J Mol Model (2011) 17:921–928
DOI 10.1007/s00894-010-0779-4

ORIGINAL PAPER

Molecular modeling and QSAR studies of a set of indole and benzimidazole derivatives as H_4 receptor antagonists

João Paulo S. Fernandes · Kerly Fernanda M. Pasqualoto · Elizabeth I. Ferreira · Carlos A. Brandt

Received: 11 January 2010 / Accepted: 7 June 2010 / Published online: 6 July 2010

Abstract Histamine is an important biogenic amine, which acts with a group of four G-protein coupled receptors (GPCRs), namely H_1 to H_4 (H_1R – H_4R) receptors. The actions of histamine at H_4R are related to immunological and inflammatory processes, particularly in pathophysiology of asthma, and H_4R ligands having antagonistic properties could be helpful as antiinflammatory agents. In this work, molecular modeling and QSAR studies of a set of 30 compounds, indole and benzimidazole derivatives, as H_4R antagonists were performed. The QSAR models were built and optimized using a genetic algorithm function and partial least squares regression (WOLF 5.5 program). The best QSAR model constructed with training set (N=25) presented the following statistical measures: r^2=0.76, q^2=0.62, LOF= 0.15, and LSE=0.07, and was validated using the LNO and y-randomization techniques. Four of five compounds of test set were well predicted by the selected QSAR model, which presented an external prediction power of 80%. These findings can be quite useful to aid the designing of new anti-H_4 compounds with improved biological response.

Keywords Antiinflammatory agents · Asthma · H_4 receptor antagonists · Molecular modeling · QSAR · Rational drug design

J. P. S. Fernandes (✉) · K. F. M. Pasqualoto · E. I. Ferreira
Laboratório de Planejamento e Síntese de Quimioterápicos Potenciais Contra Endemias Tropicais,
Faculdade de Ciências Farmacêuticas, Universidade de São Paulo,
Av. Prof. Lineu Prestes 580,
05508-900 São Paulo, SP, Brazil
e-mail: jpsfer@usp.br

C. A. Brandt
Instituto Butantan,
Av. Vital Brazil 1500,
05503-900 São Paulo, SP, Brazil

Introduction

Histamine (Fig. 1) is one of the most important biogenic amines. It is widely distributed in the body, and the highest concentrations are found stored in the inflammatory cells, mainly mast cells [1]. The physiological functions of histamine are not fully known, but they are involved in several processes such as: hypersensitivity reactions [2]; production of acid by parietal cells [3]; control of the state of wakefulness and food intake in the central nervous system (CNS) [4]; and, play an important role in inflammation [2, 3]. During inflammation, histamine is released from preformed stores in mast cells and basophils [5].

The most recently discovered histamine receptor subtype, the H_4 receptor (H_4R), expressed on blood cells [6] has a distinct pharmacological profile and is considered a promising target for the treatment of various chronic inflammatory diseases (inflammatory bowel disease, asthma, and rheumatoid arthritis) [5, 7]. The H_4R seems to be related to the activation of immune cells in inflammatory reactions, mainly dendritic cells, eosinophils, mast cells, monocytes, basophils and T-lymphocytes [5, 6]. Studies have shown that the activation of the H_4R leads to the release of cytokines and other chemotactic factors, as interleukin 16 (IL-16), which is important in the asthma pathophysiology [5]. Considering that, selective H_4R antagonists could be clinically useful to treat asthma and other inflammatory diseases.

The three-dimensional structure of the human H_4R is not yet available, but its primary aminoacid sequence is already known. Then, there are homology models which were used to perform ligand-supported homology model of human H_4R applying virtual screening and demonstrated to be suitable models [7, 8]. Although the structure-based virtual screenings (SBVS) on H_4R provided some hit compounds [8], the homology models are only an approximation of the

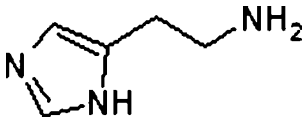


Fig. 1 Chemical structure of histamine

receptor conformations, and because of that, a ligand-based drug design (LBDD) strategy can be considered as a better alternative for planning new H_4 antagonists.

Among the H_4R antagonists reported in the literature [9–12], the indole- (**a**) and benzimidazole-piperazine-carboxamide (**b**) are compounds with good affinity as well as they are the subject of new patents [9] (see Fig. 2). Structure-activity relationships studies have shown that the best substitution in the piperazine nitrogen is a methyl group, and the benzimidazoles have a superior solubility in comparison to the indoles derivatives [10]. The alkylation of the indole nitrogen was shown to be unfavorable for the ligand-receptor interaction [11], and substitution in the aromatic ring conferred differences in the compounds' affinity [12]. Regarding these findings, quantitative structure-activity relationships (QSAR) studies were carried out to provide more information related to the substitutions mentioned and their correlation with the changes in biological response. The results may be helpful to understand the molecular mechanism of the actions of indole and benzimidazole derivatives as potential H_4R antagonists and to design more active compounds as promising leads acting on that subtype of histamine receptor.

Methods

Set of compounds and their biological activities

A set of 30 compounds, including indole- and benzimidazole-piperazine-carboxamide derivatives, were selected from reference 12 (Table 1). The binding affinity (K_i) to the human H_4R were evaluated from displacement of [^{3}H]-histamine from the recombinant human H_4R expressed in SK-N-MC cells [12]. The K_i values were expressed in negative logarithmic units, pK_i (-logK_i). The pK_i values are given in Table 1 and comprise the set of dependent variables in the QSAR analysis. The range in activity found for the compounds in Table 1 were between 3.37 and 5.40. The training set was composed of 25 compounds whereas five compounds were selected as an external validation set (test set), using hierarchical cluster analysis (HCA) [13], and are marked with an asterisk (*) in Table 1.

Molecular modeling and QSAR models

The three-dimensional models were built up in their neutral forms using the HyperChem 7.51 software [14]. The crystallized structures of the ligands 5-chloro-N-[2-(cyclopentyl-(2-hydroxyethyl)amino)-2-hydroxy-ethyl]-1H-indole-2-carboxamide and 1*H*-benzimidazole-2-carboxilic acid were retrieved from PDB (entry codes 1XOI [15] and 1FQ4 [16]) and used as starting geometries to construct the indole and benzimidazole derivatives, respectively. Each 3D-model was geometry-optimized using HyperChem 7.51 MM+force field without any restriction, followed by the AM1 semiempirical method [17] and ab initio method (Hartree-Fock/6-31G*), implemented at Gaussian 03W [18]. Partial atomic charges (ESP) were calculated based on CHELPG algorithm and the Hartree-Fock/6-31G* theory level (Gaussian 03W) [18]. The resulting 3D structures were used to calculate the independent variables considered in this QSAR study. The programs employed to generate the independent variables were: HyperChem 7.51 [14], Marvin Beans [19], and Gaussian 03W [18]. A total of 63 descriptors of distinct nature (structural, lipophilic, electronic, topologic, steric, thermodynamic) were calculated, such as: interatomic distance, bond and torsional angles, calculated *n*-octanol/water partition coefficient (Clog*P*) [20, 21], calculated *n*-octanol/water distribution coefficient (log*D*)

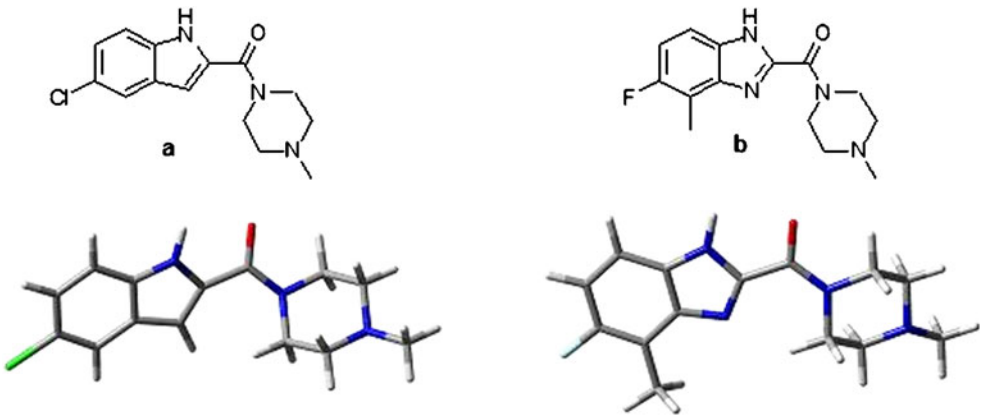
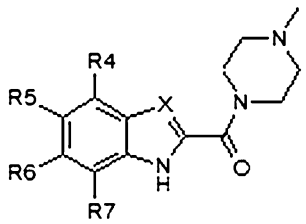


Fig. 2 Histamine H_4R antagonists

Table 1 Set of compounds selected and their binding affinities (*Ki*) [12]



Compound	R4	R5	R6	R7	X	K_i (nM)[a]	pK_i
1	H	H	H	H	CH	17±1	4.77
2	Br	H	H	H	CH	32±2	4.50
3	H	H	Br	H	CH	147±23	3.83
4	H	Br	H	H	CH	8±1	5.10
5 (**a**)	H	Cl	H	H	CH	4±1	5.40
6	H	F	H	H	CH	15±1	4.82
7*	H	CH3	H	H	CH	46±5	4.34
8	H	CF3	H	H	CH	412±94	3.39
9*	H	OH	H	H	CH	23±2	4.64
10	H	NH2	H	H	CH	15±1	4.82
11	H	H	H	Cl	CH	19±1	4.72
12*	H	H	H	Br	CH	61±5	4.21
13	H	H	H	CH3	CH	7±1	5.16
14*	H	H	H	NH2	CH	8±1	5.10
15	Cl	Cl	H	H	CH	5±3	5.30
16	CH3	F	H	H	CH	27±1	4.57
17	H	F	H	F	CH	14±1	4.85
18*	H	Cl	H	Cl	CH	11±1	4.96
19	H	CH3	H	CH3	CH	31±1	4.51
20	H	CH3	H	Cl	CH	33±3	4.48
21	H	H	H	H	N	35±9	4.46
22	CH3	H	H	H	N	30±3	4.52
23	H	NH2	H	H	N	244±44	3.61
24	H	F	H	H	N	19±7	4.72
25	H	Cl	H	H	N	26±8	4.59
26	H	CF3	H	H	N	427±80	3.37
27 (**b**)	CH3	F	H	H	N	7±3	5.16
28	F	F	H	H	N	49±21	4.31
29	Cl	H	CH3	H	N	23±6	4.64
30	Cl	H	Cl	H	N	136±74	3.87

[a]*Ki* is the geometric mean±SEM obtained from three or more independent determinations

*test set compounds

[20–22] in pH 1.5, 5.0, 7.4 and 8.0, molar refractivity (MR) [20], polarizability (α) [23], van der Waals volume (V_{vdW}) [24] and molecular surface (S_{vdW}) [24], heat of formation (HF) [17], ionization constant (p*K*a) of nitrogen atoms [25], isoelectric point (pI) [25], index of Platt [26], Randič [27], Balaban [28] and Wiener [29], steric effect of substituents [30], number of hydrogen-bond donor and acceptor groups, energy of molecular orbitals (highest occupied molecular orbital, HOMO, and lowest unoccupied molecular orbital, LUMO) [18], dipole moment (total and x, y, z) obtained by the Hartree-Fock method [18], and partial atomic charges (ESP) [31], for example. A preliminary systematic search of the most significant independent variables was carried out based on their distribution or variability against the biological data (visual inspection) using scatter plots.

The selected descriptors (26) were used to build QSAR models employing partial least squares (PLS) regression [32] and a genetic function approximation (GFA) algorithm [33], which are fitting functions available in the WOLF 5.5 program [34]. GFA can build models using not only linear polynomials but also higher-order polynomials, splines, and other nonlinear functions [35]. In this study, the top eight QSAR models were selected by the WOLF 5.5 program [34]. Linear and second-degree polynomials were the functions tested. The GFA optimizations started with 500 randomly generated models and the mutation probability over the crossover optimization cycle set as 10%. A number of genetic operations or crossovers of 50 000 to 100 000 were tested. The models are scored using Friedman's lack-of-fit (LOF) measure, which is a penalized least-squares measure. The lower the LOF value, thebetter the adjustment of the model obtained. The smoothing factor or parameter (d), which is part of the LOF definition, is the only parameter adjustable by the user [34], and it alters the balance between the number of independent variables (descriptors) in the models and the reduction in LSE measure, besides it controls *overfitting*. The default value of smoothing factor is 1.0. In this study, smoothing factor values of 1.0 to 0.1 were tested for generating the QSAR models.

Statistical measures of significance including the regression coefficient (r^2), leave-one-out (LOO) cross-validation coefficient (q^2), least squares error (LSE), and LOF measure, were calculated to test the robustness of the models. The cross-correlation descriptor matrix was examined to eliminate trial QSARs in which pairs of descriptors have cross-correlation coefficients greater than 0.5. Also, the cross-correlation matrix of residuals of fit between pairs of models was computed to determine if the top eight QSAR models provide common or distinct information. Pairs of models with highly correlated residuals of fit ($R \approx 1$) are judged to be nearly the same model, while pairs of models with poorly correlated residuals ($R < 0.5$) are distinct from one another.

In this study, the ligands of the training set whose differences in observed and predicted activities exceeded two standard deviation (SD) from the mean of a model were considered as outliers.

Approaches to QSAR model validation, including **y**-randomization or **y**-scrambling, robust internal validation strategies such as multiple leave-*N*-out (LNO) cross-validations, and external validation were applied in this study whereas only validated QSAR models can offer a meaningful mechanistic interpretation, especially in the context of design or discovery of novel chemical agents with desired properties [36–40].

The LNO procedure was carried out in triplicates and repeated till ten compounds were left out from the training set. Ideal expectation is high average q^2. In other words, if a QSAR model has a high average q^2 in LNO validation, it can be reasonably concluded that the obtained model is robust [38].

The **y**-randomization test is a widely used technique to ensure the robustness of a QSAR/QSPR model [36]. In this test, the dependent-variable vector, **y**-vector, is randomly shuffled and a new QSAR model is developed using the original independent-variable matrix. The process is repeated several times [37]. In the present study, that procedure was performed in triplicates and repeated ten times. It is expected that the resulting QSAR models should generally have low r^2 and low LOO q^2 values. Otherwise, if all QSAR models obtained in the **y**-randomization test have relatively high r^2 and LOO q^2, it implies that an acceptable QSAR model cannot be obtained for the given data set by the current modeling method, probably due to a chance correlation or structural redundancy of the training set.

It is recommended [38] that the external test set must contain at least five compounds, representing the whole range of both descriptors and activity of compounds included into the training set. In this study, the test set contains just five compounds considering the total number of compounds investigated (N=30). These five compounds were not included in the building of the QSAR models, but they were used to validate the best QSAR model constructed from the training set and to evaluate its prediction capacity. The predicted activity value (pK_i calc) of each ligand in the test set was calculated using the equation of the best model by the substitution of the descriptors' values, which were selected as the most relevant to the biological activity.

Results

The top eight models (N=25) selected by the WOLF 5.5 program, using a smoothing factor of 0.3; 10 % probability of mutation for each crossover; and 100 000 genetic operations or crossovers, presented two functions type (linear and quadratic or simply second-degree polynomial terms). The r^2 and q^2 values ranged from 0.71 to 0.76, and from 0.52 to 0.60, respectively, whereas the LOF and LSE values varied from 0.13 to 0.14, and from 0.07 to 0.08, respectively. The number of descriptors changed from 4 to 5 and all generated models did not present any outliers.

The cross-correlation matrix of residuals of fit between pairs of models was computed and they were highly correlated to one another (R=0.74 to 1.00) (Table 2). Thus, the models were judged to be nearly the same model meaning that there is a single unique model. Model 1

Table 2 Correlation matrix of residuals of fit between pairs of models found for the top eight models

Model	1	2	3	4	5	6	7	8
1	**1.00**	0.98	0.79	0.78	0.78	0.78	0.78	0.78
2		**1.00**	0.75	0.78	0.78	0.77	0.74	0.74
3			**1.00**	0.97	0.96	0.97	0.99	0.99
4				**1.00**	1.00	0.99	0.98	0.96
5					**1.00**	0.99	0.97	0.95
6						**1.00**	0.97	0.97
7							**1.00**	0.98
8								**1.00**

(Eq. 1) was selected as the best model because it did not have any outliers (see Table 3 and Fig. 3).

Model 1:

$$\begin{aligned} \mathrm{p}K_\mathrm{i} &= 5.4572 - 0.0002(\mathrm{HF} - 12.963)^2 \\ &\quad - 0.0601(\log D_{1.5} + 5.758)^2 + 0.0191(\mu_\mathrm{x} + 3.639)^2 \\ &\quad - 0.3916(\mathrm{pI} - 8.811)^2 - 1.3107\mathrm{qC5} \\ \mathrm{N} &= 25; r^2 = 0.76; q^2 = 0.60; \mathrm{LOF} = 0.13; \mathrm{LSE} = 0.07; \\ &\quad \text{outliers} = 0 \end{aligned} \tag{1}$$

Where: HF is the heat or enthalpy of formation (thermodynamic descriptor); $\log D_{1.5}$ is the calculated *n*-octanol/water distribution coefficient in pH 1.5; μ_x is the dipole moment at x coordinate; $\mathrm{pI_l}$ is the isoeletric point; and qC5 is the ESP partial atomic charge of the carbon bound to the R5 substituent.

The predicted and observed activities found for the training set (N=25) considering model 1 are plotted in Fig. 3. As already mentioned, all ligands have their activity well predicted by model 1 (no outliers). The obtained model was validated using the LNO and **y**-randomization tests. Table 4 presents the obtained q^2_{LNO} values when up to ten compounds from the training set were left out (*n* is the number of objects excluded in the internal validation process which varied from 2 to 10). Table 5 shows the resulting LOO q^2 and r^2 values when the **y**-vector was randomly shuffled ten times and new ten QSAR models were developed for the same data set, using the original independent-variable matrix and the same conditions employed in the building of the selected best QSAR model (model 1; Table 6).

Table 3 Top eight models and their statistical measures, number of descriptors and number of outliers

Model	r^2	q^2	LSE	LOF	N° descriptors	Outliers
1	0.76	0.60	0.07	0.13	5	0
2	0.76	0.59	0.07	0.13	5	0
3	0.72	0.56	0.08	0.13	4	0
4	0.71	0.54	0.08	0.13	4	0
5	0.71	0.54	0.09	0.14	4	0
6	0.71	0.54	0.09	0.14	4	0
7	0.71	0.52	0.09	0.14	4	0
8	0.71	0.55	0.09	0.14	4	0

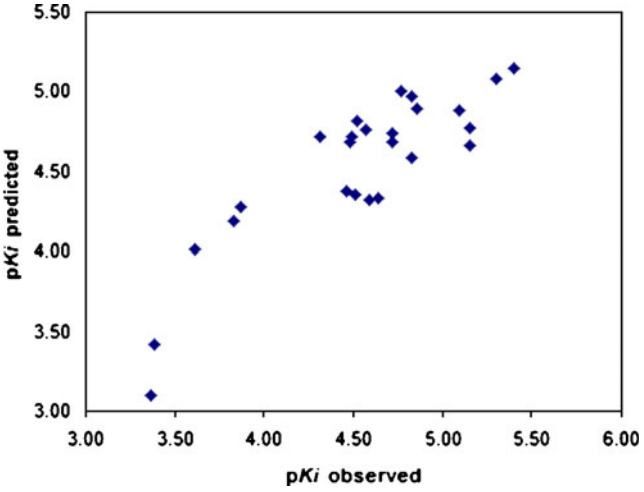


Fig. 3 Predicted or calculated ($\mathrm{p}K_i$ predicted) and observed or experimental ($\mathrm{p}K_i$ observed) activity values found for the training set (N=25), considering model 1

The external validation of the model was carried out using five compounds (* marked in Table 1) that were not used to build up the model. The $\mathrm{p}K_\mathrm{i}$ of four compounds were well predicted by the model, providing a predictability of 80% to the model.

Discussion

The compounds reported by Venable et al. [12] have a well-defined aromatic nucleus (indole or benzimidazole ring)

Table 4 Internal validation LNO procedure results for the selected best model (model 1; N=25)

n	$q^2_{LNO(1)}$	$q^2_{LNO(2)}$	$q^2_{LNO(3)}$
2	0.50	0.57	0.47
3	0.53	0.53	0.46
4	0.48	0.62	0.56
5	0.55	0.43	0.63
6	0.56	0.45	0.48
7	0.56	0.66	0.59
8	0.48	0.45	0.51
9	0.49	0.57	0.55
10	0.48	0.30	0.51

Table 5 Values of LOO q^2 and r^2 found for the ten QSAR models generated with the same data set (N=25) employing the **y**-randomization technique

Randomization	q^2_{LOO}	r^2
1	-0.17	0.19
2	-0.28	0.14
3	-0.83	0.06
4	-0.38	0.28
5	-0.33	0.08
6	-0.38	0.08
7	-0.53	0.06
8	-0.12	0.24
9	-0.39	0.11
10	-0.39	0.19

substituted in various positions as well as a piperazine-carboxamide portion as a side chain. In a previous work [10], it was shown that the best piperazine nitrogen substitution is a methyl group, and a longer alkyl chain is detrimental to the biological activity of these compounds. Also, the carbonyl group seems to be essential to the activity, but substitution in vicinal position of amine piperazine ring is not favorable to the receptor affinity [10]. Substitutions on aromatic nucleus afforded compounds that change their receptor affinity, and halogen R5-substituents increase potency, mainly associated to any other substitution. Otherwise, substitution in R7 position provides an increase of potency, but this effect is less intense [12]. Even though these findings were relevant to the understanding of H_4R antagonists' interactions, there are not quantitative structure-activity relationships established yet. In this study, a QSAR model was obtained, which estimates the contributions of the substitutions mentioned.

Model 1 has only one linear function (qC5) correlated to the pK_i values. Regarding the quadratic or second-degree polynomial and linear functions, the higher the value of the sum or the difference of the terms between parentheses, the more significant (favorable or unfavorable) the contribution to the biological activity. A brief discussion considering the descriptors in model 1 is made below.

HF is an important independent variable, which is defined as the enthalpy gained in forming a molecule from its constituent atoms. It represents the chemical stability and reactivity of a compound [41]. Here, this descriptor is probably related to the conformational arrangement and thermodynamic stability. In model 1, HF is a quadratic term, which indicates that there is possibly an optimal value between 40 and 50 kcal mol^{-1} for a maximum receptor affinity.

The log*D* value is correlated directly to the molecule ionization, and thereby to the diffusion through biological membranes [21]. In model 1, the calculated log*D* value in acidic pH (1.5) is present. Anti-H4 compounds act as anti-inflamatory agents, and the environment under inflammatory conditions is always acidic. Then, the log*D* property in acidic pH could be related to the compounds' diffusion in the inflammatory tissue and/or cells. The log$D_{1.5}$ is also a quadratic term, and to reach the best biological activity its value should be in the range -3.0 to -2.0. However, it is possible that to have a good biological activity, the compounds must have a limited penetration in the tissue and act on the cell surface, presenting the piperazine amine group positively charged in acidic pH, whereas for interacting with the receptor might be in the neutral form. The balance between these partition-electronic features must be observed.

Dipole moment indicates the polarity of a molecule, which is related to the electronegativity of the molecule's atoms and depends on the charge distribution. Dipole moment in the x-axis or coordinate (μ_x) indicates the tendency of charge distribution at the molecule's x-axis direction. Optimum value of μ_x to achieve maximum activity is about 2.0 Debye.

The isoelectric point (pI) corresponds to the defined pH in which a molecule has no net electric charge [42]. It is correlated to the charge of certain atoms or groups and can be directly associated to the receptor affinity and binding. The pI of compounds with high pK_i values is approximately 9.0. The investigated compounds have three nitrogen atoms (aromatic, amide and amine). The aromatic and amine nitrogen atoms are under influence of the ring substituents due to the resonance effect, changing the ionization degree (pK_a) of those nitrogen atoms and, consequently, the pI of the respective compound.

The role of the X group in aromatic nucleus does not seem to be determinant to the binding affinity. The compounds investigated present CH or N as X group, which are considered as bioisosteres. Regarding Table 1, in some cases, indole (X=CH) and benzimidazole (X=N) series are very similar in their binding affinities, but they also have some remarkable differences. For example, compounds **8** and **26** have almost the same pK_i value (3.39 and 3.37, respectively), but compounds **10** and **23** are completely different regarding its activity (pK_i 4.82 and 3.61). Therefore, X group does not directly affect the pK_i, but it can affect the aromatic nucleus characteristics which

Table 6 External validation (test set; N=5) of model 1

Compound	pK_i observed	pK_i predicted	Residuals
7	4.34	4.35	-0.01
9	4.64	4.52	0.12
12	4.21	4.41	-0.20
14	5.10	5.34	-0.24
18	4.96	5.37	-0.41

led to changes in molecular properties. Considering model 1, the indole derivatives have higher pI values (9.53 found for compound **8**), while benzimidazole derivatives present pI values of approximately 7 (7.70 found for compound **26**). As pI has influence on biological activity, then indole derivatives seem to be more active than benzimidazoles.

As already mentioned, model 1 shows that the ESP partial charge [31] in the carbon bound to the R5 substituent (qC5) is the only descriptor that is linearly correlated to the biological activity, having an inverse influence or unfavorable contribution. Then, more positive values found for the ESP partial charge distribution on that moiety of the molecules leads to lower pK_i values. The most active compounds must present higher electronic density on that region (C5), as can be seen in the color maps of electrostatic potential (MEPs) for compounds **5** (the most active indole derivative) and **8** (less active), respectively (Fig. 4). Considering these findings, the receptor probably has in the binding site an amino acid residue positively charged complementary to the C5 region. Thus, the substitution in R5 plays an important role in the biological activity of this set of compounds, as already observed by Venable et al. [12].

It can be observed that substitutions in R5 attribute relevant characteristics to the compounds particularly when the substituent group is halogen without any other substitutions (compounds **4**, **5**, **6**, **24** and **25**). Other kinds of substituent groups in the same position even having similar electronic features appear to be detrimental to the binding affinity (compounds **8** and **26**). However qC5 seems to be determinant to the activity, there is probably another effect (maybe the volume) that changes the affinity and is also related to the R5 substituent.

The internal validation LMO procedure (internal prediction power) and **y**-randomization technique (check for chance of correlations) were carried out to verify the robustness of model. Good QSAR models must have q^2_{LNO} values closest to the LOO q^2 value of the selected best model. Furthermore, the q^2_{LNO} values must be closest to the average, $< q^2_{\mathrm{LNO}} >$, and the oscillation range accepted is 0.1. All q^2_{LNO} values (Table 4) are closest to the LOO q^2 value (0.60). Also, the q^2_{LNO} values oscillated from 0.00 to 0.08 in comparison to the average value ($< q^2_{\mathrm{LNO}} >= 0.52$), indicating a good internal predictability.

If in the application of a simple **y**-randomization test many models present acceptable values of LOO q^2 for the same data set, there is no real structure-activity relationship and, consequently, any interpretation of those QSAR models built in this fashion is spurious. All QSAR models obtained in the **y**-randomization test have low LOO q^2 and r^2 values (Table 5), it implies that an acceptable QSAR model can be obtained for the given data set by the current modeling method, which is model 1.

To ascertain the external predictive power of model 1 based on screening a test set of compounds (* marked in Table 1), the pK_i value of each of the test set ligands was calculated using Eq. 1, as described in External validation. Four of the five ligands of the test set had residuals whose absolute values were less than or equal to the standard deviation value from the mean of the model (SD=0.2) (Table 2). This finding indicates that model 1 has a capacity of prediction of 80 %, which is a good external predictive power.

Compound **18** was not well predicted by model 1 probably due to its electronic and partition properties. The μ_x and log$D_{1.5}$ values (3.85 Debyes and -1.93, respectively) were quite distinct in comparison to those found for the compounds **7**, **9**, **12**, and **14**. Electronic and partition properties are involved in the ligand-receptor binding process and, consequently, are crucial to trigger the desirable biological response.

Conclusions

Considering the conditions adopted in this study, the QSAR model is robust (LNO and **y**-randomization validation techniques) and has a good internal and external predictability. The electronic and partition features are among the molecular properties that might be considered for designing new potential histaminergic H_4R. The electronic effect of the substituents in arene moiety, mainly in R5 position (particularly halogen), plays an important role in the receptor affinity. Also, the X position shows important influence in binding affinity, and affects the pI values.

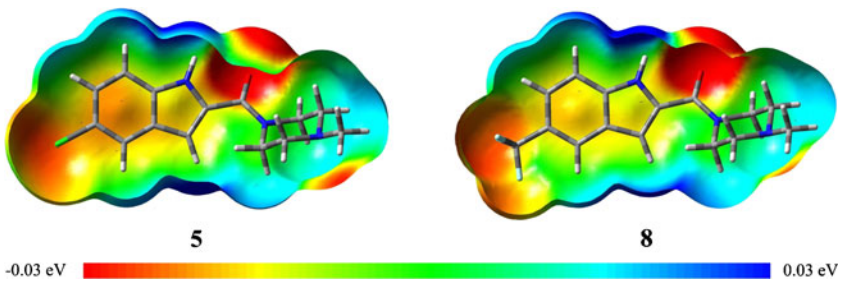


Fig. 4 Maps of electrostatic potential (MEPs) calculated for compounds **5** and **8**, respectively (Gaussian 03W). The electronic density around C5 is higher in **5** than in **8**

These findings are quite useful to the rational drug design of novel potential H_4R ligands with antagonistic properties.

References

1. Nelson WL (2002) Antihistamines and related antiallergic and antiulcer agents. In: Williams DA, Lemke TL (eds) Foye's Principles of Medicinal Chemistry, 5th edn. Lippincott Williams & Wilkins, Philadelphia, pp 794–818
2. Parsons ME, Ganellin CR (2006) Histamine and its receptors. Br J Pharmacol 147:S127–S135
3. Moessner J, Caca K (2005) Developments in the inhibition of gastric acid secretion. Eur J Clin Invest 35:469–475
4. Haas H, Panula P (2003) The role of histamine and the tuberomamillary nucleus in the nervous system. Nat Rev Neurosci 4:121–130
5. Thurmond RL, Gelfand EW, Dunford PJ (2008) The role of histamine H_1 and H_4 receptors in allergic inflammation: The search for new antihistamines. Nat Rev Drug Discov 7:41–53
6. Oda T, Morikawa N, Saito Y et al. (2000) Molecular cloning and characterization of a novel type of histamine receptor preferentially expressed in leukocytes. J Biol Chem 275:36781–36786
7. Jórjárt B, Kiss R, Viskolcz B et al. (2008) Activation mechanism of the human histamine H4 receptor – an explicit membrane molecular dynamics simulation study. J Chem Inf Model 48:1199–1210
8. Kiss R, Noszál B, Rácz A et al. (2008) Binding mode analysis and enrichment studies on homology models of the human histamine H4 receptor. Eur J Med Chem 43:1059–1070
9. Carruthers NI, Dvorak CA, Edwards JP et al. (2004) Preparation of nitrogen containing heterocyclic compounds as compounds useful for in the treatment of histamine H_4 mediated diseases. PCT Int Appl 70pp, WO 2004022060
10. Jablonowski JA, Grice CA, Chai W et al. (2003) The first potent and selective non-imidazole human histamine H_4 receptor antagonists. J Med Chem 46:3957–3960
11. Terzioglu N, Van Rijn RM, Bakker RA et al. (2004) Synthesis and structure-activity relationships of indole and benzimidazole piperazines as histamine H_4 receptor antagonists. Bioorg Med Chem Lett 14:5251–5256
12. Venable JD, Cai H, Chai W et al. (2005) Preparation and biological evaluation of indole, benzimidazole and thienopyrrole piperazine carboxamides: potent human histamine H_4 antagonists. J Med Chem 48:8289–8298
13. Brown RD, Martin YC (1996) Use of structure−activity data to compare structure-based clustering methods and descriptors for use in compound selection. J Chem Inf Comput Sci 36:572–584
14. HyperChem Program Release 7.51 for Windows (2002) Hypercube Inc, Gainesville, FL
15. Wright SW, Rath VL, Genereux PE et al. (2005) 5-Chloroindoloyl glycine amide inhibitors of glycogen phosphorylase: synthesis, in vitro, in vivo, and X-ray crystallographic characterization. Bioorg Med Chem Lett 15:459–465
16. Cronin NB, Badasso MO, Tickle IJ et al. (2000) X-ray structures of five renin inhibitors bound to saccharopepsin: exploration of active-site specificity. J Mol Biol 303:745–760
17. Dewar MJS, Zoebisch EG, Healy EF et al. (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. J Am Chem Soc 107:3902–3909
18. Gaussian 03W, version 6.0, revision A.1 (2003) Gaussian Inc, Pittsburgh, PA
19. MarvinBeans version 5.1.5 (2009) ChemAxon, Ltd, Budapest, HUN
20. Viswanadhan VN, Ghose AK, Revankar GR et al. (1989) 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 29:163–172
21. Klopman G, Li J, Wang S (1994) Computer automated logP calculations based on an extended group contribution approach. J Chem Inf Comput Sci 34:752–781
22. Csizmadia F, Tsantili-Kakoulidou A, Panderi I et al. (1997) Prediction of distribution coefficient from structure. 1. Estimation method. J Pharm Sci 86:865–871
23. Miller KJ (1990) Additivity methods in molecular polarizability. J Am Chem Soc 112:8533–8542
24. Bodor N, Gabanyi Z, Wong CK (1989) A new method for the estimation of partition coefficient. J Am Chem Soc 111:3783–3786
25. Szegezdi J, Csizmadia F (2007) A method for calculating the pKa values of small and large molecules. ACS Spring Meeting, March 25-29th
26. Platt JR (1947) Influence od neighbor bonds on additive bond properties in paraffins. J Chem Phys 15:419–420
27. Randič M (1975) Characterization of molecular branching. J Am Chem Soc 97:6609–6615
28. Balaban AT (1982) Distance connectivity index. Chem Phys Lett 89:399–404
29. Wiener H (1947) Structural determination of paraffin boiling points. J Am Chem Soc 69:17–20
30. Cao C, Liu L (2004) Topological steric effect index and its application. J Chem Inf Comput Sci 44:678–687
31. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials the need for high sampling density in formamide conformational analysis. J Comput Chem 11:361–373
32. Glen WG, Dunn WJ, Scott DR (1989) Principal components analysis and partial least-squares regression. Tetrahedron Comput Methodol 2:349–354
33. Rogers D, Hopfinger AJ (1994) Applications of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J Chem Inf Comput Sci 34:854–866
34. Rogers D (1994) WOLF Reference Manual Version 5.5. The Chem21 Group Inc, Chicago, IL
35. Shi LM, Fan Y, Myers TG et al. (1998) Mining the NCI anticancer drug discovery databases: genetic function approximation for the QSAR study of anticancer ellipticine analogues. J Chem Inf Comput Sci 38:189–199
36. Wold S, Eriksson L (1995) Statistical validation of QSAR results. In: van de Waterbeemd H (ed) Chemometric methods in molecular design. VCH, Weinheim, pp 309–318
37. Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 22:69–77
38. Golbraikh A, Tropsha A (2002) Beware of q^2! J Mol Graph Mod 20:269–276
39. Geisser S (1975) The predictive sample reuse method with application. J Am Stat Assoc 70:320–328
40. Kubinyi H (2003) QSAR in drug design. In: Gasteiger J (ed) Handbook of chemoinformatics. Wiley-VCH, Weinheim, pp 1532–1554
41. Karki RG, Kulkarni VM (2001) Three-dimensional quantitative structure–activity relationship (3D-QSAR) of 3-aryloxazolidin-2-one antibacterials. Bioorg Med Chem 9:3153–3160
42. Pripp AH (2006) Quantitative structure-activity relationship of prolyl oligopeptidase inhibitory peptides derived from β-casein using simple amino acid descriptors. J Agric Food Chem 54:224–228