European Journal of Medicinal Chemistry 46 (2011) 3499–3508

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) of some benzimidazole derivatives with trichomonicidal activity

Jaime Pérez-Villanueva [a], José L. Medina-Franco [b], Thomas R. Caulfield [b], Alicia Hernández-Campos [a], Francisco Hernández-Luis [a], Lilián Yépez-Mulia [c], Rafael Castillo [a,*]

[a] Facultad de Química, Departamento de Farmacia, UNAM, México, DF 04510, Mexico
[b] Torrey Pines Institute for Molecular Studies, Port St. Lucie, FL 34987, USA
[c] Unidad de Investigación Médica en Enfermedades Infecciosas y Parasitarias, IMSS, México, DF 06720, Mexico

ARTICLE INFO

Article history:
Received 5 November 2010
Received in revised form
6 May 2011
Accepted 7 May 2011
Available online 13 May 2011

Keywords:
Benzimidazole
3D-QSAR
CoMFA
CoMSIA
Structure–activity relationships
Trichomonas vaginalis

ABSTRACT

Trichomonosis is a common sexually transmitted infectious disease linked to reproductive health complications. Recently, the benzimidazole nucleus has emerged as a promising scaffold to develop new trichomonicidal agents. Despite the fact that large amounts of experimental data have been accumulated over the past eight years, no quantitative studies have yet been reported on this class of compounds. In our effort to develop new antiparasitic benzimidazole derivatives, we report in this paper CoMFA and CoMSIA studies with an initial set of 70 benzimidazole derivatives with trichomonicidal activity. Four CoMFA models and eight CoMSIA models were generated; ten of these models had values of $r^2 > 0.6$ and $q^2 > 0.5$. The best CoMFA model had $r^2 = 0.936$ and $q^2 = 0.634$, and the best CoMSIA model had $r^2 = 0.858$ and $q^2 = 0.642$. These models were generated by using two conformer selection methodologies (minimum energy conformations and 3D similarity), and three charge types (Mulliken, Gasteiger-Hükel and electrostatic potential atomic charges). The putative active tautomers of 1*H*-benzimidazole derivatives were selected using 3D-QSAR calculations. All models were validated via an external test set with 13 molecules. The best models satisfied additional validation criteria. The contour maps generated show the most important features that a benzimidazole derivative should have for trichomonicidal activity; they also, suggest that substituents at the 2- and 6-positions are important in the generation of derivatives with strong activity.

1. Introduction

Trichomonosis is a sexually transmitted disease caused by the protozoan *Trichomonas vaginalis*. According to the World Health Organization (WHO), more than 170 million new cases are estimated annually worldwide [1]. In women, infection can cause profound inflammation of the genital tract which has been associated with preterm labor, low-birth weight, sterility, cervical cancer and a predisposition to HIV infection [2,3]. In males, *T. vaginalis* infection is generally asymptomatic; however, in some cases it causes urethritis and prostatitis [2,3]. Metronidazole is the drug of choice against this infection; never the less, its side effects and the development of resistant strains have stimulated our interest to search for new therapeutic agents [3,4].

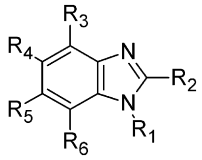
Compounds with the benzimidazole nucleus are commonly encountered in drugs having a broad diversity of pharmacological activities: anthelmintic, gastroprotective, antitumoral, antiviral, antifungical, and antibacterial [5]. As part of our efforts to find new benzimidazole derivatives as antiparasitic agents, a large number of this type of compounds against *T. vaginalis*, *Entamoeba histolytica* and *Giardia intestinalis* have previously been synthesized and tested [6–10]. The trichomonicidal benzimidazole derivatives that have been synthesized, together with that reported by another group [11,12], make up an important database in order to study the structure–activity relationships and propose new potentially active compounds using computer-aided drug design [13–16].

Comparative Molecular Field Analysis (CoMFA) [17,18] and Comparative Molecular Similarity Indices Analysis (CoMSIA) [19] are two powerful methodologies designed to study quantitative

* Corresponding author. Tel.: +52 5 56 22 52 87; fax: +52 5 56 22 53 29.
E-mail address: rafaelc@servidor.unam.mx (R. Castillo).

0223-5234/$ – see front matter © 2011 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2011.05.016

Table 1
Chemical structure and trichomonicidal activity of benzimidazole derivatives used in this study.



	R_1	R_2	R_3	R_4	R_5	R_6	pIC_{50}[a]
1	H	CF_3	H	H	H	H	5.50
2	H	CF_3	H	Cl	H	H	6.35
3[b]	H	CF_3	H	F	H	H	5.50
4	H	CF_3	H	CF_3	H	H	6.63
5	H	CF_3	H	CN	H	H	5.64
6	CH_3	CF_3	H	CF_3	H	H	5.39
7	CH_3	CF_3	H	H	CF_3	H	5.27
8	H	CF_3	H	$SCH_2CH_2CH_3$	H	H	6.46
9	CH_3	CF_3	H	$SCH_2CH_2CH_3$	H	H	6.70
10	CH_3	CF_3	H	H	$SCH_2CH_2CH_3$	H	5.59
11[b]	H	CF_3	H	COC_6H_5	H	H	4.55
12	CH_3	CF_3	H	COC_6H_5	H	H	4.53
13	CH_3	CF_3	H	H	COC_6H_5	H	4.97
14	H	H	H	H	H	H	6.44
15	H	CH_3	H	H	H	H	6.52
16	H	NH_2	H	H	H	H	6.54
17	H	SH	H	H	H	H	6.71
18	H	H	H	Cl	H	H	6.69
19	H	CH_3	H	Cl	H	H	6.79
20	H	NH_2	H	Cl	H	H	6.71
21	H	SH	H	Cl	H	H	6.87
22[b]	H	SCH_3	H	Cl	H	H	7.03
23	H	NH_2	H	Cl	Cl	H	6.98
24	H	CF_3	H	Br	H	H	5.80
25	H	CF_3	H	Br	Br	H	6.66
26	H	CF_3	Br	H	Br	H	6.72
27[b]	H	CF_3	Br	Br	Br	H	6.57
28	H	CF_3	Br	Br	Br	Br	8.70
29	H	C_2F_5	H	Cl	Cl	H	6.52
30	H	CF_3	H	NO_2	NO_2	H	6.24
31	H	C_2F_5	Br	Br	Br	Br	5.00
32	H	SCH_2CH_2OH	Cl	H	Cl	H	6.16
33	H	SCH_2CH_2OH	Br	H	Br	H	6.83
34	H	$SCH_2CH_2N(CH_3)_2$	Cl	H	Cl	H	7.14
35	H	$SCH_2CH_2N(CH_3)_2$	Br	H	Br	H	6.51
36[b]	H	$SCH_2CH_2N(C_2H_5)_2$	Cl	H	Cl	H	7.39
37	H	$SCH_2CH_2N(C_2H_5)_2$	Br	H	Br	H	7.53
38	H	$SCH_2CH_2CH_2N(CH_3)_2$	Cl	H	Cl	H	7.07
39	H	$SCH_2CH_2CH_2N(CH_3)_2$	Br	H	Br	H	7.99
40	H	SCH_2CH_2-(*N*-piperidyl)	Cl	H	Cl	H	7.14
41[b]	H	SCH_2CH_2-(*N*-piperidyl)	Br	H	Br	H	7.69
42	H	SCH_2CH_2-(*N*-morpholinyl)	Cl	H	Cl	H	7.29
43	H	SCH_2CH_2-(*N*-morpholinyl)	Br	H	Br	H	7.23
44	H	SCH_2CH_2-(*p*-nitrophenyl)	Cl	H	Cl	H	7.96
45	H	SCH_2CH_2-(*p*-nitrophenyl)	Br	H	Br	H	8.48
46[b]	CH_3	$CONH_2$	H	H	Cl	H	6.96
47	CH_3	$CONHCH_3$	H	H	Cl	H	6.98
48[b]	CH_3	$CON(CH_3)_2$	H	H	Cl	H	6.63
49	CH_3	$COOCH_2CH_3$	H	H	Cl	H	7.72
50	CH_3	$CONH_2$	H	Cl	H	H	6.73
51	CH_3	$CONHCH_3$	H	Cl	H	H	6.45
52	CH_3	$CON(CH_3)_2$	H	Cl	H	H	6.68
53[b]	CH_3	$COOCH_2CH_3$	H	Cl	H	H	7.57
54[b]	CH_3	$CONH_2$	H	Cl	Cl	H	6.87
55	CH_3	$CONHCH_3$	H	Cl	Cl	H	6.65
56	CH_3	$CON(CH_3)_2$	H	Cl	Cl	H	7.12
57	CH_3	$COOCH_2CH_3$	H	Cl	Cl	H	7.53
58	CH_3	$CONH_2$	H	H	H	H	6.78
59[b]	CH_3	$CONHCH_3$	H	H	H	H	6.98
60	CH_3	$CON(CH_3)_2$	H	H	H	H	6.37
61	CH_3	$COOCH_2CH_3$	H	H	H	H	7.07
62	CH_3	$NHCOOCH_3$	H	H	H	H	5.88
63	CH_3	$NHCOOCH_3$	H	Cl	H	H	6.36
64	CH_3	$NHCOOCH_3$	H	H	Cl	H	6.46

Table 1 (*continued*)

	R_1	R_2	R_3	R_4	R_5	R_6	pIC_{50}[a]
65[b]	CH_3	$NHCOOCH_3$	H	Cl	Cl	H	6.27
66	H	SCH_3	H	$CONH_2$	H	H	6.67
67	H	SCH_3	H	$CONHCH_3$	H	H	6.79
68	H	SCH_3	H	$CON(CH_3)_2$	H	H	7.54
69[b]	H	SCH_3	H	$CONHCH_2CH_3$	H	H	7.26
70	H	SCH_3	H	$CON(CH_2CH_3)_2$	H	H	6.05

[a] Activity data are expressed in $pIC_{50} = -\log IC_{50}$.
[b] Compounds in test set.

structure–activity relationships (QSAR). In previous studies, the CoMFA methodology has been successful when applied in the study of benzimidazole derivatives as antitumorals [20], antivirals [21], $5HT_4$ agonists [22] and recently our research group reported studies against *E. histolytica* [23]. In this paper, for the first time CoMFA and CoMSIA studies with a database of benzimidazole derivatives active against *T. vaginalis* are reported. The QSAR models show high quantitative correlations with good predictive abilities, highlighting the most important features in the design of new benzimidazole derivatives with trichomonicidal activity.

2. Materials and methods

2.1. Database and biological activity

The data set used in this study contained 70 benzimidazole derivatives which have been previously published (Table 1). All compounds have been synthesized by our group [8–10], except **24**–**45**, reported by Andrzejewska et al. [11,12]. All trichomonicidal activities were experimentally determined in the same laboratory using the *T. vaginalis* GT3 strain [8–12]. IC_{50} values were converted to $pIC_{50} = -\log IC_{50}$, using the activity in mol/L.

2.2. Training and test set selection

The initial set of 70 compounds was divided into a training set with 57 (81%) molecules and a test set with 13 (19%) molecules. The activities of the training set ranges from 2 nM ($pIC_{50} = 8.70$) to 29,512 nM ($pIC_{50} = 4.53$); median $pIC_{50} = 6.70$. The activities of the test set ranges from 20 nM ($pIC_{50} = 7.69$) to 28,440 nM ($pIC_{50} = 4.55$); median $pIC_{50} = 6.98$. Worthy of mention is the fact that the structural diversity and activity range of the test set are comparable with the training set [18,24,25].

2.3. Conformer generation

Molecular modeling calculations were completed using the Spartan'02 software package and performed on a Dell Pentium IV PC running Red Hat Enterprise Linux 4 operative system [26]. Conformational analysis was conducted by a systematic search with increments of 60° for each rotatable bond with the MMFF94 force field; the conformational search was restricted to conformers with best energies within a 10 kcal/mol range. The conformers obtained and the molecules with no rotatable bonds were further optimized using the PM3 semiempirical method. Atomic charges (Mulliken and electrostatics) [27,28] were calculated with PM3 implemented in Spartan'02 software [29]. Gasteiger-Hückel atomic charges [30] were calculated with Sybyl 7.3 software [31].

Minimum energy conformers were selected. 3D similarity conformer selection was carried out as an alternate approach. Molecular overlap was conducted with the Rapid Overlay of Chemical Structures (ROCS) approach [32], using **9**, **10** and **45** as query structures on its minimum energy conformations. The combo score, comprised of the Tanimoto shape index and color score, was used to select conformers. The Tanimoto index was calculated with equation (1),

$$T_{A,B}\frac{O_{A,B}}{O_{A,A} + O_{B,B} - O_{A,B}} \quad (1)$$

where $O_{A,A}$ is the molecule A volume, $O_{B,B}$ is the molecule B volume and $O_{A,B}$ is the overlap volume between A and B [33,34]. Color score was calculated with a color force field that included parameters for hydrogen bond donors, hydrogen bond acceptors, hydrophobic groups, anions, cations and rings [34,35].

2.4. Molecular alignment

All compounds were aligned using a benzimidazole nucleus as a template, which is a common substructure in all molecules and the minimum scaffold required for active molecules. The molecular target and action mechanism for this class of compounds are still unknown. In this paper, it is assumed that all the molecules have the same action mechanism and the same binding mode.

2.5. Tautomeric form selection

Representative tautomeric forms were selected employing the methodology called "the best predicted tautomer" [23]. This methodology is based on 3D-QSAR activity predictions and their relations with experimental activity. A first CoMFA model, including molecules with no tautomeric forms, was generated to predict the activity of molecules with tautomers. The best consensus between the experimental activity and the predicted activity for each molecule was used to select the putative active tautomeric form.

2.6. CoMFA methodology

CoMFA studies were performed with Sybyl 7.3. Steric, and electrostatic fields were obtained with Lennard-Jones potential and Coulomb potential, respectively. 3D cubic lattice with grid spacing of 2 Å and extending to 4 Å beyond the aligned molecules in all directions was used. Steric and electrostatic interactions were calculated utilizing a sp^3 carbon probe atom with a charge of +1 with a distance-dependent dielectric at each lattice point. Energy truncation values of 30 kcal/mol were set for the steric and electrostatic interactions.

2.7. CoMSIA methodology

Similarity indices based on five different properties were used. Steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor descriptors were generated using the same lattice box defined in CoMFA studies, and employing the sp^3 atom with charge +1 and radius of 1 Å. Similarity indices ($A_{F,k}$) for

the molecule j and the atom i at grid point q were calculated with equation (2),

$$A_{F,k}^{q}(j) = \sum_{i=1}^{n} \omega_{probe,k}\omega_{ik}e^{-\alpha r_{iq}^{2}} \quad (2)$$

where ω_{ik} is the actual value of the physicochemical propriety k of atom i; $\omega_{probe,k}$ indicates the probe atom with charge +1, radius 1 Å, hydrophobicity +1, hydrogen bond donor and acceptor property +1; α is the attenuation factor; r_{iq} is the mutual distance between the probe atom at grid point q and atom i of the test molecule. In this paper, steric indices are related to the third power of the atomic radii, electrostatic descriptors are derived from partial atomic charges, hydrophobic descriptors are derived from atom-based parameters [36], and H-bond donor and acceptor indices are obtained by a rule-based method based on experimental results [19]. The default value of 0.3 was used as the attenuation factor for the Gaussian-type distance r_{iq}.

2.8. Partial least squares analysis

Partial least squares analysis (PLS) implemented in Sybyl 7.3 was used to generate a linear regression that correlates descriptors with biological activities in pIC_{50}, cross validation being used to obtain the optimum number of the principal components. Cross validation was carried out by the Leave-One-Out (LOO) methodology, and the correlation coefficient q^2 was calculated with equation (3),

$$q^2 = 1 - \frac{\sum\left(Y_{pred} - Y_{\exp}\right)^2}{\sum\left(Y_{\exp} - Y_{mean}\right)^2} \quad (3)$$

where Y_{pred}, $Y_{\exp}$ and Y_{mean} are the values for the predicted activity, experimental activity and mean activity, respectively.

2.9. Validation

Validation was carried out considering the q^2 coefficient and predicting the activity of an external test set with 13 compounds. The trichomonicidal activities of the test set was predicted using all the CoMFA and CoMSIA models and were analyzed by regression analysis, applying the validation criteria proposed by Golbraikh and Tropsha [24,25]:

I. $q^2 > 0.5$
II. $R^2 > 0.6$
III. $[(R^2 - R_0^2)/R^2] < 0.1$ or $[(R^2 - R_0'^2)/R^2/] < 0.1$
IV. $0.85 \le k \le 1.15$ or $0.85 \le k' \le 1.15$

where, q^2 is the cross-validated correlation coefficient from LOO; R^2 is the correlation coefficient for experimental (y) vs. predicted ($\tilde{y}$) activities for test set molecules; ${R_0}^2$ and ${R'_0}^2$ are the correlation coefficients for the regression through origin for y vs. $\tilde{y}$ and $\tilde{y}$ vs. y respectively; k and k' are the slopes for regression through origin $y^{r0} = k\tilde{y}$ and $\tilde{y}^{r0} = k'y$ and were calculated by equations (4) and (5).

$$k = \frac{\sum y_i\tilde{y}_i}{\tilde{y}_i^2} \quad (4)$$

$$k' = \frac{\sum y_i\tilde{y}_i}{y_i^2} \quad (5)$$

The models that satisfied the four conditions above (I–IV) were considered highly predictive and close to being ideal models.

3. Results and discussion

A crucial step for the development of successful models of CoMFA and CoMSIA is the selection of conformers, especially when the biological target is unknown and the molecules in the study have substituents with several rotatable bonds [13]. Models based on low-energy structures have been useful in several papers [37–39]. In this paper, the minimum energy conformation for each compound obtained from conformational analysis was selected. In an alternate approach, there were employed 3D similarity methods based on the fact that ligand-receptor binding is generally characterized by a high molecular complementarity [40].

It is very important to consider tautomeric equilibrium because different tautomers have different hydrophobicity, pKa, 3D shape and electrostatic properties; furthermore, in some cases the stablest tautomer in solution is not necessarily the most biologically active, especially when the conversion barriers are low [41–43]. Most of the compounds in this paper can exist in two or more tautomeric forms. In order to select a representative tautomer, a CoMFA model was developed using benzimidazole derivatives with no tautomers. The model was used to predict the activity of the putative tautomeric forms of benzimidazole derivatives. All tautomers with the best prediction were selected [23]. The most important tautomeric forms considered are shown in Fig. 1.

3.1. CoMFA models constructed for tautomer selection approach

First, a CoMFA model 1 with 28 molecules and no tautomeric forms was constructed. These molecules were in their lowest

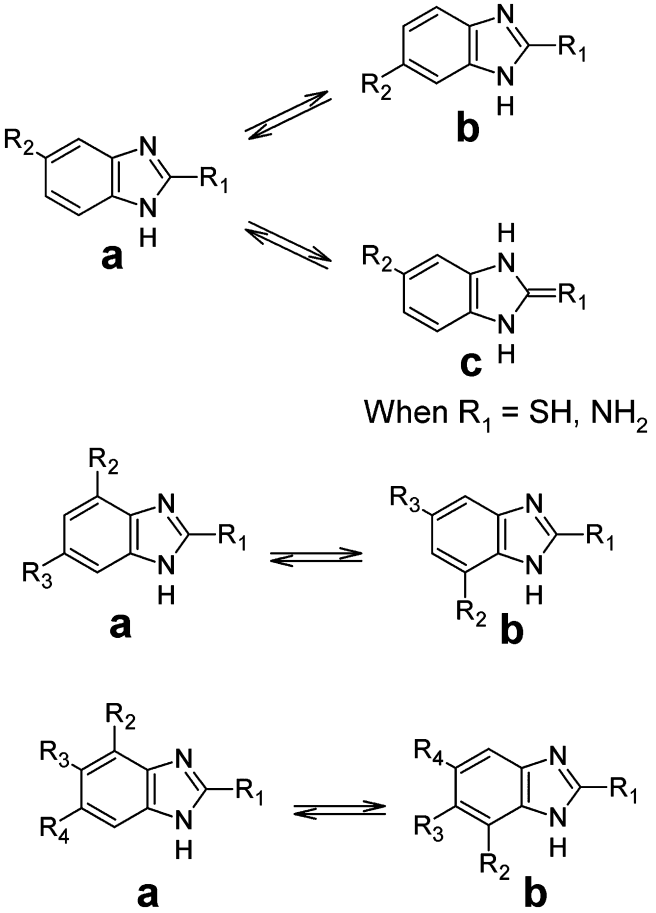


Fig. 1. Tautomeric forms considered in this paper, their identification being based on substitution patterns.

energy conformation with electrostatic potential atomic charges. The results for this model showed poor correlation and predictive ability ($r^2 = 0.443$ and $q^2 = 0.043$; Table 2). Compounds **28** and **31** were clearly responsible for the poor correlation. Then, model 2 with 26 compounds (excluding **28** and **31**) was constructed, the results showed a notable enhancement in correlation and LOO coefficient $r^2 = 0.975$ and $q^2 = 0.503$. This model was considered acceptable to predict activity values based on its q^2 value > 0.5. Interestingly, the differences in the correlations for models based on 28 and 26 compounds are very high. The presence of compounds with behavior similar to that of **28** and **31** has been extensively studied and has been associated with different action mechanism, binding mode and flexible binding sites [44–46], and the presence of activity cliffs (molecules with small changes in structure with large differences in activity) [47,48]. In all cases, the activity of these molecules cannot be explained by conventional QSAR models. Although in this work there is no experimental evidence to explain the "outlier" behavior of compounds **28** and **31**, the presence of a bromine atom at 4- to 7-positions of the benzimidazole nucleus can be an important feature to explain the discontinuous SAR [49].

An additional feature to consider is that compounds with larger substituents at the 5- or 6-positions (i.e. **12**–**13**) have two low-energy rotamers with the same energy. These rotamers are oriented at the opposite sides of the plane defined by the benzimidazole nucleus. For this reason, in addition to model 2, models 3–5 clustering the molecules by different rotamers were generated (Fig. 2). The best results were observed for model 2 ($q^2 > 0.5$), which was used to estimate the activity of molecules with tautomeric forms and select the best predicted tautomer as putatively responsible for trichomonicidal activity. The values for experimental and predicted activities are shown in Table 3.

3.2. CoMFA and CoMSIA models based on minimum energy conformers

To generate CoMFA and CoMSIA models, which are based on minimum energy conformers, only compounds with no tautomers and the best predicted tautomeric forms were employed to give a total of 55 molecules. CoMFA model 6 was generated with electrostatic potential atomic charges calculated with PM3. The model had a high correlation and LOO values with $r^2 = 0.936$ and $q^2 = 0.634$, respectively (Table 4). An important fraction value of 0.475 of electrostatic contribution was observed for this model. Based on this

Table 2
Summary of CoMFA models 1–5.

Statistics	Model				
	1	2	3	4	5
q^{2}[a]	0.043	0.503	0.465	0.472	0.433
PRESS[b]	0.895	0.615	0.638	0.618	0.640
Compounds[c]	28	26	26	26	26
ONC[d]	1	6	6	5	5
r^{2}[e]	0.443	0.975	0.975	0.964	0.971
S[f]	0.683	0.138	0.137	0.161	0.144
F[g]	20.670	123.045	124.608	107.995	135.622
Steric[h]	0.518	0.501	0.509	0.520	0.555
Electrostatic[i]	0.482	0.499	0.491	0.480	0.445

[a] Cross-validated correlation coefficient from LOO.
[b] Standard error of predictions derived from LOO method.
[c] Number of compounds in the training set.
[d] Optimum number of principal components.
[e] Non cross-validated r^2.
[f] Standard error estimate.
[g] F-test value.
[h] Steric field contribution.
[i] Electrostatic field contribution.

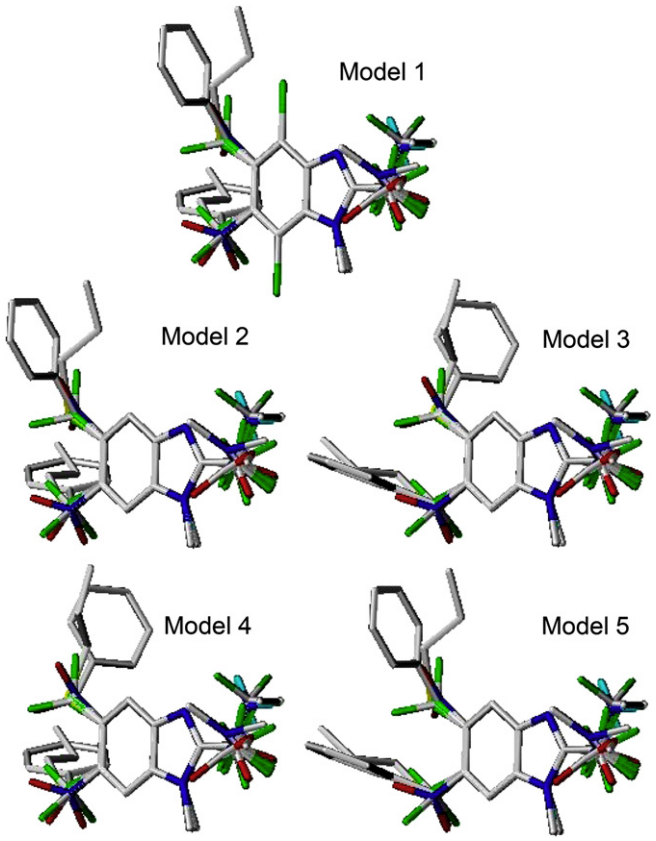


Fig. 2. Orientation of substituents at 5- to 6-positions of benzimidazole in CoMFA models 1–5. Non-polar hydrogens were removed for clarity.

observation it was decided to explore the Mulliken and Gasteiger-Hückel partial atomic charges to generate models 7 and 8, respectively. These models had high correlation ($r^2 = 0.927$ and 0.928, respectively) and good LOO values ($q^2 = 0.602$ and 0.627, respectively); however, model 6 gave the best values (Table 4).

The same molecules used in CoMFA model 6 were used to generate four additional models (9–12) using the CoMSIA methodology (Table 4). Thus, model 9 was constructed using steric and electrostatic fields, model 10 included donor and acceptor of hydrogen bonds fields and model 11 included a hydrophobic field. CoMSIA models 9–11 showed acceptable values of r^2 and q^2; particularly, model 9 had the highest values ($r^2 = 0.915$ and $q^2 = 0.617$). Model 12 was constructed using all fields, and the value of q^2 obtained increased compared with previous models. The difference in r^2 and q^2 values from models 9 and 12 was low; however the contour maps of model 12 were more informative.

3.3. CoMFA and CoMSIA studies based on conformers selected using 3D similarity

CoMFA and CoMSIA models based on minimum energy conformer selection are restricted and require that all minimum energy conformers acquire similar arrangement in 3D-space. However, a restricted space can be defined based on the fact that specific binding of small molecules to target proteins is usually similar and characterized by specific interactions and shape [40]. In this paper, ROCS methodology was used to perceive a similarity between molecules based on their three dimensional shape [34]. This methodology was used for the first time to select the best shape conformation based on query molecules for CoMFA and CoMSIA studies.

Table 3
Experimental and predicted pIC_{50} values by CoMFA model 2.

	Experimental	Predicted	Residual		Experimental	Predicted	Residual
2a	6.35	5.72	0.63	**33a**[a]	6.83	7.07	−0.24
2b[a]		5.94	0.41	**33b**		6.56	0.27
3a[a]	5.50	5.63	−0.13	**34a**[a]	7.14	7.21	−0.07
3b		5.82	−0.32	**34b**		6.66	0.48
4a	6.63	5.60	1.03	**35a**	6.51	7.20	−0.69
4b[a]		5.61	1.02	**35b**[a]		6.70	−0.19
5a[a]	5.64	5.61	0.03	**36a**[a]	7.39	7.26	0.13
5b		5.92	−0.28	**36b**		6.83	0.56
8a	6.46	7.13	−0.67	**37a**[a]	7.53	7.21	0.32
8b[a]		6.03	0.43	**37b**		6.92	0.61
11a[a]	4.55	4.88	−0.33	**38a**[a]	7.07	7.19	−0.12
11b		5.35	−0.80	**38b**		6.92	0.15
16a[a]	6.54	6.90	−0.36	**39a**[a]	7.99	7.22	0.77
16c		5.69	0.85	**39b**		6.65	1.34
17a[a]	6.71	6.37	0.34	**40a**[a]	7.14	7.32	−0.18
17c		5.01	1.70	**40b**		6.74	0.40
18a[a]	6.69	6.69	0.00	**41a**[a]	7.69	7.30	0.39
18b		6.93	−0.24	**41b**		6.77	0.92
19a[a]	6.79	6.68	0.11	**42a**[a]	7.29	7.29	0.00
19b		6.92	−0.13	**42b**		6.75	0.54
20a[a]	6.71	7.06	−0.35	**43a**[a]	7.23	7.22	0.01
20b		7.30	−0.59	**43b**		6.78	0.45
20c		6.02	0.69	**44a**[a]	7.96	7.45	0.51
21a	6.87	6.56	0.31	**44b**		6.89	1.07
21b[a]		6.79	0.08	**45a**[a]	8.48	7.45	1.03
21c		5.20	1.67	**45b**		6.96	1.52
22a	7.03	6.90	0.13	**66a**[a]	6.67	6.65	0.02
22b[a]		7.15	−0.12	**66b**		7.06	−0.39
23a[a]	6.98	7.44	−0.46	**67a**[a]	6.79	6.92	−0.13
23c		6.26	0.72	**67b**		6.98	−0.19
24a[a]	5.80	6.18	−0.38	**68a**	7.54	6.63	0.91
24b		6.32	−0.52	**68b**[a]		6.93	0.61
26a[a]	6.72	6.56	0.16	**69a**[a]	7.26	6.71	0.55
26b		6.00	0.72	**69b**		6.59	0.67
27a[a]	6.57	6.44	0.13	**70a**[a]	6.05	6.28	−0.23
27b		6.02	0.55	**70b**		6.47	−0.42
32a	6.16	7.06	−0.90				
32b[a]		6.48	−0.32				

[a] Selected tautomer based on its predicted and experimental activity.

Based on their substitution patterns, three molecules in their minimum energy conformation were used as queries in this work. The first one, **45**, was selected as query because it has the largest substituent at the 2-position and, besides it, is one of the most active compounds. The other two, **9** and **10**, were selected considering that the long chain was fixed at the 5- and 6-positions, respectively. For the other compounds with rotatable bonds shown in Table 1, the conformers were selected by considering the highest similarity value with query molecules with the same substitution pattern. It was observed that all molecules, overlaid with ROCS, showed an almost perfect alignment of the benzimidazole nucleus. Hence, this alignment was considered for the QSAR calculations.

CoMFA model 13 was constructed with conformers selected by using 3D similarity, electrostatic charges and tautomers defined by model 2. High correlation $r^2 = 0.934$ and LOO coefficient value $q^2 = 0.601$ were obtained (Table 5). These results are comparable with the minimum energy based CoMFA models. In addition, CoMSIA models 14–17 were constructed using the ROCS based conformer selection. 3D similarity based CoMSIA models were constructed using the same fields as minimum energy based models. The results showed that model 14 (with steric and electrostatic fields) and model 17 (with five fields) were the best models with the highest r^2 and q^2 values.

ROCS based models had r^2 and q^2 values very close to those of minimum energy based models. These results suggest that CoMFA models using conformer selection obtained by 3D overlay can be useful when the QSAR models based on minimum energy conformers fail to predict molecules with several rotatable bonds.

3.4. External validation of CoMFA and CoMSIA models

Validation is a crucial step in structure–activity relationship studies. Normally, cross-validation is used to assess internal predictive power of models. However, models with high values of LOO r^2 ($q^2 > 0.5$), are not always predictive [24,25]. For this reason, it is necessary to have an external set with at least five compounds with similar kind of structures and activity values similar to training set in order to have a best validation [18,24,25]. In this paper, 13 compounds were selected as a test set and their activities were calculated using models 6–17. Furthermore, the regression plots of experimental versus predicted activities and predicted versus experimental activities were constructed to apply the conditions described in the section 2.9. The first and second conditions mean that the LOO coefficient q^2 and test set correlation coefficient R^2 should be at least 0.5 and 0.6, respectively. These conditions are necessary but not sufficient to have a highly predictive model. The third condition $((|(R^2 - R_0^2)/R^2|) < 0.1$ or $|(R^2 - R_0'^2)/R^2| < 0.1)$ means that at least one of the correlation coefficients for regressions through the origin ${R_0}^2$ or ${R'_0}^2$ should be close to R^2; whereas the fourth condition ($0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$) means that at least one slope of the regression lines through the origin should be close to one. The application of these conditions suggests that our models are close to being an "ideal" model [24]. The necessary values

Table 4
Summary of statistical data and validation for CoMFA and CoMSIA models 6–12.

Statistics	CoMFA model/Charges			CoMSIA model/Charges			
	6	7	8	9	10	11	12
	Electrostatic potential	Mulliken	Gasteiger-Hückel	Electrostatic potential			
q^2	0.634	0.602	0.627	0.617	0.507	0.503	0.642
PRESS	0.468	0.488	0.472	0.474	0.527	0.535	0.454
ONC	5	5	5	4	2	3	3
r^2	0.936	0.927	0.928	0.915	0.573	0.706	0.858
S	0.196	0.209	0.208	0.223	0.491	0.412	0.286
F	142.893	124.974	125.819	134.483	34.844	40.730	102.338
Steric	0.525	0.554	0.556	0.231			0.084
Electrostatic	0.475	0.446	0.444	0.769			0.285
Donor[a]					0.596		0.227
Acceptor[a]					0.404		0.177
Hydrophobic[b]						1	0.226
R^2[c]	0.890	0.890	0.842	0.873	0.562	0.759	0.809
${R_0}^2$[d]	0.828	0.843	0.809	0.777	0.509	0.663	0.749
k[e]	1.002	0.996	1.004	0.993	0.994	0.999	0.998
$(R^2 - {R_0}^2)/R^2$	0.069	0.052	0.039	0.110	0.094	0.126	0.074
${R'}_0^2$[d]	0.645	0.706	0.650	0.434	−0.790	−0.05	0.416
k'[e]	0.996	1.001	0.993	1.003	0.999	0.996	0.998
$(R^2 - {R'}_0^2)/R'^2$	0.275	0.206	0.228	0.503	2.406	1.066	0.486

q^2, PRESS, ONC, r^2, S, F, steric and electrostatic have the same meaning as in Table 2.
[a] Donor and acceptor of hydrogen bond fields, respectively.
[b] Hydrophobic field contribution.
[c] Correlation coefficient derived from predictions of test set molecules.
[d] Correlation coefficients for the regression through origin for experimental vs. predicted and predicted vs. experimental activity, respectively.
[e] Slopes for regression through origin from experimental vs. predicted and predicted vs. experimental, respectively.

obtained from correlation charts and the results of the application of the conditions are shown in Tables 4 and 5. An example of the regressions for test set predictions for model 6 is shown in Fig. 3. Although both plots may appear redundant, these plots could be characterized by different statistics.

CoMFA models 6–8 satisfied the validation conditions; model 6 in particular showed the best q^2 and R^2 values. In this case model 6 is in agreement that the best q^2 model has higher R^2. On the other hand, CoMSIA models 9–12 showed acceptable q^2 values. Especially model 12, which had the highest q^2 value and the second best R^2 value; however, models 9 and 11 failed to satisfy the third condition and model 10 failed to satisfy the second condition. Therefore, only model 12 is highly predictive. For 3D similarity based models 13 through 17, all of the validation rules were satisfied with the exception of model 15. Although, models 13 through 17 have lower values of q^2 and R^2 than models 6 to 12, they better fulfill the third condition. The predicted values for the test set using the best models are shown in the Table 6.

Table 5
Summary of statistical data and validation for CoMFA and CoMSIA models 13–17.

Statistics	CoMFA model	CoMSIA model			
	13	14	15	16	17
q^2	0.601	0.565	0.482	0.506	0.585
PRESS	0.489	0.510	0.551	0.538	0.484
ONC	5	5	4	4	2
r^2	0.934	0.914	0.669	0.748	0.805
S	0.198	0.227	0.441	0.385	0.332
F	139.397	104.593	25.296	37.061	107.175
Steric	0.499	0.245			0.095
Electrostatic	0.501	0.755			0.307
Donor			0.603		0.247
Acceptor			0.397		0.121
Hydrophobic				1	0.230
R^2	0.825	0.729	0.689	0.678	0.760
${R_0}^2$	0.795	0.692	0.589	0.644	0.726
k	0.995	0.993	0.996	1.009	0.995
$(R^2 - {R_0}^2)/R^2$	0.036	0.051	0.145	0.050	0.045
${R'}_0^2$	0.625	0.301	−0.54	0.135	0.426
k'	1.002	1.002	0.998	0.986	1.001
$(R^2 - {R'}_0^2)/R'^2$	0.242	0.587	1.784	0.801	0.439

All values have the same meaning as they do in Tables 3, 4.

3.5. CoMFA and CoMSIA contour maps

Contour maps for the best CoMFA and CoMSIA models (6 and 12, respectively) are shown in Fig. 4. The discussion of the contour maps is presented in the following six sections.

3.5.1. Steric CoMFA contour maps

Steric contour maps (Fig. 4a) are shown in green when the activity increases with large or bulky substituents. The opposite is true for yellow contours. There is a yellow contour close to the 1-position in the benzimidazole nucleus, which suggests that trichomonicidal activity decreases with methyl substitution. This trend is observed for all benzimidazole derivatives substituted at other positions by small groups (**1**, **2**, **4**–**7**, **14**–**21**, **23**–**26**, and **29**, **30**), where the 1-methylbenzimidazole derivatives **6** and **7** are the least active compounds (pIC_{50} = 5.39 and 5.27, respectively). Indeed, the most active compounds ($pIC_{50} > 7$) have hydrogen at the 1-position with the exception of compounds **49**, **56**, **57**, and **61**, where the substituent at the 2-position plays an important role in activity. This trend is different for compounds **8**–**10**, **12**, and **13**, where the larger substituents at the 5- and 6-positions determine the differences in activity. A favorable steric contour is found at the 2-position of the benzimidazole nucleus. This region is associated with the high activity of the 2-alkylthiobenzimidazole derivatives **32**–**35**, **37**–**40**, **42**–**45**, and the 2-ethyl ester derivatives **49**, **57**, and **61** (pIC_{50} values 6.16–8.48). This group contains most of the active compounds in the data set. A yellow steric unfavorable region at the 5-position and a green contour close to the 6-position indicate that compounds substituted for at the 6-position are more active than those substituted for at the 5-position. This is observed with 1-methylbenzimidazole derivatives **12**, **13**, **47**, **51**, **63**, and **64**. However, this trend is not so clear with compounds **6** and **7** because they have almost the same activity. The opposite trend, on the other

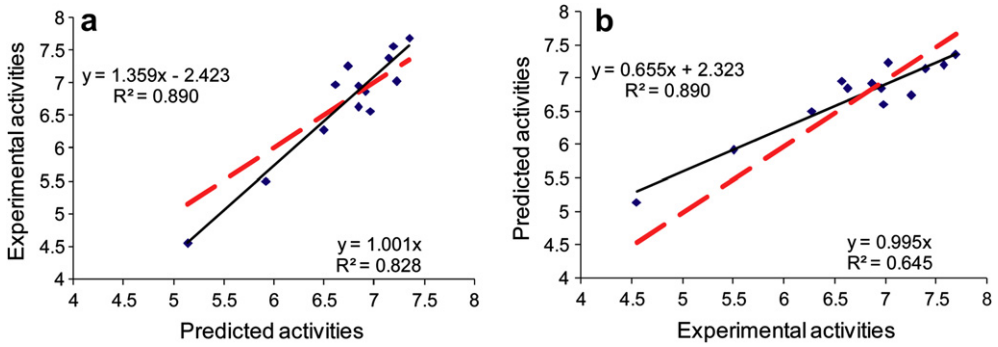


Fig. 3. Validation for CoMFA model 6 using the test set. (a) Experimental vs. predicted activities. (b) Predicted vs. experimental activities.

hand, is observed for compounds **9** and **10**. The difference in activity between the latter two compounds (pIC_{50} values of 6.7 and 5.59, respectively) can explain the yellow contour at the border of the 6-position, thus suggesting that large substituents at this position decreases the activity.

3.5.2. Electrostatic CoMFA contour maps

Electrostatic contour maps are shown in red when the activity increases with negative atomic charges and in blue when the activity decreases (Fig. 4b). In general, red contour maps are close to heteroatoms such as nitrogen and oxygen, whose partial atomic charges are highly negative. Three negative favorable red regions are observed close to the substituents at the 2-position. One of them is close to the oxygen in the carbonyl group of compounds **47**, **49**–**52**, **55**–**58**, and **60**–**61**. These compounds showed moderate activity for amide derivatives (pIC_{50} 6.37–7.12) and strong activity for ester derivatives (pIC_{50} 7.07–7.72). A second one is extended from the amide nitrogen and the alkyl oxygen, for the same group of compounds (**47**, **49**–**52**, **55**–**58**, and **60**–**61**), to the nitrogen of the amino group in the high active compounds **34**, **37**–**40**, and **42**–**43** ($pIC_{50} > 7$). The last red contour is at the extreme of the 2-position. It can be associated with the nitro group in two of the more active compounds **44** and **45** ($pIC_{50} = 7.96$ and 8.48, respectively). Besides, a blue contour at the 2-position is shown to be close to the aliphatic region of the moderately to highly active compounds **32**–**35**, **37**–**40**, and **42**–**45** (pIC_{50} values 6.16–8.48). The same contour is also close to the ethyl chain in the potent ethyl ester derivatives **49**, **57**, and **61** ($pIC_{50} > 7$). Three red contours close to the 5- and 6-positions suggest that negative charged groups such as halogen atoms increase the activity as compared to the hydrogen atom. Furthermore, larger substituents are associated with a blue contour, which suggests the favorable effect of the aliphatic portion in the substituent propylthio **8**–**10** ($pIC_{50} = 5.59$–6.70), and alkyl amides **68** and **70** ($pIC_{50} = 6.05$–7.54) as compared with the electron rich benzoyl substituent in the less active compounds **12**, **13** ($pIC_{50} = 4.53$ and 4.97, respectively).

3.5.3. Steric CoMSIA contour maps

Coloring in CoMSIA steric contour maps (Fig. 4c) is similar to the color schemes for the CoMFA maps; however, although both maps have similar appearance, that of CoMSIA offer a greater level of detail. An unfavorable yellow CoMSIA contour is observed close to the 1-position of the benzimidazole with a better definition than that of CoMFA and suggests the favorable effect of hydrogen substitution. A green contour at the 2-position is shown, suggesting the favorable substitution effect in this position with large substituents. In agreement with CoMFA, a prominent unfavorable yellow contour is observed at the 5- and 6-positions, indicating the unfavorable effect of the substituents at the 5-position and the large substituents at the 6-position.

3.5.4. Electrostatic CoMSIA contour maps

The colors of the electrostatic contour maps in the CoMSIA models (Fig. 4d) have the same meaning as those of the CoMFA models. Similar to CoMFA, a blue contour map is observed at the 2-position of the benzimidazole. However, unlike CoMFA, the CoMSIA contour is larger and covers the aliphatic region of **32**–**35**, **37**–**40** and **42**–**43**, and the deactivated aromatic ring of **44** and **45**. Also similar to CoMFA, a second blue contour is shown at the 5- and 6-positions. This contour is close to the aromatic ring of inactive compound **13** ($pIC_{50} = 4.97$) and to the oxygen in carbonyl groups of **12**, **66**, **67**, and **70** ($pIC_{50} = 4.53$–6.79, median $pIC_{50} = 6.36$).

3.5.5. Hydrogen bond donor and acceptor CoMSIA contour maps

Areas favored by donors and acceptors are shown in cyan and magenta respectively; unfavorable areas are in purple and red, respectively (Fig. 4e). At the 1-position of the benzimidazole nucleus the contour is indicated in cyan, which suggests favorable contributions of hydrogen in the 1*H*-benzimidazole derivatives. This observation is in agreement with the yellow CoMFA steric contour generated by the 1-methylbenzimidazole derivatives. This contour can be associated with the higher activity of most of the 1*H*-benzimidazole derivatives in the database. A purple unfavorable hydrogen bond donor contour map is close to the substituents at the 2-position. The purple contour is associated with the decrement in the activity of the hydroxyl substituents of **32** and **33** ($pIC_{50} < 7$). Furthermore, this contour is close to the *N*-H of the amide type of compounds that are less active than are the ester

Table 6
Training set predictions using models 6, 12, 13 and 17.

Model		6		12		13		17	
	Exp.[a]	Pred.[b]	Res.[c]	Pred.	Res.	Pred.	Res.	Pred.	Res.
3a	5.50	5.92	−0.42	6.18	−0.68	5.88	−0.38	6.09	−0.59
11a	4.55	5.14	−0.59	5.4	−0.85	5.17	−0.62	5.33	−0.78
22b	7.03	7.23	−0.2	7.32	−0.29	7.24	−0.21	7.16	−0.13
27a	6.57	6.96	−0.39	6.71	−0.14	6.89	−0.32	6.56	0.01
36a	7.39	7.14	0.25	7.36	0.03	7.41	−0.02	7.51	−0.12
41a	7.69	7.35	0.34	7.52	0.17	7.28	0.41	7.42	0.27
46	6.96	6.84	0.12	6.81	0.15	7.04	−0.08	6.91	0.05
48	6.63	6.84	−0.21	6.97	−0.34	6.8	−0.17	7.25	−0.62
53	7.57	7.19	0.38	7.09	0.48	7.26	0.31	7.13	0.44
54	6.87	6.91	−0.04	6.81	0.06	7.18	−0.31	6.92	−0.05
59	6.98	6.61	0.37	6.74	0.24	6.3	0.68	6.77	0.21
65	6.27	6.50	−0.23	6.25	0.02	6.62	−0.35	6.46	−0.19
69a	7.26	6.74	0.52	6.53	0.73	6.8	0.46	6.38	0.88

[a] Experimental activity.
[b] Predicted activity.
[c] Residual.

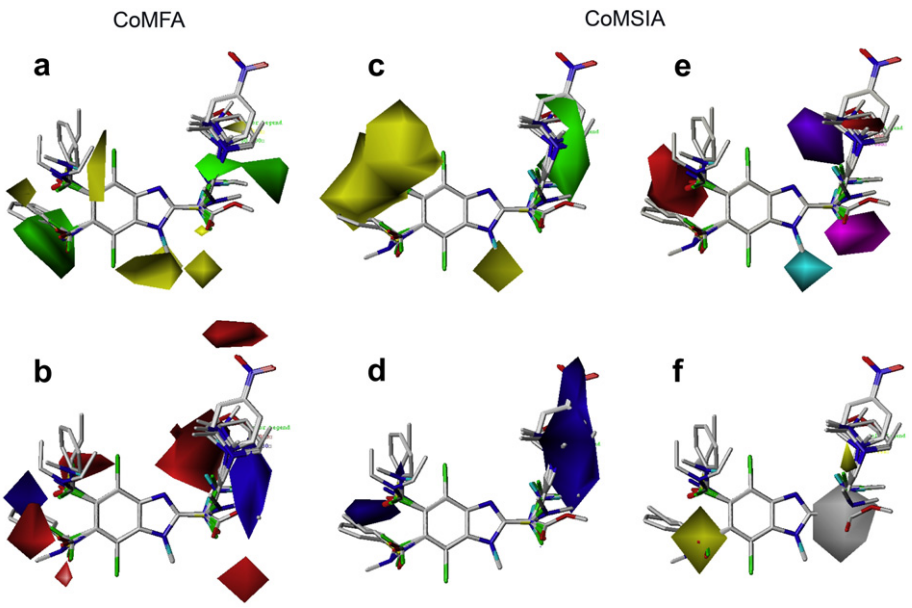


Fig. 4. CoMFA and CoMSIA STDEV*COEFF contour maps for the best models. CoMFA model 6: (a) Sterically favored areas are in green, and sterically unfavored areas are in yellow. (b) Negative charge favored areas are in red and unfavored areas are in blue. CoMSIA Model 12: The colors in (c) and (d) have the same meaning as do CoMFA contour maps. (e) Donor and acceptor favored areas are in cyan and magenta, respectively, and donor and acceptor unfavored areas are in purple and red, respectively. (f) Hydrophobic favored areas are in yellow and unfavored areas in gray. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

analogues. A magenta favorable hydrogen bond acceptor contour is shown at the 2-position of benzimidazole and is associated with the carbonyl group in the amide and ester substituents. This contour is in agreement with the CoMFA red electrostatic contour in the same area (Fig. 4b). A hydrogen bond acceptor unfavorable contour in red is shown at the 2-position of the benzimidazole. This contour can be associated with an increment in activity when the linker carbon chain, between acceptor groups and the benzimidazole nucleus, changes from two to three methylene groups in compounds **34**, **35** and **38**, **39** (pIC_{50} 6.51–7.14 to 7.07–7.99, respectively). A second red contour is close to the carbonyl group at the 5-position in some amides and benzoyl derivatives.

3.5.6. *Hydrophobic CoMSIA contour maps*

Hydrophobic contour maps (Fig. 4f) are shown in yellow when the activity increases with hydrophobic groups and in gray when the activity decreases. A gray contour map is shown at the 2-position of the benzimidazole nucleus generated by the poor activity of trifluoromethyl and pentafluoroethyl substituted benzimidazoles (pIC_{50} 4.53–6.98, median pIC_{50} = 6.24). A small favorable yellow contour map is observed at the 2-position, close to the aliphatic portion of the substituents in **38**, **39**. A second favorable yellow hydrophobic contour map is close to the 6-position of the benzimidazole nucleus. The contour can be explained by the presence of the halogen atom, which in most cases produces more active compounds. This contour is in agreement with the green contour at the same position in CoMFA (Fig. 4a).

4. Conclusions

The benzimidazole derivatives included in this study had a reasonable SAR and strong quantitative correlations. However, some exceptions were observed in compounds **28** and **31**, which have small changes in structure but large changes in activity.

The CoMFA and CoMSIA models were generated and showed good r^2 and q^2 values in almost all models. Eight models revealed a good response to test set validation. CoMFA models based on minimum energy conformations presented better test set predictions than did CoMSIA. However, the CoMSIA models offered valuable information, since they explored additional kinds of fields and in some cases gave clearer contour maps.

The CoMFA and CoMSIA models, generated using 3D similarity conformer selection, showed good results and complemented models based on minimum energy. These models could be useful to estimate the activity of some molecules with high rotational degrees of freedom on substituents at the 2-, 5-, and 6-positions.

The best CoMFA model (model 6, $r^2 = 0.936$, $q^2 = 0.634$), based on the steric and electrostatic fields, included electrostatic potential atomic charges and minimum energy conformers. The best CoMSIA model (model 12, $r^2 = 0.858$, $q^2 = 0.642$), based on five fields, included the same charge type and energy conformations as the CoMFA model 6. The analysis of the contour maps for the best CoMFA and CoMSIA models indicates that hydrogen at the 1-position seems to have an important role in activity. Large groups at the 2-position with hydrophobic segments generate potent compounds. Also, heteroatoms at the 2-position can generate regions with negative charges that could act as hydrogen bond acceptors involving the binding site. In general, substituents at the 6-position are more favored than are substituents at the 5-position. Substitution at the 6-position with chlorine or bromine atom is favored over other substitutions in the current data set.

These models were used in our group to design and screen a new series of benzimidazole derivatives that are currently being synthesized, the results of which will be reported in further publications.

Acknowledgements

This work was supported by project 80093 from CONACyT. J.P-V. acknowledges CONACyT scholarship 173896/173896 and Carso Health Institute for the PhD student fellowship S0710443. J.L.M-F acknowledges the State of Florida for funding. The authors are grateful to OpenEye Scientific Software, Inc., for providing ROCS.

References

[1] Global Prevalence and Incidence of Selected Curable Sexually Transmitted Infections: Overview and Estimates. World Health Organization, Geneva, 2001.

[2] J.R. Schwebke, Update of trichomoniasis, Sex. Transm. Infect. 78 (2002) 378–379.

[3] P. Upcroft, J.A. Upcroft, Drug targets and mechanisms of resistance in the anaerobic protozoa, Clin. Microbiol. Rev. 14 (2001) 150–164.

[4] S.L. Cudmore, K.L. Delgaty, S.F. Hayward-McClelland, D.P. Petrin, G.E. Garber, Treatment of infections caused by metronidazole-resistant *Trichomonas vaginalis*, Clin. Microbiol. Rev. 17 (2004) 783–793.

[5] M. Boiani, M. Gonzalez, Imidazole and benzimidazole derivatives as chemotherapeutic agents, Mini Rev. Med. Chem. 5 (2005) 409–424.

[6] G. Navarrete-Vázquez, R. Cedillo, A. Hernández-Campos, L. Yépez, F. Hernández-Luis, J. Valdez, R. Morales, R. Cortés, M. Hernández, R. Castillo, Synthesis and antiparasitic activity of 2-(trifluoromethyl)-benzimidazole derivatives, Bioorg. Med. Chem. Lett. 11 (2001) 187–190.

[7] J. Valdez, R. Cedillo, A. Hernández-Campos, L. Yépez, F. Hernández-Luis, G. Navarrete-Vázquez, A. Tapia, R. Cortés, M. Hernández, R. Castillo, Synthesis and antiparasitic activity of 1*H*-benzimidazole derivatives, Bioorg. Med. Chem. Lett. 12 (2002) 2221–2224.

[8] G. Navarrete-Vázquez, L. Yépez, A. Hernández-Campos, A. Tapia, F. Hernández-Luis, R. Cedillo, J. González, A. Martínez-Fernández, M. Martínez-Grueiro, R. Castillo, Synthesis and antiparasitic activity of albendazole and mebendazole analogues, Bioorg. Med. Chem. 11 (2003) 4615–4622.

[9] G. Navarrete-Vázquez, M.D. Rojano-Vilchis, L. Yépez-Mulia, V. Meléndez, L. Gerena, A. Hernández-Campos, R. Castillo, F. Hernández-Luis, Synthesis and antiprotozoal activity of some 2-(trifluoromethyl)-1*H*-benzimidazole bioisosteres, Eur. J. Med. Chem. 41 (2006) 135–141.

[10] D. Valdez-Padilla, S. Rodríguez-Morales, A. Hernández-Campos, F. Hernández-Luis, L. Yépez-Mulia, A. Tapia-Contreras, R. Castillo, Synthesis and antiprotozoal activity of novel 1-methylbenzimidazole derivatives, Bioorg. Med. Chem. 17 (2009) 1724–1730.

[11] M. Andrzejewska, L. Yépez-Mulia, R. Cedillo-Rivera, A. Tapia, L. Vilpo, J. Vilpo, Z. Kazimierczuk, Synthesis, antiprotozoal and anticancer activity of substituted 2-trifluoromethyl- and 2-pentafluoroethylbenzimidazoles, Eur. J. Med. Chem. 37 (2002) 973–978.

[12] M. Andrzejewska, L. Yépez-Mulia, A. Tapia, R. Cedillo-Rivera, A.E. Laudy, B.J. Starosciak, Z. Kazimierczuk, Synthesis, and antiprotozoal and antibacterial activities of S-substituted 4,6-dibromo- and 4,6-dichloro-2-mercaptobenzimidazoles, Eur. J. Pharm. Sci. 21 (2004) 323–329.

[13] F. Ooms, Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry, Curr. Med. Chem. 7 (2000) 141–158.

[14] W.L. Jorgensen, The many roles of computation in drug discovery, Science 303 (2004) 1813–1818.

[15] J.L. Medina-Franco, F. López-Vallejo, R. Castillo, Diseño de fármacos asistido por computadora, Educ. Quím 17 (2006) 452–457.

[16] W.L. Jorgensen, Efficient drug Lead discovery and Optimization, Acc. Chem. Res. 42 (2009) 724–733.

[17] R.D. Cramer, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis. 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.

[18] H. Kubinyi, Comparative molecular field analysis (CoMFA). in: P. v. R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. SchaeferIII, P.R. Schreiner (Eds.), The Encyclopedia of Computational Chemistry. John Wiley & Sons, Chichester, 1998, pp. 448–460.

[19] G. Klebe, U. Abraham, Comparative Molecular Similarity Index Analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries, J. Comput. Aided Mol. Des 13 (1999) 1–10.

[20] O. Temiz-Arpaci, B. Tekiner-Gulbas, I. Yildiz, E. Aki-Sener, I. Yalcin, 3D-QSAR analysis on benzazole derivatives as eukaryotic topoisomerase II inhibitors by using comparative molecular field analysis method, Bioorg. Med. Chem. 13 (2005) 6354–6359.

[21] P.D. Patel, M.R. Patel, N. Kaushik-Basu, T.T. Talele, 3D QSAR and molecular docking studies of benzimidazole derivatives as hepatitis C virus NS5B polymerase inhibitors, J. Chem. Inf. Model. 48 (2008) 42–55.

[22] M.L. López-Rodríguez, M. Murcia, B. Benhamú, A. Viso, M. Campillo, L. Pardo, Benzimidazole derivatives. 3. 3D-QSAR/CoMFA model and computational simulation for the recognition of 5-HT4 receptor antagonists, J. Med. Chem. 45 (2002) 4806–4815.

[23] F. López-Vallejo, J.L. Medina-Franco, A. Hernández-Campos, S. Rodríguez-Morales, L. Yépez, R. Cedillo, R. Castillo, Molecular modeling of some 1*H*-benzimidazole derivatives with biological activity against *Entamoeba histolytica*: a comparative molecular field analysis study, Bioorg. Med. Chem. 15 (2007) 1117–1126.

[24] A. Golbraikh, A. Tropsha, Beware of q(2)!, J. Mol. Graph. Model. 20 (2002) 269–276.

[25] A. Golbraikh, A. Tropsha, Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection, J. Comput.-Aided Mol. Des 16 (2002) 357–369.

[26] Spartan, Version 02, Wavefunction Inc., Irvine, CA.

[27] R.S. Mulliken, Electronic population analysis on LCAO-MO molecular wave functions, J. Chem. Phys. 23 (1955) 1833–1840.

[28] S.R. Cox, D.E. Williams, Representation of the molecular electrostatic potential by a net atomic charge model, J. Comput. Chem. 2 (1981) 304–323.

[29] J.J.P. Stewart, Optimization of parameters for semiempirical methods I. Method, J. Comput. Chem. 10 (1989) 209–220.

[30] J. Gasteiger, M. Marsili, Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges, Tetrahedron 36 (1980) 3219–3228.

[31] Sybyl, Version 7.3, Tripos Inc., St. Louis, MO.

[32] Rocs, Version 2.3.1, OpenEye Scientific Software Inc., Santa Fe, NM.

[33] P. Willett, J.M. Barnard, G.M. Downs, Chemical similarity searching, J. Chem. Inform. Comput. Sci. 38 (1998) 983–996.

[34] T.S. Rush, J.A. Grant, L. Mosyak, A. Nicholls, A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction, J. Med. Chem. 48 (2005) 1489–1495.

[35] M.J. Sykes, M.J. Sorich, J.O. Miners, Molecular modeling approaches for the prediction of the nonspecific binding of drugs to hepatic microsomes, J. Chem. Inf. Model. 46 (2006) 2661–2673.

[36] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics, J. Chem. Inf. Comput. Sci. 29 (1989) 163–172.

[37] S.S. Kulkarni, A.H. Newman, W.J. Houlihan, Three-dimensional quantitative structure-activity relationships of Mazindol analogues at the dopamine transporter, J. Med. Chem. 45 (2002) 4119–4127.

[38] V. Ravichandran, R.K. Agrawal, Predicting anti-HIV activity of PETT derivatives: CoMFA approach, Bioorg. Med. Chem. Lett. 17 (2007) 2197–2202.

[39] A.K. Chakraborti, B. Gopalakrishnan, M.E. Sobhia, A. Malde, Comparative molecular field analysis (CoMFA) of phthalazine derivatives as phosphodiesterase IV inhibitors, Bioorg. Med. Chem. Lett. 13 (2003) 2473–2479.

[40] G. Klebe, Recent developments in structure-based drug design, J. Mol. Med. 78 (2000) 269–281.

[41] Y.C. Martin, Let's not forget tautomers, J. Comput.-Aided Mol. Des 23 (2009) 693–704.

[42] A.R. Katritzky, C.D. Hall, B. El-Dien, M. El-Gendy, B. Draghici, Tautomerism in drug discovery, J. Comput.-Aided Mol. Des 24 (2010) 475–484.

[43] F. Milletti, A. Vulpetti, Tautomer preference in PDB complexes and its impact on structure-based drug discovery, J. Chem. Inf. Model. 50 (2010) 1062–1074.

[44] K.H. Kim, Outliers in SAR and QSAR: is unusual binding mode a possible source of outliers? J. Comput.-Aided Mol. Des 21 (2007) 63–86.

[45] K.H. Kim, Outliers in SAR and QSAR: 2. Is a flexible binding site a possible source of outliers? J. Comput.-Aided Mol. Des 21 (2007) 421–435.

[46] M.T. Sisay, L. Peltason, J. Bajorath, Structural interpretation of activity cliffs revealed by systematic analysis of structure-activity relationships in analog series, J. Chem. Inf. Model. 49 (2009) 2179–2189.

[47] G.M. Maggiora, On outliers and activity cliffs - Why QSAR often disappoints, J. Chem. Inf. Model. 46 (2006) 1535.

[48] J.L. Medina-Franco, K. Martínez-Mayorga, A. Bender, R.M. Marin, M.A. Giulianotti, C. Pinilla, R.A. Houghten, Characterization of activity Landscapes using 2D and 3D similarity methods: consensus activity cliffs, J. Chem. Inf. Model. 49 (2009) 477–491.

[49] J. Pérez-Villanueva, R. Santos, A. Hernández-Campos, M.A. Giulianotti, R. Castillo, J.L. Medina-Franco, Towards a systematic characterization of the antiprotozoal activity landscape of benzimidazole derivatives, Bioorg. Med. Chem. 18 (2010) 7380–7391.