kcal/mol. Protein Explore[37] was used to generate contact surface maps of the nucleic acid–ligand complex.

2.4. Mutual alignment

A crucial decision in 3D QSAR studies is that of aligning the molecules so that their 3D conformation resembles their 'bio-active'[25] conformation. The alignment of molecules with respect to each other was done using two rigid body approaches: (1) atoms 11, 14, and 13 (Fig. 1) of all molecules were used for RMS fitting on corresponding atoms on **DB867** (most active compound). (2) Each molecule was aligned to the template by rotation and translation with an objective of minimizing the RMSD between atoms 3, 4, 7, 10, 12, 15, 18, and 21 (atomic positions with respect to Fig. 2) using DATABASEALIGN tool in SYBYL. Again, **DB867** was used as the template. Both these alignments were used to generate CoMFA and CoMSIA models. Models obtained from manual RMS alignment (as in one above) are referred to as M1-CoMFA/CoMSIA, and models obtained from alignments using the DATABASEALIGN tool in SYBYL (as in two above) are referred to as M2-CoMFA/CoMSIA.

3. Results and discussion

3.1. Biological assays

As described above, biological testing of the compound set against *Trypanosoma brucei rhodesiense* (TBR) was initially done after compound synthesis. As part of this study, the testing was repeated four times to allow statistical analysis of the variation and to obtain the most accurate biological data for QSAR studies. The results obtained in the two testing sets are very similar both qualitatively (as in the QSAR models they generated) and quantitatively (Supplementary materials), with a Pearson's correlation coefficient of 0.951. An average of the pairwise (2, 1; 3, 2; and 4, 3) Pearson's correlation coefficient was 0.936 suggesting the high reproducibility and accuracy of the biological testing data. Hence, four compounds synthesized later, and biological assays were performed once, were also included in the study. An average of the four tests was used for the 22 compounds synthesized earlier.

3.2. Statistical analysis of CoMFA/CoMSIA models

The two alignments, as detailed above, were used to compare the results obtained by CoMFA and CoMSIA methods. To make an initial comparison, electrostatic and steric fields were used to describe the biological testing results. Four models, M1-CoMFA, M1-CoMSIA(II), M2-CoMFA, and M2-CoMSIA(II), were generated (Section 3.2.1; Table 2). Both alignments were also used to generate two CoMSIA models, M1-CoMSIA(V), M2-CoMSIA(V), with electrostatic, steric, hydrophobic, acceptor, and donor descriptors (Section 3.2.2; Table 2). Both CoMFA and CoMSIA use PLS[26] analyses to compute the predictive models. The number of components used in a PLS analysis is an index of the degree of complexity of the model. The model that uses the minimum set of components required to describe the dataset is always preferred over models with higher dimensionality. The number of principal components sufficient to explain activity is calculated using the SAMPLS[38] routine in SYBYL. This is based on cross-validated results using only the independent variables (for faster processing) and this value is used in the final model.

3.2.1. Comparison of CoMFA and CoMSIA models. Comparison of M1-CoMSIA and M2-CoMSIA (Table 2) indicates that the CoMSIA method gave results that were relatively insensitive to the alignment used. Table 3 summarizes the electrostatic and steric contributions to the overall activity, according to the various models. The most statistically significant model among all six, M2-CoMSIA(V) (Section 3.2.2; Table 2), suggests a realistic representation of the weights that can be placed on each descriptor. As compared to this model, both M1-CoMSIA(II) and M2-CoMSIA(II), with two descriptors, overemphasized the importance of the electrostatic descriptor (Table 3). On the other hand, CoMFA models gave a realistic picture of the actual weights that could be placed on the electrostatic field, as observed earlier. As observed in,[27] CoMSIA models facilitate the distribution of variance across H-bonding fields, while maintaining spatial context with respect to compound design. To enhance our capability to better visu-

Table 2. CoMFA/CoMSIA statistical results[a,b]

	q^2	r^2	No. of com	SE	F
M1-CoMFA	0.66	0.772	1	0.267	81.213
M1-CoMSIA(II)	0.673	0.764	1	0.271	77.58
M2-CoMFA	0.656	0.84	2	0.228	60.57
M2-CoMSIA(II)	0.662	0.757	1	0.275	74.7
M1-CoMSIA(V)	0.682	0.969	6	0.111	97.852
M2-CoMSIA(V)	0.699	0.974	6	0.1	120.038

[a] q^2 = leave-one-out cross-validated r^2 value; r^2 = non-cross-validated regression coefficient; No. of com = number of components; SE = standard error; F = F-statistic.

[b] M1-CoMFA/M1-CoMSIA(II/V): CoMFA/CoMSIA models derived from model-1; M2-CoMFA/M2-CoMSIA(II/V): CoMFA/CoMSIA models derived from model-2.

Table 3. CoMFA/CoMSIA field contributions[a]

	Steric	Electrostatic	Hydrophobic	Donor	Acceptor
M1-CoMFA	0.983	0.017	—	—	—
M1-CoMSIA (II)	0.839	0.161	—	—	—
M2-CoMFA	0.972	0.028	—	—	—
M2-CoMSIA (II)	0.852	0.148	—	—	—
M1-CoMSIA (V)	0.14	0.076	0.271	0.339	0.175
M2-CoMSIA (V)	0.134	0.071	0.279	0.348	0.168

[a] M1-CoMFA/M1-CoMSIA(II/V): CoMFA/CoMSIA models derived from model-1; M2-CoMFA/M2-CoMSIA(II/V): CoMFA/CoMSIA models derived from model-2.

alize the contribution of various functional groups, toward activity, we performed CoMSIA with the five descriptors defined above, that help in partitioning the property fields with respect to the 3D grid.

3.2.2. CoMSIA models with five descriptors. We performed CoMSIA using various values of the attenuation factor (α) Eq. 1 ranging from 0.2 to 0.8 and found relatively small differences in the predictive quality of the models. Hence, for this particular dataset, the addition of hydrophobic, donor, and acceptor fields was primarily responsible for the better predictive results obtained using CoMSIA as opposed to the difference in the functions used to calculate the distance-dependence between the probe atom and the molecule atoms. CoMSIA models with five descriptors had low residuals for activity predictions (Fig. 2). It was observed that most compounds had appreciably lower residuals as compared to CoMFA models (not shown). All compounds were used in the CoMSIA since they provide useful information to the overall model without compromising the reliability of the models (as suggested by high q^2 and r^2 values). Similar to the two descriptor CoMSIA models, there were no significant differences in the model based on the alignment used. model-2 ($r^2 = 0.974$, $q^2 = 0.699$) performed slightly better than model-1 ($r^2 = 0.969$, $q^2 = 0.682$). The analysis in Section 3.3 is with reference to M2-CoMSIA(V) and a plot of the actual versus the predicted activity values is presented in Figure 3 for this model.

The reason for larger errors (Table 2) in models using two descriptors can be rationalized by analyses of the compounds and descriptors. All compounds studied have two amidine groups and a +2 charge that is centered on the amidines. Both M1-CoMSIA(V) and M2-CoMSIA(V) suggest lower weights assigned to electrostatic effects (Table 3). This is expected since the positively charged amidine groups are common to all compounds. It has been shown elsewhere[1] that removal of one or both charged groups results in the loss of activity. The electrostatic effects come essentially completely from these two charged amidine-type groups and other electrostatic effects are relatively insignificant.

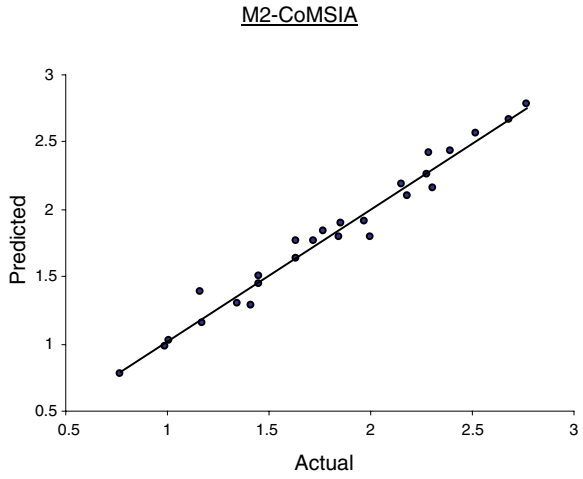


Figure 3. Predicted (*y*-axis) versus actual (*x*-axis) plots of M2-CoMSIA(V).

3.3. Analysis of contour maps

The M2-CoMSIA results gave the best statistical fit to the biological testing results and contour maps obtained from the M2 model are as shown in Figure 4. The results obtained from M1-CoMSIA are similar to M2-CoMSIA maps. The individual contributions from each of the five fields used in the final CoMSIA model are shown in Table 3. The positioning of donor atoms and hydrophobicity explains 62.7% of the biological activity. The small electrostatic contribution is expected because of the constancy of the terminal cationic amidine groups.

3.3.1. Analysis of steric field maps. The contour maps for steric fields (Figs. 4a and b) are particularly helpful in explaining the variation in biological activity for compounds with low activity that have differences in substitutions on the amidine groups. The presence of large moieties in the yellow regions correlates well with low activity. The green regions display areas in the 3D space that, when occupied, encourage higher biological activity. The lower activity of **DB249** (Fig. 4a) and **DB568** is correlated with the presence of side chains in the unfavorable yellow regions. As expected, the activities of the compounds improve as the overlap of this region and the substituents decrease in size. This can be seen in the case of **DB518** (Fig. 4b) and **DB312** that have large substituents in the green region but low overlap with the yellow region. The presence of green and yellow regions at close proximity presents a design problem that has been solved with some compounds in this library.

3.3.2. Analysis of hydrophobic field maps. The hydrophobic maps (Fig. 4c) indicate that the presence of hydrophilic atoms, such as oxygen, is not favored at the center region of the inner face of the molecule that interacts with base pairs at the bottom of the minor groove. The yellow region indicates domains in space that favor hydrophobic substituents and the gray regions are places where hydrophilic groups are favored. These maps suggest that the presence of polar

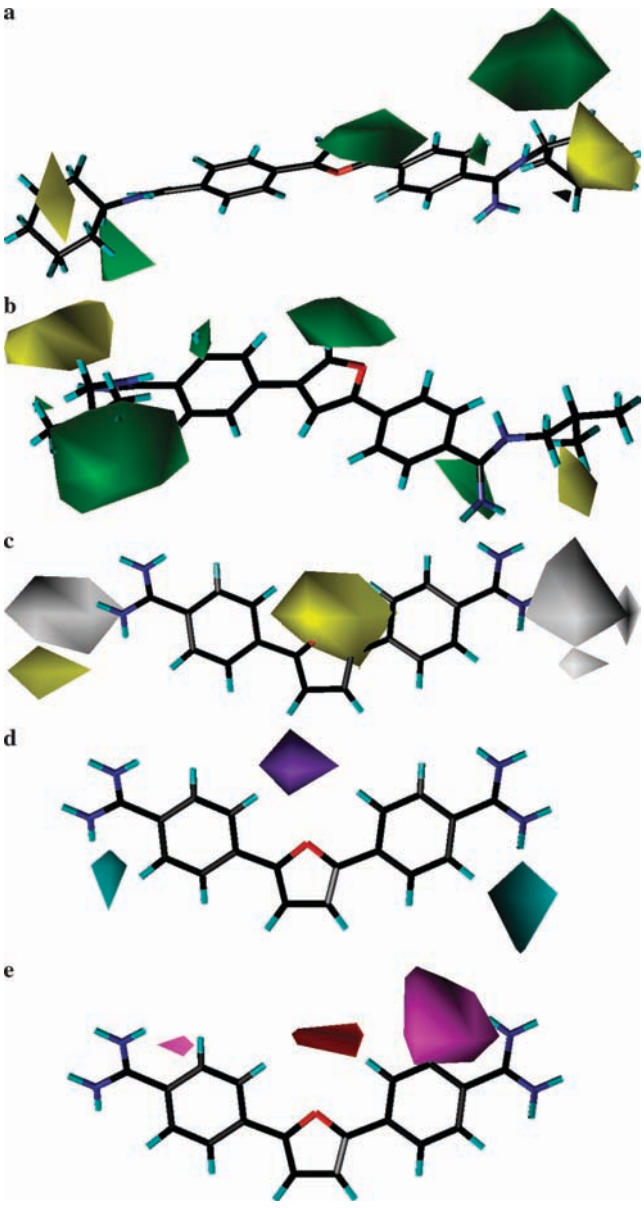


Figure 4. CoMSIA generated contour maps. (a) and (b) Steric maps of **DB518** and **DB568**. Yellow regions indicate areas where bulky side chains are not favored and green regions favor bulky side chains; (c) hydrophobicity maps: yellow regions favor the presence of hydrophobic atoms and gray regions favor hydrophilic substituents; (d) donor: cyan regions favor the presence of donors and purple regions disfavor the presence of donors; (e) acceptor: magenta regions favor the presence of acceptors and red regions do not.

atoms at position 11 (Fig. 1) is detrimental to activity. **DB351** has sulfur (thiophene) at the center and its activity is higher than for **DB75** with a central furan. This can be attributed to sulfur being less polar than oxygen since other components are the same for these two compounds. The large hydrophilic regions (gray in Fig. 4c) at both ends of the structure probably indicate exposed parts of the molecule that are likely to interact with water and hence hydrophilic substituents are favorable to activity.

3.3.3. Analysis of donor and acceptor field maps. Donor and acceptor maps are shown in Figures 4d and e. In Figure 4d, cyan and magenta regions favor the presence of donor and acceptor substitutions in the respective domains they represent. Similarly, the purple and red regions show positions that should not be occupied by donor and acceptor atoms. The cyan regions suggest that the amidines are involved in hydrogen-bonding interactions. The purple regions indicate that atoms that are able to participate in donor interactions should not be present at position 11 (Fig. 1). This agrees with the hydrophobic maps and this observation is exemplified in the case of **DB262** that has NH (pyrrole) at the center position. The activity of **DB262** decreases to 1.8 as compared to 2.4 with **DB75**.

As with the hydrophobic maps, acceptor contour maps do not favor the presence of acceptors at position 11 (small red region at the center). The magenta areas indicate regions where acceptor atoms are favored. With respect to a planar ring geometry and considering **DB867** and **DB994**, this suggests that the activity is improved when nitrogens are on one side of the molecule across the vertical symmetry. Even though that part of the inference is clear, it is not obvious where the nitrogens are favored across the horizontal symmetry, that is, if N is preferred at position 5 or 9. Since the alignments are limited to one conformation, we had chosen the nitrogen to be facing the groove (position 5). If we had started with the nitrogen at position 9, the situation might have been reversed.

To further investigate this point, we generated contact surface maps using Protein Explorer.[37] Contact surfaces for the DNA–ligand complex, obtained from crystallographic structures, are shown in Figure 5. X-ray structures for five compounds complexed with DNA are available (Table 1) and contact surfaces for all five complexes were generated. Since the maps are similar in all cases, one of the five maps is shown in Figure 4 for reference. The pink surfaces (Fig. 5) show regions on the DNA minor groove that are close enough to the ligand to favor interactions. Also note the surface of DNA that is close enough to the ligand to participate in hydrophobic interactions. These regions are represented by light colors,

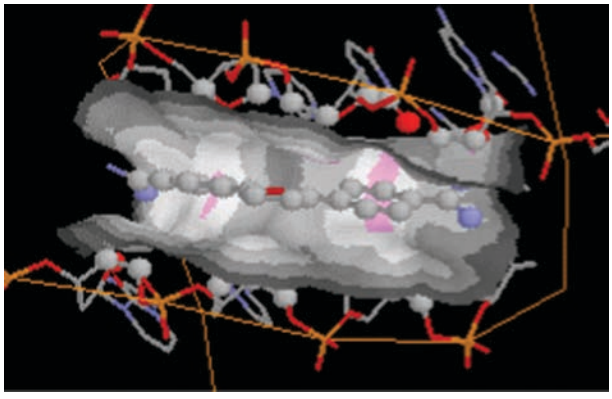


Figure 5. Contact surface maps of **DB75** (from Protein Explorer). Pink regions denote areas that are close enough to participate in hydrogen bond interactions and light regions (light gray and white) are regions that are close enough to encourage hydrophobic activity. Atoms represented by balls are within a 7 Å distance from the receptor.

namely, light gray and white. Regions close to the center (position 11 in Fig. 1) are light gray, and this agrees with conclusions derived from hydrophobic contour maps (Fig. 4c) which suggested that polar substituents are not favorable in this region. The two pink regions are close to position 17 (rather than 19) and position 5 (rather than 9) with reference to Figure 1. This explains why **DB867** and **DB994** have better activities than **DB820** and **DB829**. **DB820** and **DB829** have polar substituents in hydrophobic regions (white regions in Fig. 4), while **DB867** and **DB994** have nitrogen close to the pink regions. This suggests that the nitrogen atoms (positions 5 and 17, respectively) adopt a conformation that places them close to the minor groove. On the other hand, inspection of the AT base pair receptor environment in the compound binding site of the minor groove indicates that there are no donor groups for the formation of hydrogen bonds with the compound in this region. There are potential hydrogen bonding acceptors on the bases, the adenine nitrogen N3 and thymine O2 groups, but it is not clear how they would interact favorably with the unprotonated pyridine N. One possibility to consider, however, is that the pyridine ring N of the compound could be protonated when bound to DNA, as has been observed for other minor groove binding compounds with basic groups.[39] With the pyridine protonated the compound would become an H-bond donor group for interaction with the acceptors on the AT base pairs. Without further experiments with nitrogen permuted at various positions in the compounds, however, it is not possible to tell conclusively as to which is the adopted bio-active conformation of the pyridine-substituted compound.

3.3.4. Analysis of electrostatic field maps. As described above, due to the constant +2 charge for all compounds in the library, electrostatic field maps were not found to be very useful in predicting biological activity. They contributed only 3.5% of the variance and the maps were not analyzed in detail due to this low value.

4. Conclusions

The design of new compounds for finding leads with better activity has been enhanced by the availability of the robust 3D QSAR maps generated in this project. We have been able to specifically derive chemical properties that are important to activity and hence adopt a rational approach toward the selection of substituents at various positions in our scaffold. Work is in progress to synthesize and test new compounds that implement the optimum features from the CoMSIA maps. Such molecules should show improved target interactions and biological activity.

Acknowledgments

This work was supported by awards from the Bill and Melinda Gates Foundation and from NIH (NIAID RO1A1 46365). Computers and software used in the research were funded in part by the Georgia Research Alliance. Patricia Ruiz was supported by a Colciencias scholarship, Colombia, South America.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2005.12.029.

References and notes

1. Tidwell, R. R.; Boykin, W. D.. In Demeunynck, M., Bailly, C., Eds.; Small Molecule DNA and RNA Binders: From Synthesis to Nucleic Acid Complexes; Wiley-VCH, Weinheim: Germany, 2003; Vol. 2, pp 414–460.
2. Fairlamb, A. H. *Trends Parasitol.* **2003**, *19*, 488.
3. Bouteille, B.; Oukem, O.; Bisser, S.; Dumas, M. *Fundam. Clin. Pharmacol.* **2003**, *17*, 171.
4. Singh, H.; Sivakumar, R. *J. Infect. Chemother.* **2004**, *10*, 307.
5. Jannin, J.; Cattand, P. *Curr. Opin. Infect. Dis.* **2004**, *17*, 565.
6. Zhou, L.; Lee, K.; Takker, D. R.; Boykin, D. W.; Tidwell, R. R.; Hall, J. E. *Pharm. Res.* **2002**, *19*, 1689.
7. Ansede, J. H.; Anbazhagan, M.; Brun, R.; Easterbrook, J. D.; Hall, J. E.; Boykin, D. W. *J. Med. Chem.* **2004**, *47*, 4335.
8. Wilson, D. W.; Nguyen, B.; Tanious, F.; Mathis, A.; Hall, J. E.; Stephans, C. E.; Boykin, D. W. *Curr. Med. Chem.* **2005**, 5, in press.
9. Bailly, C.; Chaires, J. B. *Bioconjugate Chem.* **1998**, *9*, 513.
10. Henderson, D.; Hurley, L. H. *Nat. Med.* **1995**, *1*, 522.
11. Baraldi, P. G.; Bovero, A.; Fruttarolo, F.; Preti, D.; Tabrizi, M. A.; Pavani, M. G.; Romagnoli, R. *Med. Res. Rev.* **2004**, *24*, 475.
12. Cory, M.; Tidwell, R. R.; Fairley, T. A. *J. Med. Chem.* **1992**, *35*, 431.
13. Wilson, D. W.; Tanious, F.; Ding, D.; Kumar, A.; Boykin, W. D.; Colson, P.; Houssier, C.; Bailley, C. *J. Am. Chem. Soc.* **1998**, *120*, 10310.
14. Matovu, E.; Stewart, M. L.; Geiser, F.; Brun, R.; Maser, P.; Wallace, L. J. M.; Burchmore, R. J.; Enyaru, J. C. K.; Barrett, M. P.; Kaminsky, R.; Seeback, T.; de Koning, H. P. *Eukaryotic Cell* **2003**, *2*, 1003.
15. de Koning, H. P.; Anderson, L. F.; Stewart, M.; Burchmore, R. J.; Wallace, L. J.; Barrett, M. *Antimicrob. Agents Chemother.* **2004**, *48*, 1515.
16. Shapiro, T. A.; Englund, P. T. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 950.
17. Shapiro, T. A.; Englund, P. T. *Annu. Rev. Microbiol.* **1995**, *49*, 117.
18. Yeates, C. *IDrugs* **2003**, *6*, 1086.
19. Laughton, C. A.; Tanious, F.; Nunn, C. M.; Boykin, W. D.; Wilson, D. W.; Neidle, S. *Biochemistry* **1996**, *35*, 5655.
20. Neidle, S. *Nat. Prod. Rep.* **2001**, *18*, 291.
21. González, M. P.; González, H. D.; Molina, R. R.; Cabrera, M. A.; Ramos de Armas, R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192.
22. Molina, E.; González, H. D.; González, M. P.; Rodríguez, E.; Uriarte, E. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 515.
23. Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
24. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.
25. Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A.. In *Computer Assisted Drug Design, ACS Symposia*; Olson, E. C., Christofferson, R. E., Eds.; American Chemical Society: Washington, DC, 1979; 122, pp 205–226.
26. Geladi, P.; Kowalski, B. R. *Anal. Chem. Acta* **1986**, *185*, 1.

27. Klebe, G.; Abraham, U. *J. Comput. Aided Mol. Des.* **1999**, *13*, 1.
28. Klebe, G. In *3D QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 173–199.
29. Folkers, G.; Merz, A.; Rognan, D. In *QSAR in Drug Design*; Kubinyi, H., Ed.; ESCOM: Leiden, 1993; pp 583–618.
30. Böhm, M.; Stürzebecher, J.; Klebe, G. *J. Med. Chem.* **1999**, *42*, 458.
31. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
32. Baltz, T.; Baltz, D.; Giroud, C.; Crockett, J. *EMBO J.* **1985**, *4*, 1273.
33. Räz, B.; Iten, M.; Grether-Bühler, Y.; Kaminsky, R.; Brun, R. *Acta Trop.* **1997**, *68*, 139.
34. Sybyl, Version 6. 9. 2; Tripos Associates: St. Louis, MO, 2003.
35. Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. *J. Comp. Chem.* **1989**, *10*, 982.
36. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. In *Numerical Recipes in C, The Art of Scientific Computing*; Cambridge University Press: New York, 1988; pp 324–326.
37. Martz, E. *Trends Biochem. Sci.* **2002**, *27*, 107, <http://proteinexplorer.org/>.
38. Bush, B. L.; Nachbar, R. B. *J. Comput. Aided Mol. Des.* **1993**, *7*, 587.
39. Nguyen, B.; Lee, M. P. H.; Hamelberg, D.; Bailey, C.; Colson, P.; Stanek, J.; Brun, R.; Neidle, S.; Wilson, D. W. *Biophys. J.* **2004**, *86*, 1028.