Original Article

DOI: 10.5582/ddt.2012.v6.2.69

In silico ligand based design of indolylpiperidinyl derivatives as novel histamine H_1 receptor antagonists

Sarvesh Paliwal*, Supriya Singh, Mahima Pal

Department of Pharmacy, Banasthali University, Banasthali, Rajasthan, India.

ABSTRACT: Histamine H_1 receptor antagonists play a vital role in the first line treatment of a broad range of allergic diseases. Frequent dosing of the antagonist results in side effects like sedation and cardiovascular toxicity. The present study highlights the important structural requirement and mechanistic interpretation of novel indolylpiperidinyl derivatives as H_1 receptor antagonists so as to facilitate the design of newer antihistaminics with increased duration of action and comparatively reduced side effects. The significance of the developed quantitative structure-activity relationship (QSAR) models were evaluated on the basis of statistical values of square of correlation coefficient (r^2); (multiple linear regression (MLR), 0.86; and partial least squares (PLS), 0.85). The predictive ability of the resulting QSAR models was evaluated with cross-validated correlation coefficient (r^2cv) values (MLR, 0.82; PLS, 0.82) generated for the training set and r^2 values (MLR, 0.763; PLS, 0.855) derived for test set. The final models comprised of multidimensional steric (verloop length, verloop B_3), electronic (total dipole moment) and steric (KAlpha1 index) descriptors. The study indicates that antihistaminic activity is largely explained by steric and electronic parameters. In line with parameters entered in the model some indolylpiperidines derivatives were designed with good antihistaminic properties and pharmacokinetic profiles.

Keywords: Quantitative structure-activity relationship (QSAR), tools for structure activity relationship (TSAR), multiple linear regression (MLR), partial least square (PLS)

**Address correspondence to:*
Dr. Sarvesh Paliwal, Department of Pharmacy, Banasthali University, Banasthali, Rajasthan-304022, India.
E-mail: paliwalsarvesh@yahoo.com

1. Introduction

Among a wide range of mediators which are involved in the pathophysiology of allergic diseases, histamine remains the principal one and plays a fundamental role in the genesis of these diseases, particularly rhinitis and urticaria (*1*). Chronically, histamine also affects inflammatory cells and causes cellular activation (mast cells, basophils, and eosinophils) and release of proinflammatory mediators *e.g.*, leukotrienes and cytokines and results in an increase in the expression of class II human histocompatibility molecules (HLA) and vascular endothelial adhesion molecules (*2-4*).

Antihistaminics prevent symptoms associated with histamine release such as rhinorrhea, nasal and conjunctival itching, and lacrimation, although they do not control symptoms of nasal congestion (*5*). Depending on their action on the central nervous system (CNS), they are classified as "classical", or first-generation, and "non-classical", or second-generation.

In general, first-generation H_1 antihistamines (for example dexchlorpheniramine and hydroxyzine) cross the blood brain barrier (BBB), bind with ease to the cerebral H_1 receptors and also possess anticholinergic properties (*3,6*). Their principal side effect is sedation, dry mouth, and lack of receptor specificity at therapeutic doses. This has led to the development of a second-generation of H_1-antagonists (*7*).

Second-generation H_1-antagonists exhibit high potency, long-lasting effects and minimal adverse effects. They are unlikely to cross the BBB and rarely cause sedation. Some findings suggest that second generation H_1 receptor antagonists produce CNS depressant effects due to their liability to penetrate into the CNS (*8*). Also, adverse cardiac effects have been reported with second-generation H_1 antihistamines namely astemizole and terfenadine (*9,10*).

In line to the above discussion we felt that there is a need to reevaluate the binding requirements of antihistaminics by employing a computational approach. One of the most promising techniques to get insight into the structural requirements is quantitative structure-activity relationship (QSAR), which came into existence in the 1960's for the first time. QSAR is

a mathematical relationship linking chemical structure and pharmacological activity in a quantitative manner for a series of compounds representing hydrophobic, electronic, steric and other effects using multiple regression correlation methodology. QSAR increases the probability of success and reduces the time and cost involved in the drug discovery process (*11*). The aim of this investigation was to develop QSAR models using the multivariate analysis derived from global descriptors with the purpose to elucidate the contribution of substitutions of antihistamine antagonists and the forces involved in drug receptor interaction. The study of *de novo* contributions of the groups helped us to design and predict activity of some newer analogs.

2. Materials and Methods

A series of 53 indolylpiperidinyl benzoic acid derivatives was obtained from the literature for the present QSAR studies (*12*). For the current QSAR investigation 51 compounds were used and the remaining two compounds (**8** and **39**) were excluded in view of uncertain activity data since such data cannot be used for model development using tools for structure activity relationship (TSAR) software. Experimentally determined biological activity of compounds in the series was reported as a half minimum inhibitory concentration (IC_{50}) value. Since biological activities are generally skewed and are measures of the free energy of binding, the reported inhibitory constant values were converted into a corresponding negative log value.

The negative log IC_{50} values were used as the dependent variable and various quantifying parameters as the independent variable. All computational studies were performed using TSAR (version 3.3) software (*13*).

2.1. *Drawing and optimization of structures*

The organic structure of 51 indolylpiperidinyl benzoic acid derivatives selected for QSAR studies, were sketched using ChemDraw Ultra 10.0. All sketched chemical structures were imported to a TSAR 3.3 spread sheet *via .mol* files. Structure entry and substitutions defining is an important stage in QSAR methodology. The substituents of each chemical structure were defined into six substituents namely R_1, R_2, R_3, R_4, R_5, and R_6. All the substituents were numbered according to their position in molecules, and each molecule had a defined number of substituents attached to the nucleus by a single bond. The substituent pattern opted for is given in Supplemental Table S1 (*http://www.ddtjournal.com/getabstract.php?id=538*).

TSAR has a built in program CORINA, which was used to convert all molecular structures and substitutions to 3D structures. The 3D structure concept was developed by Hiller (*14*). The three dimensional structure of a molecule is closely related to a large variety of chemical, physical and biological properties. The CORINA automatically generates 3D atomic co-ordinates from the constitution of a molecule as expressed by a connection table or linear string (*15*).

The Cosmic module was used to optimize the structure of compounds. Cosmic calculates molecular energies by summing bond length, bond angle, torsion angle, van der waals, and coulombic terms for all appropriate sets of atoms. These calculations involve the valence electrons of the atoms of the molecule. These lead to further development of semi-empirical molecular orbital (MO) calculations (*16*). The calculations were terminated when the energy difference or the energy gradient were smaller than 1×10^{-5} and 1×10^{-10} kcal/mol, respectively (*17*).

The data set consisting of compounds with molecular structures and their biological activities were divided into a training and test set. Twenty percent of the compounds were selected with a maximum dissimilarity algorithm and assigned to the test set; with the remaining 80% assigned to the training set (*18*). The training set comprised of 41 molecules was used for QSAR model development and the test set of 10 molecules was used for model validation.

2.2. *Calculation of molecular descriptors*

Descriptors can be defined as numerical quantities that have been generated to represent the molecular configuration. It also contributes toward better understanding of structural, steric, electronic and multidimensional properties responsible for activity. The aim of calculating molecular descriptors is to provide all useful information about all chemical structures and respective substituents to build a good and predictive QSAR model. The nature of descriptors used and the extent to which they encode structural features related to biological activity is a crucial part of a QSAR study. The success of the QSAR analysis is significantly dependent on the accurate definition and appropriate use of molecular descriptors. Molecular descriptors are terms that characterize a specific aspect of a molecule.

TSAR can calculate up to 500 descriptors (topological, geometrical and electrostatic) derived from whole structures as well as substitution of compounds under consideration. Since a large pool of descriptors was calculated, there is a significant requirement for data reduction to eliminate chance correlation. A correlation matrix was used to reduce the number of descriptors and to identify the best subset of descriptors with minimum intercorrelation. The goal was to remove redundancy among the descriptors and to detect chance effects during model development.

A correlation coefficient describes the degree of linear correlation between two variables. Pair wise correlation coefficients were calculated for all pairs of descriptors. If an intercorrelation coefficient > 0.5 was detected, the descriptor with high correlation with biological activity was kept and others were discarded. In the next

phase data reduction was performed on the remaining descriptors on the basis of *t*-values using a backward elimination technique. Stepwise regressions were developed and the descriptors having lower *t*-values were discarded from the data set (*19*).

The reduction process was repeated a number of times and finally four independent descriptors: Verloop L (Substituent 5), Verloop B_3 (Substituent 5), total dipole moment (whole molecule), and Kalpha1 index (whole molecule), which were highly correlated with biological activity and exhibited minimum intercorrelation with each other, were retrieved. The statistical values for the most relevant descriptors and their corresponding *t*-values used in QSAR development are shown in Table 1 and the correlation matrix for the four descriptors is depicted in Table 2.

2.3. *Multivariate statistical analysis*

Multivariate statistical analysis is a set of statistical tools used for modeling a set of dependent variables, such as biological activity, and molecular descriptors as the independent variables (*20*). The relationship between the structural parameters (global descriptors) and the biological activities was quantified by multiple linear regressions (MLR) and partial least squares (PLS). Values for F-to-enter and F-to-leave were set to 4. Outliers in QSAR can be very important and interesting, especially when the observed biological activity is higher than the predicted one by the developed model (higher residual value). Outliers may be present due to inappropriate calculation of the parameter values used. The mathematical model may not be appropriate. There may be a lack of certain descriptors or parameters to describe QSAR for entire compounds. A different mechanism mode may even be a reason (*21*). Four outliers (**6**, **7**, **10**, and **23**) were detected with the help of a regression line equation. There could be various reasons for the observed outliers. The acceptable and robust QSAR model was selected on the basis of various statistical significant parameters like correlation coefficient (r), square of correlation coefficient (r^2), cross-validated correlation coefficient (r^2cv), Fisher ratio (F), and standard deviation (SD). The resulting models were validated by a leave-one out cross-validation procedure and a test set prediction to check their predictability and robustness.

In addition to MLR, PLS has also been performed to check the robustness of the developed MLR model. PLS was developed in the 1960's by Herman Wold as an econometric technique. PLS is a robust multivariate regression method suitable for overcoming general problems in MLR related to over-abundant descriptors and therefore comparable predictive models can be obtained by the PLS method (*22*). The results of PLS were evaluated on the basis of r^2, r^2cv, and statistical significance.

3. Results and Discussion

Multivariate regression analysis such as MLR and PLS were carried out in order to discover the contribution of the whole molecule and substituents for biological activity. More specifically, the understanding of the *de novo* contribution of the functional groups helped to understand interactions between substitution on the N- chain of the indole ring and the transmembrane region of the H_1-receptor active site. The set of 41 compounds were subjected to stepwise multiple linear regression analysis in order to develop the QSAR model, considering antihistaminic activity as dependent variables and all the reduced set of descriptors as independent variables. When the multiple regression analysis was performed without deleting any outlier, we retrieved a statistically insignificant model. In order to improve the predictability and reliability of the model the outliers were detected and it was found that four compounds (**6**, **7**, **10**, and **23**) had high residual values and were too far away from the regression line. Various models developed after deleting the four compounds one by one and in combination exhibited a high value

Table 1. Statistical values for the most relevant descriptors used in QSAR

	Abbreviation	Coefficient	Jacknife	*t*-Value	Covariance SE
Verloop L (subs.5)	X_1	–0.24	0.02	–9.86	0.02
Verloop B_3 (subs.5)	X_2	–0.30	0.02	–9.09	0.03
Total dipole moment (Whole molecule)	X_3	0.12	0.01	7.71	0.01
KAlpha1 index (Whole molecule)	X_4	0.06	0.02	3.72	0.02

Table 2. Correlation matrix of classical descriptors used in QSAR models

	Verloop L (subs.5)	Verloop B_3 (subs.5)	Total dipole moment (Whole molecule)	KAlpha1 index (Whole molecule)	Log IC_{50}
Verloop L (subs.5)	1				
Verloop B_3 (subs.5)	–0.55	1			
Total dipole moment (Whole molecule)	0.36	–0.40	1		
KAlpha1 index (Whole molecule)	0.18	0.23	0.09	1	
Log IC_{50}	–0.13	–0.51	0.61	–0.01	1

of r^2 (0.86) and r^2cv (0.82) for MLR analysis as shown in Table 3. In addition to MLR, the data set was also subjected to PLS analysis to check soundness of results obtained from MLR analysis. The PLS model generated showed comparable results to that of MLR (*23*) as exhibited by the high value of r^2 (0.85) and r^2cv (0.82) as shown in Table 4.

3.1. *Validation of statistical output of multivariate equation (PLS and MLR)*

The model developed using MLR and PLS were validated and checked for predictability and robustness on the basis of the r^2, r^2cv, F value and its standard error S value.

3.1.1. *Squared correlation coefficient (r^2)*

In statistics, the coefficient of determination r^2 is the proportion of variability in a data set that is accounted for by a statistical QSAR model. In this determination, the term "variability" is defined as the sum of squares. There are equivalent expressions for r^2 based on analysis of the fractionof total variance in the data, which is explained by the regression model (*24*). It can be calculated using the following formula.

$$r^2 = 1 - \sum\Delta^2/Syy$$

Where, $Syy = \sum(y_{obs} - y_{mean})$, $\sum\Delta^2 = \sum(y_{obs} - y_{cal})^2$. Where "$y_{obs}$" is observed biological activities, "y_{mean}" is mean of biological activities value, and "y_{cal}" is calculated biological activity used in the equation. Also r^2 is used to describe the goodness of fit of the data. Its value should be greater than 0.7 for a sound model. The statistical significance of the generated QSAR model was evaluated in terms of r^2 values (MLR = 0.8657 and PLS = 0.859) which explained a 86.57% and 85.90% variance in biological activity. Moreover the r^2 values for the test set (MLR = 0.76 and PLS = 0.85) were also found to be significant.

3.1.2. *Cross validated regression coefficient (r^2cv)*

Cross-validation is an important tool to avoid over fitting of data, as over-fitting will give low accuracy on validation. The predictive value of the models were evaluated by LOO cross validation (*24*). The cross validated coefficient r^2cv was calculated using the formula for PLS analysis.

$$r^2cv = 1 - \sum(Y_{pred} - Y_{obs})^2 / \sum(Y_{pred} - Y_{mean})$$

Where Y_{pred}, Y_{obs}, and Y_{mean} are predicted, actual and mean values of the target propert (pIC_{50}), respectively. $\sum(Y_{pred} - Y_{obs})^2$ is the predictive sum of squares (PRESS). And for MLR, it can be calculated using the following formula.

$$r^2cv = SD - PRESS/SD$$

Where SD is standard deviation and predicted residual sum of square (PRESS) = $\sum(y_{obs} - y_{cal})^2$. This method is used to predict the property value for a compound from the data set, which in turn is predicted from the regression equation calculated from the data for all of the compounds. For evaluation, predicted values can be used for squared correlation coefficient criteria. The r^2cv should be more than 0.6 and a small difference between r^2 and r^2cv indicates a good internal predictive ability of developed model. A small difference in r^2 (0.86) and r^2cv (0.82) for MLR values is indicative of the high predictive ability of the developed model. Similarly the r^2 (0.85) and r^2cv (0.82) values generated by PLS were also very close.

3.1.3. *S value (SD)*

SD is measured as the error mean square, which expresses the variation of the residuals or the variation about the regression line. It indicates, how well the function derived by the QSAR analysis predicts the

Table 3. Various equations derived after removing of outliers by the MLR method

S. No.	Equation	r	r^2	r^2cv	S value	F value	Name of outliers
1.	$Y = -0.23X_1 - 0.32X_2 + 0.11X_3 + 0.05X_4 - 1.49$	0.86	0.75	0.25	0.16	27.40	ND
2.	$Y = -0.24X_1 - 0.30X_2 + 0.12X_3 + 0.06X_4 - 1.64$	0.90	0.82	0.73	0.13	40.20	**7**
3.	$Y = -0.23X_1 - 0.29X_2 + 0.12X_3 + 0.06X_4 - 1.72$	0.92	0.84	0.76	0.13	44.47	**7**, **23**
4.	$Y = -0.23X_1 - 0.32X_2 + 0.10X_3 + 0.05X_4 - 1.43$	0.89	0.79	0.70	0.15	32.29	**23**, **6**
5.	$Y = -0.23X_1 - 0.30X_2 + 0.11X_3 + 0.06X_4 - 1.72$	0.93	0.86	0.82	0.12	51.58	**6**, **7**, **10**, **23**

Table 4. Various equations derived after removing of outliers by the PLS method

S. No.	Equation	Statistical significance	Cross validation (r^2cv)	Fraction of variance (r^2)	Name of outliers
1.	$Y = -0.23X_1 - 0.30X_2 + 0.12X_3 + 0.06X_4 - 1.78$	1.07	0.65	0.75	ND
2.	$Y = -0.23X_1 - 0.28X_2 + 0.13X_3 + 0.06X_4 - 1.92$	0.99	0.77	0.82	**7**
3.	$Y = -0.23X_1 - 0.28X_2 + 0.12X_3 + 0.06X_4 - 1.95$	1.02	0.77	0.84	**7**, **23**
4.	$Y = -0.22X_1 - 0.30X_2 + 0.12X_3 + 0.06X_4 - 1.73$	0.95	0.75	0.78	**23**, **6**
5.	$Y = -0.22X_1 - 0.27X_2 + 0.12X_3 + 0.07X_4 - 2.03$	0.97	0.82	0.86	**6**, **7**, **10**, **23**

observed biological activity. Its value considers the number of object 'n' and the number of variable 'k'. Hereby, SD depends not only on the quality of fit but also on the number of degrees of freedom. The low values for the standard error estimate (0.118) of the developed model further testify about the statistical significance of the developed model. It can be calculated using the following formula.

$$SD = \sqrt{\sum(y_{obs} - y_{cal})^2/n - k - 1}$$

3.1.4. *F value (Fischer value)*

It is a measure of the statistical significance of the regression model. It can be calculated using the following formula.

$$F = r^2(n - k - 1)/k(1 - r^2)$$

A high value of F (51.584) indicates that the model is statistically significant and a lower S value also supported the quality of the model.

3.2. *Test set prediction*

3.2.1. *Internal test set prediction*

The internal predictive capability of the QSAR model was also checked using test sets of compounds that were excluded during model development. All the compounds in the test set were treated in a manner analogous to the compounds in the training set. The r^2 value of MLR = 0.763 and PLS = 0.855 derived for the test set illustrate the high predictive ability of the developed model. The actual and predicted activity obtained after MLR and PLS analysis for the training and test set of antihistamines are given in Supplemental Tables S2 and S3 (*http://www.ddtjournal.com/getabstract.php?id=538*) and their corresponding graphs are shown in Figures 1 and 2.

3.2.2. *External test set prediction*

A model is claimed to be validated when it is also able to predict the activities of external test set compounds. So, the generated model was further evaluated by an external test set comprising of H_1 receptor antagonists from the literature (*25*). Thirteen compounds were selected for the external test set showing diversity in an activity range as given in Table 5. The overall squared correlation coefficient was 0.68 (MLR) and 0.68 (PLS) for the external test set (Figures 3 and 4) testifying to the validity of the model.

3.3. *Interpretation of the generated multivariate model*

The highly predictive and robust MLR and PLS models were selected on the basis of statistical significance of the regression equations (Tables 3 and 4) obtained from training set compounds, which explains the relationship between biological activity and structures of molecules. Descriptors obtained from the final model (both MLR and PLS) were namely Verloop L (substitution R_5), Verloop B_3 (substitution R_5), Total dipole moment (whole molecule), and KAlpha1 index (whole molecule).

The Verloop L and Verloop B_3 parameters are steric parameters developed by Verloop and co-workers and are used to characterize geometry of substitution groups in the molecule. The length parameter, Verloop L, is defined as calculated length of the substituent along the axis of the bond connecting substituent with parent molecule (*26*). Width parameter Verloop B_3 characterizes distribution of atoms in the substituent with respect to the connecting bond. In addition, the Verloop B_3 parameter describes width of the substituent in the direction perpendicular to L (*27*). The equation clearly reveals that a decrease in length and width of substitution R_5 will increase binding affinity of compounds.

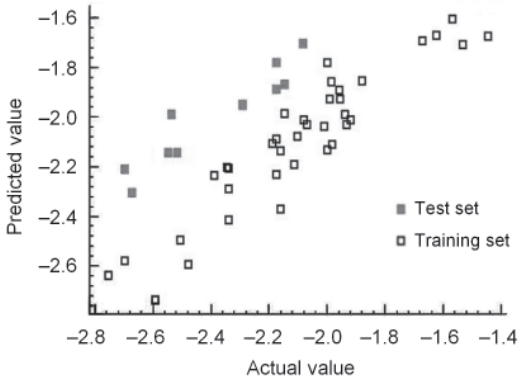


Figure 1. Plot of actual versus predicted values for the training set and test set molecules with the help of the MLR statistical method.

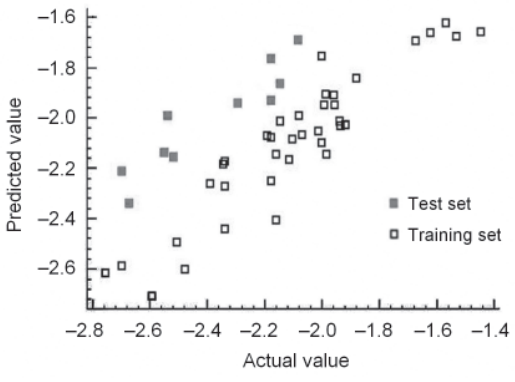
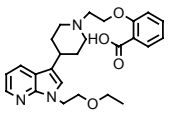
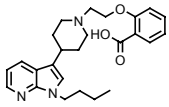
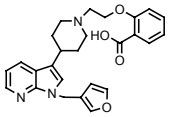
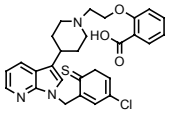
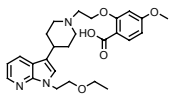
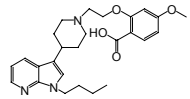
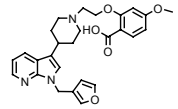
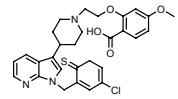
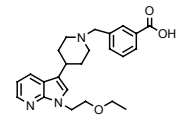
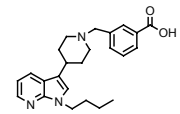
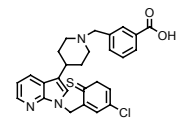
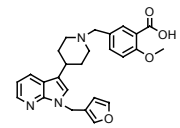
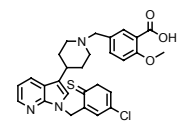


Figure 2. Plot of actual versus predicted values for the training set and test set molecules with the help of the PLS statistical method.

Table 5. Structures along with –log IC_{50} values and estimated value of the 13 piperidinylpyrrolopyridines and its derivatives as H_1 antagonists for external validation

Compound name	Structure of compound	Actual activity IC_{50} (nM)	Actual activity in $-\log IC_{50}$	Estimated activity by MLR	Estimated activity by PLS
1		190	–2.27875	–2.0532	–2.08818
2		235	–2.37107	–1.95353	–2.01655
3		225	–2.35218	–1.89567	–1.92184
4		295	–2.46982	–2.35658	–2.36357
5		185	–2.26717	–2.10389	–2.15123
6		215	–2.33244	–1.86682	–2.00285
7		170	–2.23045	–1.73856	–1.84707
8		240	–2.38021	–2.16555	–2.26474
9		403	–2.6053	–2.46071	–2.33642
10		275	–2.43933	–2.13215	–2.15197
11		330	–2.51851	–2.58772	–2.57525
12		205	–2.31175	–1.9529	–2.00152
13		365	–2.56229	–2.27669	–2.33247

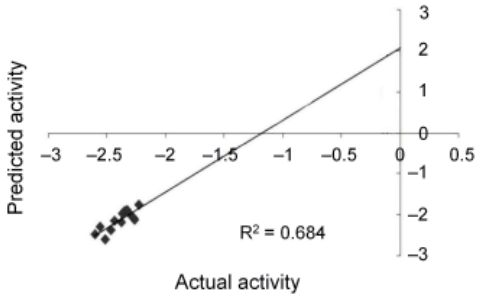


Figure 3. Plot of the correlation graph between experimental and estimated activities of the external test set molecules by MLR.

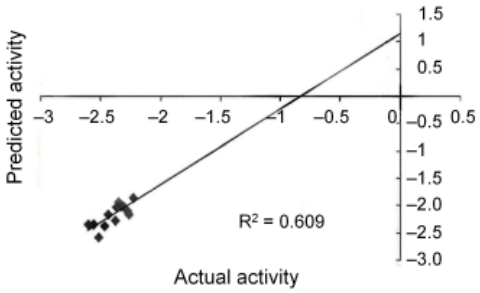


Figure 4. Plot of the correlation graph between experimental and estimated activities of the external test set molecules by PLS.

The equation in Tables 3 and 4 clearly shows that substitution on the N- chain of the indole ring is negatively correlated with H_1-antihistamine activity. This apparently explains the difference in activity of compounds **13** and **27** which differ only in substitution on the N- chain of the indole ring. Compound **13** exhibited a higher antihistaminic activity because it possesses a lower value for length and width parameters at this position when compared to compound **27**.

The decrease in H_1 antihistaminic activity with the length of substituents can be further demonstrated by the activity of compounds **1**, **26**, and **27**, which differ only in a single substitution. Compound **1** is more active than compounds **26**, and **27** as its R_5 substitution is shorter than the 2-[1,4]-dioxan-2-ylethyl of compound **26**, and the 2-pyridin-2-ylethyl of compound **27**. Compounds **26** and **27** have more or less the same length value. The above discussion reveals that while designing newer molecules care must be taken to keep the length and width of substitution at R_5 in a limited range because length and width of the substitution will cause steric hindrance between drug and receptor and consequently will decrease activity.

Total dipole moment for a whole molecule may be approximated as the vector sum of individual bond dipole moments. Moreover, dipole moment is a partial charge-dependent parameter calculated on the basis of center of charge over the substitution as the origin. The total dipole moment describes the electrostatic interaction between drug and receptor. It can be calculated using the following formula.

$$\mu = e\sum r_i q_i$$

Where r_i is the distance of the ith atom from the origin and q_i is the atomic charge of the ith atom. The descriptor uses Debye units. The total dipole moment is often considered as the direct characteristic of polarity of a molecule (*28*). The present series have an indolylpiperidine ring, the aromatic π system of which is formed by π electrons. Quantum mechanical calculations showed distribution of electron density among the atoms of the indole ring which results in a large molecular dipole moment which allows possibility of dipole-dipole interactions between the H_1 receptor and compounds under consideration. The basic nitrogen of the indole ring is confined to the region accessible to its counter ion on the histamine H_1-receptor, *i.e.*, the carboxylate group of aspartic acid (Asp116) (*29*).

In the regression equation total dipole moment of whole molecules is positively correlated with biological activity. It means that larger electronic properties of the compounds play a key role in the interaction between drug and receptor. Increasing total dipole moment of the whole molecule can increase binding between molecules and receptor and thus can increase interaction between molecules and the H_1 receptor. The above interpretation is in line with mutational studies done on cetrizine, a potent H_1 receptor antagonist. Cetrizine contains a COOH group which interacts with the protonated lysine (Lys200) present in the transmembrane domain V of the H_1 receptor (*30*). Likewise Indolylpiperidine derivatives containing a COOH group form an ionic bond/H bond with the positively charged amino group of lysine.

Another descriptor entering in the model was KAlpha1 which helps to differentiate the molecules according to their size, degree of branching, flexibility, and overall shape (*31*). The KAlpha1 index encodes information about several attributes of molecular shape, based upon atom count and path count of various orders. This index explains the elongated nature of molecules and their branches at the end. Each K index value would imply shape identity. Atoms other than Sp^3 hybrids make a different size contribution to a molecule, thereby influencing its overall shape. Non-carbon Sp^3 atoms should be counted more or less than one, the increment or decrement, called 'alpha' being based on the size contribution of the atom in question relative to a carbon Sp^3. One basis of evaluating alpha is to use covalent radius of the atom. The quality of the equation and easy interpretation of the indices make this result useful for prediction of possible antihistaminic potency of some newer designed compounds (*32*).

The multivariate regression equation reveals that the KAlpha1 index of the whole molecule of the compound

is correlated positively with biological activity. It means that if we increase the shape and branching of the lead compound there will be an increase in H_1-antihisaminic activity because shape or steric configuration of a molecule have a potential influence on physical properties and biological activity. The QSAR studies suggested that the catalytic site of the protein and the steric requirements of the compounds play an important factor in the antagonist's potency.

This QSAR study clearly indicates that optimum steric and electronic properties of whole molecules are important for favorable interaction with the receptor. On the basis of the final multivariate regression model, some newer compounds were designed to find molecules with higher antihistaminic potencies than the existing series of indolylpiperidinyl benzoic acid derivatives.

4. Conclusion

On the basis of present study, it can be concluded that physicochemical descriptors have sufficient reliability to relate to the biological activity of indolylpiperidinyl benzoic acid molecules with their structural features. This physicochemical study suggests that the electronic and steric interaction between drug and receptors are dominant in whole molecules, whereas the multidimensional steric interaction is dominant due to substitution at R_5 in this series. It means that low length and width of substitution R_5 is important for interaction between drug and receptors. The study provides better understanding of the structural features and their binding affinities to the histamine H_1 receptor and helped in the design of orally active more potent H_1 receptor antagonists.

Acknowledgements

Computational resources were provided by Banasthali University, and the authors thank the Vice Chancellor, for extending all the necessary facilities.

References

1. Simons FE. H_1-Antihistamines: More relevant than ever in the treatment of allergic disorders. J Allerg Clin Immunol. 2003; 112:S42-S52.
2. Das AK, Yoshimura S, Mishima R, Fujimoto K, Mizuguchi H, Dev S, Wakayama Y, Kitamura Y, Horio S, Takeda N, Fukui H. Stimulation of histamine H_1 receptor up-regulates histamine H_1 receptor itself through activation of receptor gene transcription. J Pharmacol Sci. 2007; 103:374-382.
3. Bousquet J, Van Cauwenberge P, Khaltaev N; Aria Workshop Group; World Health Organization. Allergic rhinitis and its impact on asthma. J Allergy Clin Immunol. 2001; 108(Suppl 5):S147-S334.
4. Holgate ST, Canonica GW, Simons FE, Taglialatela M, Tharp M, Timmerman H, Yanai K; Consensus Group on New-Generation Antihistamines. Consensus Group on New-Generation Antihistamines (CONGA): Present status and recommendations. Clin Exp Allergy. 2003; 33:1305-1324.
5. Scadding GK. Clinical assessment of antihistamines in rhinitis. Clin Exp Allergy. 1999; 29:77-81.
6. Simons FE, Simons KJ. The pharmacology and use of H_1-receptor-antagonist drugs. N Engl J Med. 1994; 330:1663-1670.
7. Kay GG. The effects of antihistamines on cognition and performance. J Allergy Clin Immunol. 2000; 105:S622-S627.
8. Shigemoto Y, Shinomiya K, Mio M, Azuma N, Kamei C. Effects of second-generation histamine H_1 receptor antagonists on the sleep-wakefulness cycle in rats. Eur J Pharmacol. 2004; 494:161-165.
9. Barbey JT, Anderson M, Ciprandi G, Frew AJ, Morad M, Priori SG. Cardiovascular safety of second-generation antihistamines. Am J Rhinol. 1999; 13:235-243.
10. DuBuske LM. Second-generation antihistamines: The risk of ventricular arrhythmias. Clin Ther. 1999; 21:281-295.
11. Hansch C, Verma RP. Overcoming tumor drug resistance with C2- modified 10-deacetyl 7-propionyl cephalomannines – A QSAR study. Mol Pharmaceut. 2009; 6:849-860.
12. Fonquerna S, Miralpeix M, Pagès L, *et al*. Synthesis and structure-activity relationships of novel histamine H_1 antagonists: Indolylpiperidinyl benzoic acid derivatives. J Med Chem. 2004; 47:6326-6337.
13. Tsar 3.3, Oxford Molecular Ltd., The Medawar Centre, Oxford Science Park, Oxford, 2000.
14. Hendrickson MA, Nicklaus MC, Milne GWA, Zaharevitz D. CONCORD and CAMBRIDGE: Comparison of computer-generated chemical structures with X-ray crystallographic data. J Chem Inf Comput Sci. 1993; 33:155-163.
15. Sadowski J, Gasteiger J. From atoms and bonds to three-dimensional atomic coordinates: Automatic model builders. Chem Rev. 1993; 93:2567-2581.
16. Molecular Modeling and Drug Design (Vinter JG, Gardner M, eds.). CRC Press Inc., Boca Raton, FL, USA, 1994.
17. Kovatcheva A, Buchbauer G, Golbraikh A, Wolschann P. QSAR modeling of r-campholenic derivatives with sandalwood odor. J Chem Inf Comput Sci. 2003; 43:259-266.
18. Kovatcheva A, Golbraikh A, Oloff S, Xiao Y, Zheng W, Wolschan P, Buchbaucer G, Tropsha A. Combinatorial QSAR of ambergris fragnance compounds. J Chem Inf Comput Sci. 2004; 44:582-595.
19. Paliwal SK, Pal M, Siddiqui AA. Quantitative structure activity relationship analysis of angiotensin II AT1 receptor antagonists. Med Chem Res. 2010; 19:475-489.
20. Shen M, LeTiran A, Xiao Y, Golbraikh A, Kohn H, Tropsha A. Quantitative structure-activity relationship analysis of functionalized amino acid anticonvulsant agents using κ nearest-neighbor and simulated annealing PLS methods. J Med Chem. 2002; 7:2811-2823.
21. Kim KH. Outliers in SAR and QSAR: 2. Is a flexible binding site a possible source of outliers? J Comput Aided Mol Des. 2007; 21:421-435.
22. Paliwal S, Narayan A, Paliwal S. Quantitative structure activity relationship analysis of dicationic diphenylisoxazole as potent anti-trypanosomal agents. QSAR Comb Sci. 2009; 28:1367-1375.
23. Cramer RD. Partial least squares (PLS): Its strengths and limitations. Perspect Drug Discov Des. 1993; 1:269-278.
24. Hawkins DM, Basak SC, Mills D. Assessing model fit by

cross-validation. J Chem Inf Comput Sci. 2003; 43:579-586.
25. Fonquerna S, Miralpeix M, Pagès L, Puig C, Cardús A, Antón F, Vilella D, Aparici M, Prieto J, Warrellow G, Beleta J, Ryder H. Synthesis and structure-activity relationships of piperidinylpyrrolopyridine derivatives as potent and selective H_1 antagonists. Bio Med Chem Lett. 2005; 15:1165-1167.
26. Paliwal S, Das S, Yadav D, Saxena M, Paliwal S. Quantitative structure activity relationship (QSAR) of N^6-substituted adenosine receptor agonists as potential antihypertensive agents. Med Chem Res. 2011; 20:1643-1649.
27. Molecular Descriptors in QSAR/QSPR (Karelson M, ed.). Wiley Interscience, New York, NY, USA, 2000; pp. 220-221.
28. Paliwal S, Seth D, Yadav D, Yadav R, Paliwal S. Development of a robust QSAR model to predict the affinity of pyrrolidine analogs for dipeptidyl peptidase IV (DPP- IV). J Enzyme Inhib Med Chem. 2011; 26:129-140.
29. Ohta K, Hayashi H, Mizuguchi H, Kagamiyama H, Fujimoto K, Fukui H. Site-directed mutagenesis of the histamine H_1 receptor: Roles of aspartic acid 107, asparagine198 and threonine194. Biochem Biophys Res Commun. 1994; 203:1096-1101.
30. Gillard M, Van Der Perren C, Moguilevsky N, Massingham R, Chatelain P. Binding characteristics of cetirizine and levocetirizine to human H(1) histamine receptors: Contribution of Lys(191) and Thr(194). Mol Pharmacol. 2002; 61:391-399.
31. Kier LB. Shape indexes of orders one and three from molecular graphs. Mol Inform. 1986; 5:1-7.
32. Kier LB. Distinguishing atom differences in a molecular graph shape index. Mol Inform. 1986; 5:7-12.

(Received December 5, 2011; Revised February 10, 2012; Re-revised March 16, 2012; Accepted April 3, 2012)