Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 15 (2007) 1117–1126
Bioorganic & Medicinal Chemistry

Molecular modeling of some 1*H*-benzimidazole derivatives with biological activity against *Entamoeba histolytica*: A comparative molecular field analysis study☆

Fabian López-Vallejo,[a] José Luis Medina-Franco,[b] Alicia Hernández-Campos,[a] Sergio Rodríguez-Morales,[a] Lilián Yépez,[c] Roberto Cedillo[c] and Rafael Castillo[a,*]

[a]*Departamento de Farmacia, Facultad de Química, UNAM, CU, D.F. 04510, Mexico*
[b]*BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA*
[c]*Unidad de Investigación Médica en Enfermedades Infecciosas y Parasitarias, IMSS. México, DF 06720, Mexico*

Received 15 July 2006; revised 23 September 2006; accepted 11 October 2006
Available online 18 October 2006

Abstract—Comparative molecular field analysis (CoMFA) was performed on a set of 1*H*-benzimidazole derivatives. Molecular modeling and 3D-QSAR were employed to determine the tautomeric form that would probably fit a target receptor in *Entamoeba histolytica*. CoMFA results suggest that the antiamoebic activity is favored with steric bulk at position 5 of the benzimidazole ring and low electron density on the group at position 2. To the best of our knowledge this is the first 3D-QSAR study performed for benzimidazoles as antiamoebic agents. The CoMFA models derived will be very valuable to design new and more potent compounds against *E. histolytica*.

1. Introduction

Parasitic infections such as amoebiosis and other protozooses are still major threats against public health, especially in developing countries. The infections caused by *Entamoeba histolytica* mainly affect the infant population with high morbidity and mortality indices due to severe diarrhea and dehydration.[1] Although drug therapy exists for the control of these infections, such as the nitroheterocycles, they produce severe side effects and in some cases resistance by the parasites have also been reported.[2] Consequently, the search for new therapeutic agents is of utmost importance.

We have recently reported that some 1*H*-benzimidazole derivatives with hydrogen at position 1, non 2-carbamates, were very active in vitro against *E. histolytica*.[3–6] Since there are no reports on the target receptor for these types of compounds and in order to obtain more information regarding the structural requirements of benzimidazole derivatives to be active against *E. histolytica*, computational studies were carried out.

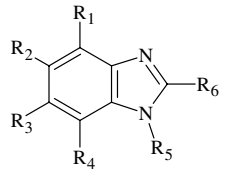
Three-dimensional quantitative structure–activity relationship (3D-QSAR) methods, such as comparative molecular field analysis (CoMFA),[7] have been applied in many cases for the design of new active molecules; especially, when the binding site and mechanism of action are known.[8] Additionally, it is also useful when the binding site is unknown.[9] Although CoMFA studies have been conducted for benzimidazole derivatives,[10–14] there are no computational analyses that study the activity against *E. histolytica.*

In order to rationalize the structure–activity relationships of a set of 41 1*H*-benzimidazole derivatives (cf., Table 1), in this paper 3D-QSAR studies using CoMFA were carried out. A particular challenge for the present CoMFA study is to deal with the tautomeric forms of compounds **3**, **8**, **13**, **29**, **32**, **36**, and **40**; and the tautomeric forms of compounds **6**, **7**, **9**, **10**, **17**, **21**, and **23** shown in Figures 1 and 2, respectively. The challenge is to determine which tautomer is most likely to be the active one when the QSAR studies are carried out. Each

Keywords: Amoebiosis; CoMFA; QSAR validation; Tautomers.
☆ Taken in part from the Ph.D. dissertation of Fabian López-Vallejo.
* Corresponding author. Tel./fax: +52 5622 5329; e-mail: rafaelc@servidor.unam.mx

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2006.10.019

Table 1. Benzimidazole derivatives used in the CoMFA studies



Compound	R_1	R_2	R_3	R_4	R_5	R_6	pIC_{50}
1	H	H	H	H	H	CH_3	8.15
2	H	H	H	H	H	NH_2	6.94
3	H	H	H	H	H	SH	6.88
4	H	H	H	H	H	SCH_3	6.41
5	H	H	H	H	H	H	7.38
6	H	Cl	H	H	H	CH_3	7.08
7	H	Cl	H	H	H	NH_2	6.90
8	H	Cl	H	H	H	SH	8.30
9	H	Cl	H	H	H	SCH_3	6.72
10	H	Cl	H	H	H	H	7.41
11	H	Cl	Cl	H	H	CH_3	7.60
12	H	Cl	Cl	H	H	NH_2	7.23
13	H	Cl	Cl	H	H	SH	7.26
14	H	Cl	Cl	H	H	SCH_3	6.45
15	H	Cl	Cl	H	H	H	7.02
16	H	H	H	H	H	CF_3	7.16
17	H	Cl	H	H	H	CF_3	7.66
18	H	H	H	H	CH_3	CF_3	7.40
19	H	Cl	H	H	CH_3	CF_3	7.34
20	H	Cl	Cl	H	CH_3	CF_3	7.48
21	H	Br	H	H	H	CF_3	5.88
22	H	Br	Br	H	H	CF_3	6.37
23	Br	H	Br	H	H	CF_3	7.30
24	Br	Br	Br	Br	H	CF_3	6.52
25	H	Cl	Cl	H	H	C_2F_5	6.68
26	H	NO_2	NO_2	H	H	CF_3	5.87
27	Br	Br	Br	Br	H	C_2F_5	6.67
28	H	H	H	H	CH_3	CH_3	7.21
29	H	H	H	H	CH_3	SH	8.22
30	H	H	H	H	CH_3	SCH_3	6.71
31	H	Cl	H	H	CH_3	CH_3	8.30
32	H	Cl	H	H	CH_3	SH	8.00
33	H	Cl	H	H	CH_3	SCH_3	7.43
34	H	H	Cl	H	CH_3	CH_3	6.69
35	H	H	Cl	H	CH_3	NH_2	6.61
36	H	H	Cl	H	CH_3	SH	5.42
37	H	H	Cl	H	CH_3	SCH_3	6.24
38	H	Cl	Cl	H	CH_3	CH_3	7.89
39	H	Cl	Cl	H	CH_3	NH_2	7.43
40	H	Cl	Cl	H	CH_3	SH	6.89
41	H	Cl	Cl	H	CH_3	SCH_3	5.98

tautomeric form represents a structure with different properties, only one of which may have greater affinity for the target, and hence, greater biological activity than the other. Notably, this tautomeric form is not necessarily the lowest in energy. Tautomers are often disregarded in computer-aided drug design (CADD) applications. Since tautomeric forms differ in functional groups, shapes, hydrogen-bonding pattern, surfaces, and other properties, the corresponding descriptors used in QSAR will have different values for each tautomeric form. Considering that tautomerism in CADD has a huge impact on the development of QSAR or affinity receptor–ligand models,[15] this paper deals with a density functional theory (DFT) and semi-empirical (PM3) calculations to help in the task of finding the putative bioactive tautomer of several 1*H*-benzimidazoles under study.

2. Results and discussion

2.1. Tautomeric form selection

Several compounds in the data set may exist as two or even three tautomeric forms. In Figure 1, the tautomeric equilibria of benzimidazole-2-thione derivatives are proposed. In Figure 2, all tautomeric equilibria of 1*H*-benzimidazole derivatives with hydrogen at position 1 are depicted. For CoMFA studies, tautomeric forms represent a challenge since each tautomer has different values for its electrostatic and steric descriptors. Regarding the molecules in this paper, the rearrangement of one hydrogen atom of the same compound leads to an alteration of the benzimidazole ring, leading to a new molecular skeleton (Figs. 1 and 2). Different molecular skeletons may have different steric and electrostatic complementarities to the target. Thus, incorrectly selected tautomers for the CoMFA study will produce misleading correlations.

The determination of the most suitable tautomeric form for 3D-QSAR studies is not trivial. The tautomeric equilibrium can be influenced by different factors.[15] In the case of a ligand–receptor interaction this equilibrium must be shifted toward the tautomeric form that stabilizes the ligand–receptor complex, but this tautomeric form is not necessarily the lowest in energy.

In order to explore the putative bioactive tautomeric form of compounds in Figures 1 and 2, and to develop a predictable CoMFA model for benzimidazole derivatives in Table 1, two approaches were followed. In the first one, the tautomer of lowest energy was considered the bioactive tautomeric form. In the other one, the bioactive tautomeric form was selected by a CoMFA model with predictive ability, which in turn was developed using benzimidazole derivatives with no tautomeric forms. In both approaches, high accurate evaluation of the atomic partial charges was carried out by means of quantum chemical calculation at DFT/B3LYP level theory using natural charges.

2.1.1. CoMFA study based on lowest energy tautomers. In this approach, a theoretical study was carried out to determine the lowest energy tautomers in the gas phase at DFT level. The change in free energy of the tautomeric equilibrium and the tautomeric constant were calculated according to Eq. 1. This approximation has been demonstrated to be acceptable for tautomeric equilibria.[16–23] The results of these calculations are shown in Table 2.

$$\Delta G = -RT \cdot \ln K_T \Rightarrow K_T = e^{-(\Delta G/RT)} \qquad (1)$$

As can be seen in Table 2, thione forms **3b**, **8b**, **13b**, **29b**, **32b**, **36b**, and **40b** are more stable than their corresponding tautomers. The lowest $-pK_T$ is 98.0 for compound

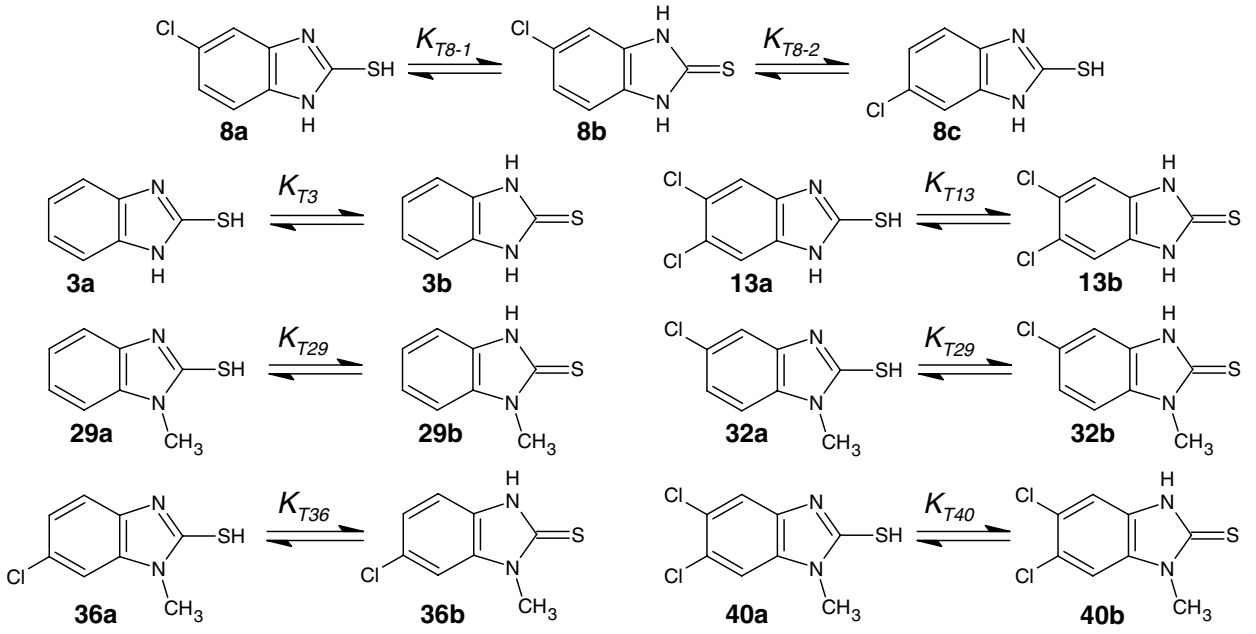


Figure 1. Tautomeric equilibria for benzimidazole-2-thione derivatives.

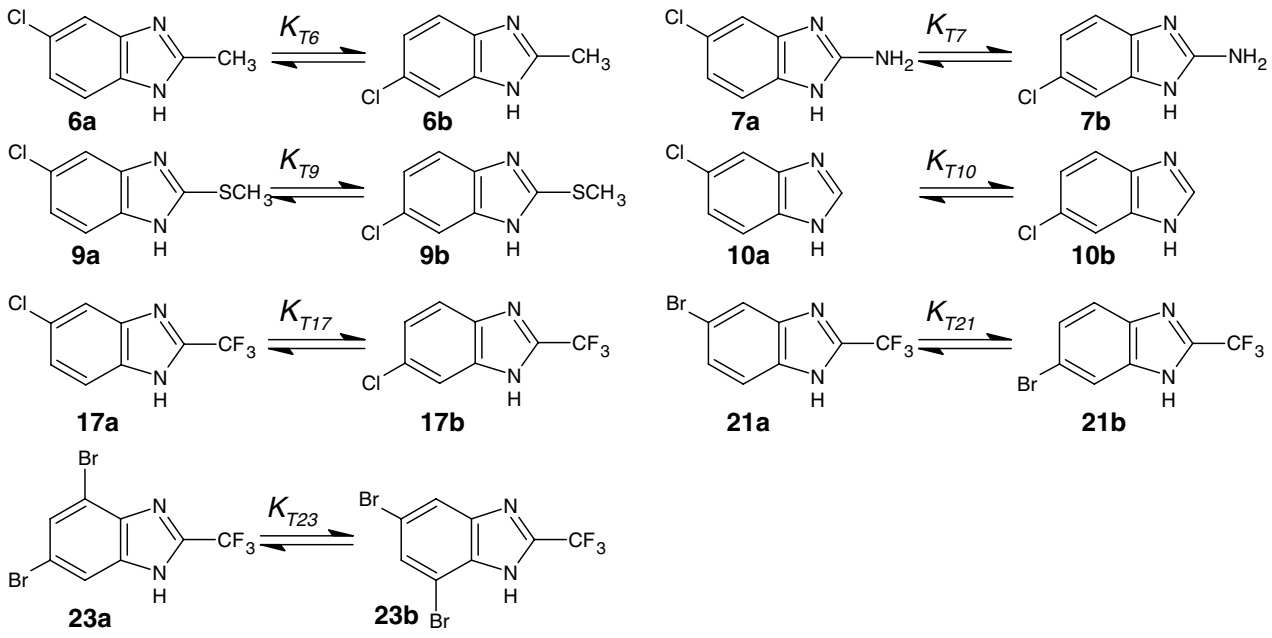


Figure 2. Tautomeric equilibria for 1-*H* benzimidazole derivatives with hydrogen at position 1.

13. These results are in agreement with the data of the ^{1}H NMR spectra for these compounds, in which essentially 100% of the thione form is predominant.[3] For 1*H*-benzimidazole derivatives with hydrogen at position 1, the equilibrium is slightly favored toward the tautomeric form **6b**, **7b**, **9b**, **10b**, **17b**, **21b**, and **23b**. In all these compounds ΔG is less than 1.45 kcal/mol (the largest $-\mathrm{p}K_\mathrm{T}$ is 10.6 for compound **23**).

Once the lowest energy tautomers were determined, the CoMFA model was developed based on all 41 compounds in data set (Table 1). The tautomeric form with the calculated lowest energy '**b**' for compounds in Table 2 was used to develop this model plus the remaining compounds listed in Table 1 that have no tautomeric form. Results are summarized in Table 3. The poor statistics of this model ($q^2 = 0.292$; $r^2 = 0.634$) suggest that some tautomeric forms considered in this first approach could not have steric and electrostatic complementarity with the target. This observation further confirms the fact that the lowest energy tautomer is not necessarily the bioactive one.

2.1.2. CoMFA study based on best-predicted tautomers. In this approach, a CoMFA model was developed using as a training set 27 compounds in Table 1 that have no

Table 2. DFT (B3LYP) predicted tautomeric equilibrium constants in the gas phase ($\varepsilon = 1$) at 298.15 °K for 1*H*-benzimidazole derivatives

Tautomers	DFT energy (a.u)	$-\Delta G$ (kcal/mol)	$-pK_T$[a]
3a	−778.053844	15.00	109.928
3b	−778.076873		
6a	−878.785931	0.09	0.664
6b	−878.786070		
7a	−894.822511	1.28	9.375
7b	−894.824475		
8a	−1237.649110	14.30	104.778
8b	−1237.671060		
8c	−1237.649280	14.18	103.966
9a	−1276.967820	0.01	0.095
9b	−1276.967840		
10a	−839.461439	0.13	0.931
10b	−839.461634		
13a	−1697.238750	13.38	98.047
13b	−1697.259290		
17a	−1176.494250	0.21	1.575
17b	−1176.494580		
21a	−3290.179390	0.14	1.002
21b	−3290.179600		
23a	−5863.456620	1.45	10.597
23b	−5863.458840		
29a	−817.366243	15.62	114.497
29b	−817.390229		
32a	−1276.961710	14.77	108.262
32b	−1276.984390		
36a	−1276.961880	14.66	107.451
36b	−1276.984390		
40a	−1736.551500	14.18	103.966
40b	−1736.573280		

[a] $pK_T = -\log K_T$

tautomeric forms. The PLS analysis gave a model with q^2 and r^2 values of 0.455 and 0.794, respectively (Table 3, model 2). Exclusion of one compound (an outlier) from model 2 increased the q^2 and r^2 values (Table 3, model 3). The omission of the outlier was done mainly because it was not well predicted by model 2. Although CoMFA model 3 has a higher value of q^2, it is still relatively low (i.e., 0.490), however, it has acceptable values of r^2, s and N (Table 3).[24,25]

CoMFA model 3 was used to predict the biological activities of the 28 tautomeric forms in Table 2. Results are summarized in Table 4, where the tautomeric forms with predicted pIC_{50} values that are closer to the experimental ones are marked with an asterisk. Tautomers with the lowest residual values might correspond to the bioactive form. Notably, seven of them, **3b**, **6b**, **7b**, **21b**, **23b**, **36b**, and **40b**, correspond to the calculated lowest energy tautomers (Table 2). In the case of the 2-thione tautomeric forms (**8b**, **13b**, **29b**, and **32b**), which are energetically more stable, the CoMFA model 3 predicts that the thiol form is the bioactive one. The change in the ΔG for these tautomeric forms is about 14 kcal/mol (Table 2), a relatively high energetic cost. However, **8a**, **13a**, **29a**, and **32a** have higher pIC_{50} than the thione forms **3b**, **36b** and **40b** selected by CoMFA model 3. In the case of tautomeric forms **6a**, **7a**, **9a**, **10a**, **17a**, **21a**, and **23a** for 1-*H* benzimidazole derivatives with hydrogen at position 1, the change in the ΔG is not higher than 1.45 kcal/mol.

Once the bioactive tautomeric forms, marked with an asterisk in Table 4, were predicted by model 3, a fourth CoMFA model was then developed with all 41 molecules and the results were compared with those obtained from model 1. For best correlation, the outlier was included in both models. Results are summarized in Table 3 (model 4). It is clear from this table that model 4 is significantly better than model 1. The good q^2 and r^2 obtained in model 4 further suggest that selected tautomers could be the bioactive form.

The biological activities of the selected tautomeric forms, especially for tautomers **3b**, **8a**, **13a**, **29a**, **32a**, **36b**, and **40b**, showed an interesting correlation with the net natural charge on the group at position 2 (Fig. 3).

Interestingly, compounds with a positive net charge (**8a**, **13a**, **29a**, and **32a**) have a high biological activity (e.g., pIC_{50} greater than 7.26), and compounds with a negative net charge (**3b**, **36b**, and **40b**) have a low biological activity (e.g., pIC_{50} less than 6.89).

Table 3. Summary of CoMFA models related to the tautomeric forms selection

Statistics	CoMFA study based on lowest energy tautomers	CoMFA study based on best-predicted tautomers		
	Model 1	Model 2	Model 3	Model 4
q^{2a}	0.292	0.455	0.490	0.707
r^{2b}	0.634	0.794	0.815	0.868
Steric contribution	0.604	0.654	0.660	0.620
Electrostatic contribution	0.396	0.346	0.340	0.380
s^c	0.450	0.320	0.310	0.273
F^d	21.389	29.613	32.325	59.339
N^e	3	3	3	4
n^f	41	27	26	41

[a] Cross-validated r^2.
[b] Non-cross-validated r^2.
[c] Standard error estimate.
[d] Fraction of explained versus unexplained variance.
[e] Optimum number of principal components.
[f] Number of compounds in the training set.

Table 4. Experimental and predicted pIC_{50} for all tautomeric forms using CoMFA model 3

Tautomer	pIC_{50} Experimental	pIC_{50} Predicted	Residual
3a	6.876	7.361	−0.485
3b*		6.786	0.090
6a	7.076	8.053	−0.977
6b*		6.978	0.098
7a	6.903	8.014	−1.111
7b*		7.046	−0.143
8a*	8.301	7.841	0.460
8b		6.771	1.530
8c		7.267	1.034
9a*	6.717	6.888	−0.171
9b		5.813	0.904
10a*	7.409	7.831	−0.422
10b		6.763	0.646
13a*	7.260	7.294	−0.034
13b		6.744	0.516
17a*	7.658	7.238	0.420
17b		6.144	1.514
21a	5.876	7.238	−1.362
21b*		6.116	−0.240
23a	7.301	6.278	1.023
23b*		7.097	0.204
29a*	8.222	7.500	0.722
29b		6.928	1.294
32a*	8.000	7.952	0.048
32b		7.411	0.589
36a	5.423	6.878	−1.455
36b*		5.898	−0.475
40a	6.893	7.440	−0.547
40b*		6.889	0.004

In order to further explore the electronic distribution of selected tautomeric forms based on model 3, the electrostatic potential density surface at DFT level was calculated to give an isovalue of 0.002. Figure 4 shows the electrostatic potential density surface for all tautomeric forms of the benzimidazole-2-thiol derivatives. As expected, all thione forms have high electron density at sulfur at position 2, and all thiol forms have low electron density in this group at this position. In this figure it can also be seen that tautomeric forms predicted by the CoMFA model 3 have a different density charge on the group at position 2. In each tautomeric pair there is a tautomer with a high residual value (Table 4). Notably, the best-predicted tautomers, having an activity greater than 7.26 (i.e., **8a**, **13a**, **29a**, and **32a**), have a low electron density on sulfur at position 2 (e.g., a positive net charge. See also Fig. 3). In turn, the best-predicted tautomers that have an activity lower than 7 (i.e., **3b**, **36b**, and **40b**) have a high electron density on sulfur at position 2 (e.g., a negative net charge. See also Fig. 3). This observation suggests that activity is favored with low electron density in this region.

2.2. Development of robust and thoroughly validated CoMFA models

Once we determined the bioactive tautomeric forms described before, we developed six new CoMFA models. For this, five compounds with a different substitution pattern in Table 1 were selected randomly as a test set, as recommended.[24] The new CoMFA models were developed based on a training set of 36 compounds, 23 from Table 1 that have no tautomeric forms and the 13 bioactive tautomers determined by CoMFA model 3. Partial atomic charges determined by DFT and PM3 calculations were used. Results are summarized in Table 5.

From the data in Table 5 it can be seen that CoMFA models 5–10 have a q^2 higher than 0.6, except CoMFA model 8 developed from Mülliken charges. In all models r^2 is higher than 8.0. The best statistical results were for model 7, obtained from DFT calculations and natural charges; it has a q^2 value of 0.72 and a r^2 value of 0.873. Considering semi-empirical calculations, the best CoMFA model developed is model 9, since it uses only 4 principal components, while model 10 uses 5.

Validation is a crucial aspect of any QSAR study. The leave-one-out R^2 (LOO q^2) cross-validation approach is commonly used as a criterion for both robustness and predictive ability; however, according to Golbraikh and Tropsha, this approach per se is not sufficient to assure the predictability of any QSAR model. Low values of q^2 can indeed serve as indicators of low predictive ability of the model, but as a matter of fact, values of high q^2 do not necessarily imply high predictive ability

Tautomeric form	*Net charge on position 2	pIC50
3b	-0.224	6.880
8a	0.213	8.300
13a	0.221	7.260
29a	0.199	8.220
32a	0.208	8.000
36b	-0.374	5.420
40b	-0.204	6.890

* Natural charges from DFT calculation

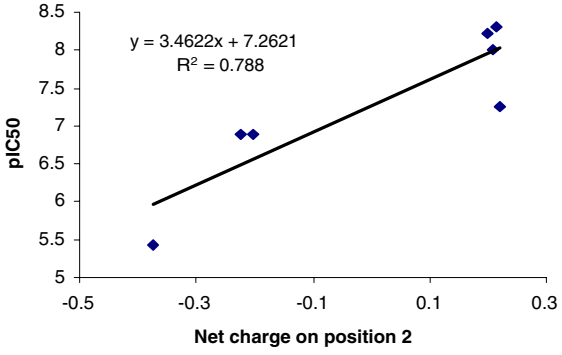


Figure 3. pIC_{50} versus net charge on position 2 for tautomeric form selected from CoMFA model 3 (Table 3).

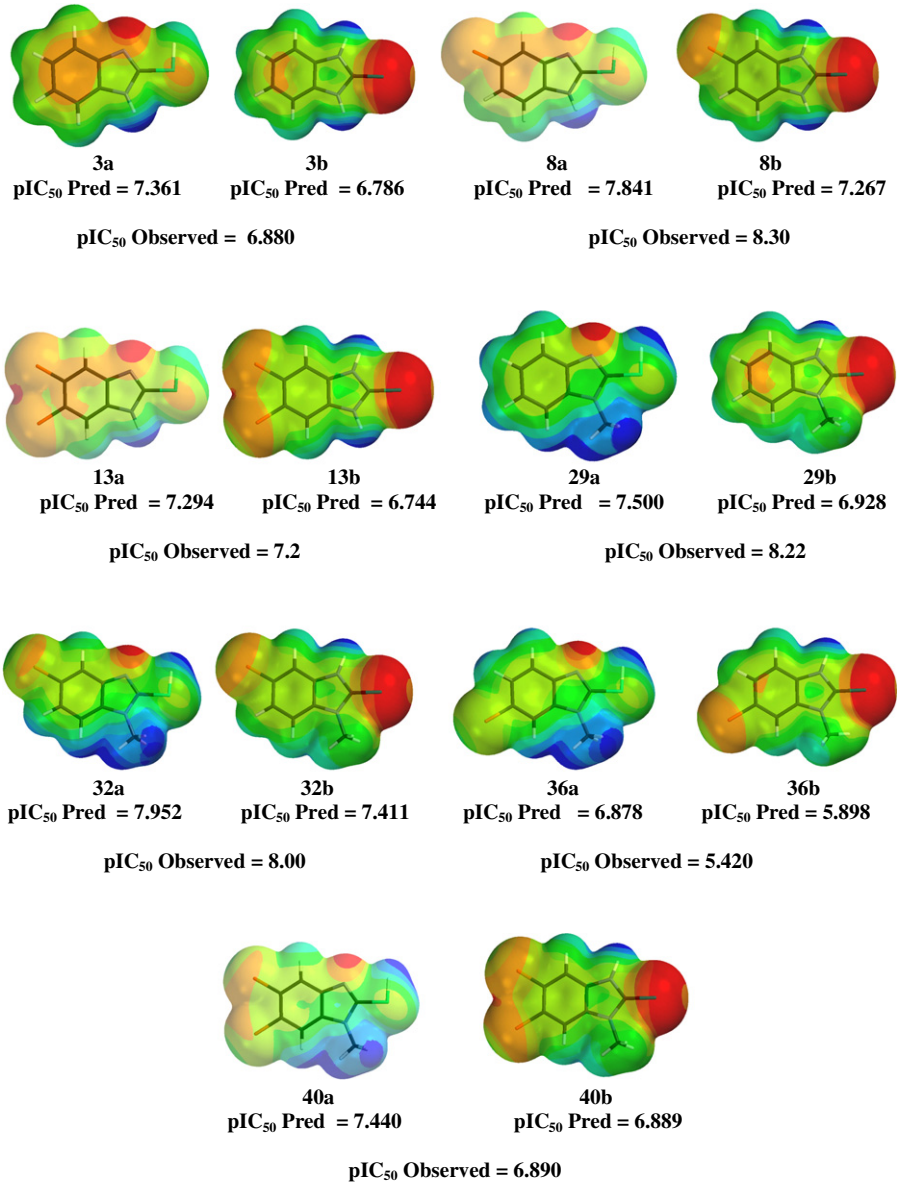


Figure 4. Potential electrostatic density surface for tautomeric forms of the benzimidazole-2-thiol derivatives. Regions in red and orange indicate that the sulfur atom at position 2 has a high electron density, while regions in light green indicate that the sulfur atom on the same position has a low electron density. Regions in yellow and green have intermediate electron density, all of them obtained from DFT (B3LYP/6-31G*) calculation. Values of predicted pIC_{50} were obtained from CoMFA model 3.

of the QSAR model, since in addition, it must meet the criteria discussed in Section 4.5.

Therefore, in addition to the q^2 computed for models 5–10, all models in Table 5 were further validated by predicting the activities of compounds in the test set. Results of this validation are also reported in Table 5.

Table 6 shows the predicted activities and residual values made by the CoMFA model 7 for the test set. Figure 5 shows the validation graphics made by the CoMFA model 7.

All models in Table 5, except models 8 and 10, for which $(R^2 - R_0^2)/R^2$ and $(R'^2 - R_0'^2)/R'^2 > 0.1$ meet the minimal validation criteria presented in Section 4.5. In this table, it can be seen that the CoMFA model 7, although it has the highest q^2 value (0.720), its R^2 value (0.9517) is lower than the respective R^2 value for models 9 and 10 (0.9632 and 0.9820, respectively). The lack of correlation between q^2 with R^2 has been previously noted.[24] Although model 10 has the highest r^2 and R^2 values (0.8910 and 0.9820, respectively) it does not satisfy all the validation criteria since $(R^2 - R_0^2)/R^2$ and $(R'^2 - R_0'^2)/R'^2 > 0.1$. These observations further support the convenience of assessing all criteria presented in Section 4.5 to explore the predictive ability of CoMFA models rigorously.

The best CoMFA model from DFT calculations was developed using natural charges (model 7), and the best CoMFA model from PM3 calculations was developed

Table 5. Different CoMFA models obtained from quantum chemistry levels for the electrostatic calculations

	DFT calculations B3LYP/6-31G*//HF/6-31G* Charges			Semi-empiric calculations PM3//MMFF Charges		
	Mülliken Model 5	Electrostatic potential Model 6	Natural Model 7	Mülliken Model 8	MOPAC Model 9	Gasteiger-Hückel Model 10
q^2	0.686	0.633	0.720	0.595	0.633	0.666
r^2	0.855	0.849	0.873	0.846	0.849	0.891
n	36	36	36	36	36	36
N	3	4	4	4	4	5
F	63.544	43.58	53.080	42.487	43.696	48.824
s	0.283	0.295	0.212	0.298	0.294	0.255
Steric contribution	0.580	0.567	0.662	0.679	0.708	0.600
Electrostatic contribution	0.420	0.433	0.338	0.321	0.292	0.400
R^{2g}	0.954	0.943	0.951	0.922	0.963	0.982
R_0^{2h}	0.860	0.849	0.877	0.795	0.934	0.838
k^i	0.961	0.951	0.963	0.940	0.914	0.917
$(R^2 - R_0^2)/R^2$	0.098	0.099	0.077	0.127	0.029	0.146
$R_0'^{2j}$	0.692	0.662	0.754	0.443	0.901	0.571
k'^k	1.037	1.048	1.035	1.058	1.092	1.085
$(R'^2 - R_0'^2)/R'^2$	0.274	0.298	0.277	0.519	0.064	0.417

q^2, r^2, n, N, F, and s have the same meaning as in Table 3. [g]Correlation coefficient R for the regression of ÿ versus y and y versus ÿ. [h,j]Correlation coefficient R for the regression through the origin. [i,k]The slopes for the regression through the origin.

using MOPAC charges (model 9). Both models meet the validation criteria. Models 7 and 9 can, therefore, be used to predict the biological activity of compounds not included in the training set. One of the advantages of the model obtained from PM3 calculations is that it is less expensive computationally. Hence, model 9 should be a good tool for the design of novel and more active molecules with antiamoebic activity. On the other hand, from the data in Table 5 it can also be seen that the CoMFA models developed from high levels of theory DFT calculations do not always lead to noticeable improvement of the CoMFA models as shown in the paper recently reported by Occhiato.[26]

Table 6. Predictions of CoMFA model 7 for the test set

Compound	Experimental activity	Predicted activity	Residual
27	6.67	6.09	0.58
32a	8.00	8.29	−0.29
33	7.43	7.14	0.29
37	6.24	5.72	0.51
38	7.89	7.53	0.35

2.3. CoMFA contour maps

The stdev* coefficient contour maps for the best CoMFA models developed from DFT and PM3 calculations are depicted in Figure 6. As an aid in visualization, compound **31** is displayed on the maps, which represent those areas where changes in the field values of the compounds in the training set have a strong influence over the biological activity. Figure 6 shows areas where high electron density is needed in the ligand to increase (red) or decrease (blue) the biological activity and the areas where an increased steric volume (green) in the ligand is needed to favor biological activity. In the case of the benzimidazole derivatives under study, it is seen that the electrostatic contour maps, using either DFT or PM3 calculations, show a large blue contour map close to position 2 of the benzimidazole ring. This contour indicates that a group with high electron density at this position decreases the biological activity. This fact explains why the 2-methylbenzimidazole derivative **1** (8.15) is more active than the predicted thione form **3b** (6.88) or why the predicted thiol form **8a** (8.30) is more active than the 2-methylbenzimidazole derivative **6** (7.08). High electron density on groups at position 2

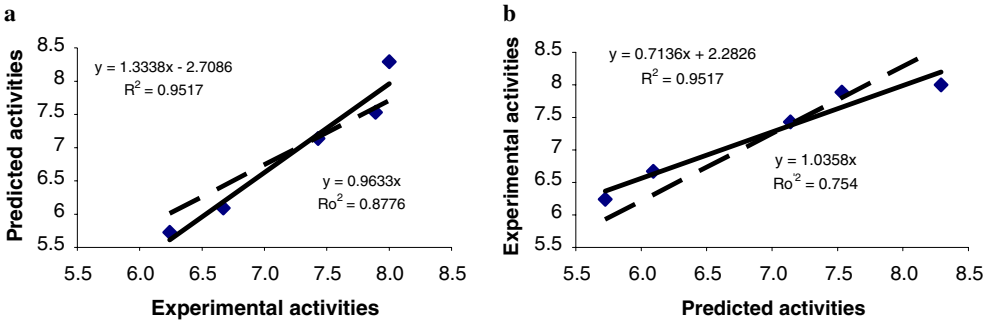


Figure 5. Validation of CoMFA model 7 graphics. (a) Predicted versus experimental activities and (b) Experimental versus predicted activities to the best CoMFA model obtained from the approach 2.

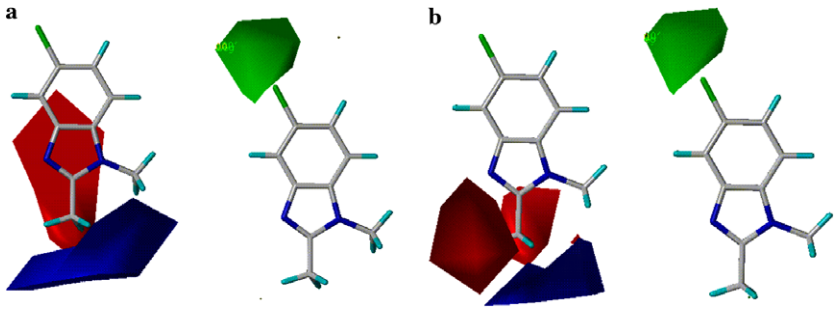


Figure 6. Electrostatic and steric CoMFA STDEV*COEFF contour maps for compound 31 obtained from (a) DFT (model 7) and (b) PM3 (model 9) calculations. For an increase of the biological activity, the positive charge, the negative charge, and the molecular volume have to be increased in blue, red, and green regions, respectively.

of the benzimidazole ring can also be seen with electron-withdrawing groups at this position, such as trifluoromethyl; therefore, one would expect this compound **16** (7.16) to be less active than compound **1** (8.15) with a hydrogen at this position.

The green contour obtained by semi-empiric and DFT calculations (Fig. 6) indicates that an increase in size surrounding the substituent at position 5 of the benzimidazole ring favors the biological activity. This information can explain why the benzimidazole derivatives with a chlorine at position 5 are more active than their corresponding analogues with hydrogen at that position; thus, compounds **9a** (6.72), **10a** (7.41), **31** (8.30), and **33** (7.43) with a chlorine at position 5 are more active than **4** (6.41), **5** (7.38), **28** (7.21), and **30** (6.71), respectively, with a hydrogen at that position. In the case of compounds **6** and **7**, the predicted bioactive tautomeric forms are **6b** (7.08) and **7b** (6.90), which have the chlorine substituent at position 6 instead of position 5 and, therefore, are less active than **11** (7.60) and **12** (7.23), respectively, with a chlorine substituent at position 5.

3. Summary and conclusions

A CoMFA study was carried out on 41 1*H*-benzimidazole derivatives, most of which were synthesized in our laboratory and all of which had been tested by our research group in previous studies and found to be active in vitro against *E. histolytica*. As far as we know, this is the first computational study aimed at exploring the structural requirements for the activity of benzimidazole derivatives as antiamoebic agents. A particular challenge of the present CoMFA study was to determine the putative bioactive tautomeric form of several of these compounds in the data set. Theoretical calculations using DFT were employed to explore the most stable tautomers in gas phase; however, it was found that not all of the most stable tautomers led to predictive CoMFA models. A second approach was then undertaken consisting of the determination of the preferred tautomeric forms by QSAR-based predictions. As a result a CoMFA guided approach to select putative tautomeric forms with biological significance is presented.

Using best-predicted tautomers, CoMFA models were developed and validated; different charge models were determined from DFT and PM3 calculations. The validation studies included not only the most-common criterion q^2 but also several statistical criteria recently proposed by Tropsha and Golbraikh, based on predictions for an external set. The best CoMFA model developed from DFT calculations has a q^2 value of 0.72, and a R^2 value for an external set of 0.95. The best CoMFA model developed from PM3 calculations has a q^2 value of 0.633, and a R^2 value for an external set of 0.96. Such CoMFA models help to rationalize, at the molecular level, the structure–activity relationships of some 1*H*-benzimidazole derivatives against *E. histolytica*. Results suggest that antiamoebic activity is favored with steric bulk at position 5 of the benzimidazole ring and low electron density of the group at position 2. CoMFA models 7 and 9 are very valuable and could be useful in the design of new and more potent molecules against *E. histolytica*.

4. Materials and methods

4.1. Data set and biological data

The data set consists of 41 1*H*-benzimidazole derivatives (**1**–**41**) shown in Table 1. All biological activities of the 41 compounds used in this paper were determined in the same laboratory, using the same protocol. The antiamoebic activity of compounds **1**–**20** was recently reported by our research group,[3,4] and compounds **21**–**27** were reported by Andrzejewska et al.[27] As for compounds **28**–**41**, they were also synthesized and tested in vitro against *E. histolytica* by our research group, using the same procedure employed to prepare and test **1**–**20**. All biological activities were converted into the corresponding $-\log IC_{50}$ values ($-\log IC_{50} = pIC_{50}$).

4.2. Molecular modeling

All molecular modeling calculations and visualizations were carried out on a Silicon Graphics Octane2 workstation running under the IRIX 6.5 operating system, and a DELL Pentium IV, 3.0 MHz Pc running under LINUX Red Hat 7.0 operating system. All molecules were constructed using the sketch module implemented in SYBYL 6.8 (Tripos) molecular modeling software.[28] Semi-empirical calculations were carried out as follows. All the molecules were sketched, and geometric optimi-

zations were performed for each one using the MM TRIPOS force field with a 0.005 kcal/mol energy gradient convergence criterion. After that, a second geometric optimization for each molecule was performed using the semi-empiric PM3 implemented in MOPAC SYBYL 6.8 module. Then, Mopac, Mülliken and Gasteiger-Hückel atomic partial charges were calculated for each molecule.[29–31] DFT calculations were computed as follows. The coordinates of the molecules from semi-empirical calculations were then imported into SPARTAN'02 modeling software.[32] Geometric optimization was performed again using Hartree-Fock level with a 6-31G* basis set. Finally, a single point was employed to calculate the energy applying density functional theory (DFT), using the B3LYP hybrid method with a 6-31G* basis set. Spartan'02 allows the assignment of atomic partial charges under three different schemes. Natural, Mülliken, and electrostatic potential charges were calculated for all molecules.[33,34]

4.3. CoMFA methodology and structure alignment

The CoMFA studies were carried out with SYBYL 6.8 software. All molecules were aligned using the fit option over all carbon atoms of compound **5**, which was used as a template. It was assumed that hydrogen and methyl at N-1 in all compounds occupy the same position upon binding to the receptor. Electrostatic and steric CoMFA fields were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å in all three dimensions within a defined region. The grid was extended to 4.0 Å in every direction away from the molecule. A sp^3 carbon as a steric probe atom and a +1 charge as electrostatic probe were used. The maximum field values were truncated at 30 kcal/mol for the steric fields and to ±30 kcal/mol for the electrostatic fields.

4.4. Partial least squares (PLS) analysis

The partial least squares analysis was used to generate a linear regression equation that correlates changes in the electrostatic and steric fields with changes in biological activities (pIC_{50}). This analysis was also used in conjunction with cross-validation to obtain the optimal number of principal components. The cross-validation analysis was performed using the Leave-One-Out (LOO) method.

4.5. CoMFA models validation

For the evaluation of the predictive ability of the final CoMFA models, a set of criteria defined by Golbraikh and Tropsha was used.[24,25] By this approach the experimental activity (y) is plotted against the predicted activity ($\ddot{y}$). For an ideal QSAR model the regression line presented in Eq. 2 bisects the angle formed by positive directions of the orthogonal axes y and $\ddot{y}$; the slope is equal to 1, the intercept is equal to 0, and the correlation coefficient R for the regression of $\ddot{y}$ versus y is equal to 1.

$$y = m\ddot{y} + b, \tag{2}$$

where m is the slope and b is the intercept on axis y.

When a real QSAR model is close to an ideal model, the former model is able to predict with high precision the biological activity values for a test set. In this case R between $\ddot{y}$ versus y and y versus $\ddot{y}$ must be close to 1 and the regressions of $\ddot{y}$ versus y and y versus $\ddot{y}$ through the origin ($y^\circ = k\ddot{y}$ and $\ddot{y}^\circ = k'y$) should have at least either k or k' close to 1. The slopes k and k' are characterized by the following expressions:

$$k = \frac{\sum y\ddot{y}}{\sum \ddot{y}^2} \tag{3}$$

$$k_{,} = \frac{\sum y\ddot{y}}{\sum y^2} \tag{4}$$

where summations are overall compounds in the test set. In addition, the corresponding determination coefficients R_0^2 and $R_0'^2$ must be close to R^2.

In summary, the final CoMFA models are considered acceptable if they meet all of the following criteria:[24]

- $q^2 > 0.5$
- $R^2 > 0.6$
- $[(R^2 - R_0^2)/R^2] < 0.1$ or $[(R'^2 - R_0'^2)/R'^2] < 0.1$
- $0.85 \leqslant k \leqslant 1.15$ or $0.85 \leqslant k' \leqslant 1.15$

Acknowledgments

Fabian López-Vallejo is grateful to DGAPA, UNAM, for the Ph.D. scholarship. We are very grateful for the financial support by CONACyT of Projects Nos. G-34851-M and V43629M, and DGAPA No. IN202101 and IX244704.

References and notes

1. Mora, L.; García, A.; DeDonato, M. *Kasmera* **2005**, *33*, 36.
2. (a) Byington, C. L.; Dunbrack, R. L.; Whitby, F. G.; Cohen, F. E.; Agabian, N. *Exp. Parasitol.* **1997**, *87*, 194; (b) Johnson, P. J. *Parasitol. Today* **1993**, *9*, 183; (c) Upcroft, J. A.; Upcroft, P. *Parasitol. Today* **1993**, *9*, 187; (d) Voolmann, T.; Boreham, P. F. L. *Med. J. Aust.* **1993**, *159*, 490.
3. Valdez, J.; Cedillo, R.; Hernández-Campos, A.; Yépez, L.; Hernández-Luis, F.; Navarrete-Vázquez, G.; Hernández, M.; Castillo, R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2221.
4. Navarrete-Vázquez, G.; Cedillo, R.; Hernández-Campos, A.; Yépez, L.; Hernández-Luis, F.; Valdez, J.; Morales, R.; Hernández, M.; Castillo, R. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 187.
5. Navarrete-Vázquez, G.; Yépez, L.; Hernández-Campos, A.; Tapia, A.; Hernández-Luis, F.; Cedillo, R.; Gonzáles, J.; Fernández, M. A.; Martínez-Grueiro, M.; Castillo, R. *Bioorg. Med. Chem.* **2003**, *11*, 4615.
6. Navarrete-Vázquez, G.; Rojano-Vilchis, M.; Yépez, L.; Meléndez, V.; Gerena, L.; Hernández-Campos, A.; Hernández-Luis, F.; Castillo, R. *Eur. J. Med. Chem.* **2006**, *41*, 135.
7. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.

8. Medina-Franco, J. L.; Rodríguez-Morales; Juárez-Gordiano, C.; Hernández-Campos, A.; Castillo, R. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 345.
9. Yamagami, C.; Akamatsu, M.; Motohashi, N.; Hamada, S.; Tanahashi, T. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2845.
10. López-Rodríguez, M.; Murcia, M.; Benhamú, B.; Viso, A.; Campillo, M.; Pardo, L. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2807.
11. López-Rodríguez, M.; Murcia, M.; Benhamú, B.; Viso, A.; Campillo, M.; Pardo, L. *J. Med. Chem.* **2002**, *45*, 4806.
12. Vaz, R. J.; Maynard, G. D.; Kudlacz, E. M.; Bratton, L. D.; Kane, J. M.; Shatzer, S. A.; Knippenberg, R. W. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2825.
13. Temiz-Arpaci, O.; Tekiner-Gulbas, B.; Yildiz, I.; Aki-Sener, E.; Yalcin, I. *Bioorg. Med. Chem.* **2005**, *13*, 6354.
14. Bhongade, B. A.; Gadad, A. K. *Bioorg. Med. Chem* **2004**, *12*, 2797.
15. Pospisil, P.; Ballmer, P.; Scapozza, L.; Folkers, G. *J. Recept. Signal Transduct.* **2003**, *23*, 36.
16. Wiench, J. W.; Koprowski, M.; Stefaniak, L.; Webb, G. A. *Polish J. Chem.* **2002**, *76*, 525.
17. Houben, L.; Ramaekers, R.; Adamowicz, L.; Maes, G. *Internet Electron. J. Mol. Des.* **2004**, *3*, 163.
18. Luyten, I.; Pankieewicz, K. W.; Watanabe, K. A.; Chattopadhyaya, J. *J. Org. Chem.* **1998**, *63*, 1033.
19. Anandan, K.; Kolandaivel, P.; Kumaresan, R. *J. Mol. Struct.: THEOCHEM* **2004**, *686*, 83.
20. Duarte, H. A.; Carvalho, S.; Paniago, E. B.; Simas, A. M. *J. Pharm. Sci.* **1999**, *88*, 111.
21. Civcir, P. Ü. *J. Mol. Struct.: THEOCHEM* **2000**, *532*, 157.
22. Senthilkumar, L.; Kolandaivel, P. *J. Mol. Struct.: THEOCHEM.* **2003**, *638*, 69.
23. Kwiatkoski, M. J.; Leszczynski, J. *Chem. Phys. Lett.* **1993**, *204*, 430.
24. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Mod.* **2002**, *20*, 269.
25. Golbraikh, A.; Tropsha, A. *J. Comput-Aid. Mol. Des.* **2002**, *16*, 357.
26. Occhiato, E. G.; Ferrali, A.; Menchi, G.; Guarna, A.; Danza, G.; Comerci, A.; Mancina, R.; Serio, M.; Recanatini, M.; Gianni Garotta, A.; de Vivo Cavalli, M. *J. Med. Chem.* **2004**, *47*, 3546.
27. Andrzejewska, M.; Yépez-Mulia, L.; Cedillo-Rivera, R.; Tapia, A.; Vilpo, L.; Kazimierczuk, J. *Eur. J. Med. Chem.* **2002**, *12*, 973.
28. Tripos, Inc., St. Louis, MO.
29. Gasteiger, J.; Marsilli, . M. *Tetrahedron* **1980**, *36*, 3219.
30. Purcel, W.; Singer, P. *J. Chem. Eng. Data* **1967**, *12*, 235.
31. Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
32. Wavefunction, Inc., Irvine, CA.
33. Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.
34. Stewart, J. P. *J. Comput. Chem.* **1989**, *10*, 209.