

A NOVEL QSAR MODEL FOR EVALUATING AND PREDICTING THE INHIBITION ACTIVITY OF H₁-RECEPTOR ANTAGONISTS: A SERIES OF THIENOPYRIMIDINE DERIVATIVES*Meenakshi Singh^a, Sushil K Singh^a, Mahesh T Chhabria^b^aDepartment of Pharmaceutics, Institute of Technology, Banaras Hindu University (BHU), Varanasi-221005, U.P., India^bDepartment of Pharmaceutical Chemistry, L.M. College of Pharmacy, Navrangpura, Ahmedabad-380009, Gujarat, India*Corresponding Author's Email: meenu1013@gmail.com

Received 22 Nov 2011; Revised 26 Dec 2011; Accepted 28 Dec 2012, Available online 20 Jan 2012

ABSTRACT

A Quantitative Structure Activity Relationship (QSAR) study has been established using combination of most influencing physicochemical parameters viz. thermodynamic, electronic, geometric & quantum mechanical descriptors, and H₁-antihistaminic activity of a series of thienopyrimidines, the novel Histamine H₁ receptor antagonists. Genetic function approximation (GFA) technique was used to identify the descriptors that have influence on biological activity. Dipole, AlogP 98, Jurs and LUMO descriptors were found to influence biological activity significantly. Lipophilicity of compounds was found to have a significant role in H₁ Histaminic inhibition along with other thermodynamic, spatial and electronic descriptors. Positive contribution of Dipole, AlogP 98 descriptors suggests that molecules with lipophilic-electronic substituents are more likely to improve the potency. Developed models were found to be significant and predictive as evidenced from their internal and external cross-validation statistics.

Keywords: H₁-receptor antagonists; thienopyrimidines; molecular descriptor; genetic function approximations; cross-validation; quantitative structure activity relationship

Abbreviations:

QSAR : Quantitative structure activity relationship
GFA : Genetic function approximation

LOF : Friedman's lack of fit
VIF : Variance inflation factor

INTRODUCTION

Histamine, chemically 5-(2-aminoethyl) imidazole, generally known as local hormone or autocooid is the first vasoactive amine identified in the body, that is produced locally in response to some stimulus¹. The H₁ histamine receptor (H₁HR), one of the G protein-coupled receptors (GPCRs) mediates the functional effects of histamine in the multiple cell types through activation of the G_{q/11} heterotrimeric G protein. These histamine-induced intracellular messengers promote diverse functions in multiple cell types, including smooth muscle and non smooth muscle contraction^{2,3} and exocytotic release of neurotransmitters and various autocrine/paracrine factors^{4,5}, both of which can contribute to inflammation and inflammatory disease processes^{6,7}.

Compounds containing thienopyrimidine nucleus represent a very important chemical class of antagonists in drug discovery due to their wide range of pharmacological properties, including antiallergic, anti-inflammatory, analgesic, antispasmodic, antibacterial, antifungal etc⁸. Therefore, the interest in developing quantitative structure activity relationships (QSAR) for this class of bioactive compounds remains high in medicinal chemistry.

Quantitative structure activity relationship (QSAR) modeling is an area of computational research that constructs models to correlate structural features with biological activity, which provides information that is useful for drug design and discovery. The underlying assumption is that the variations of biological activity within a series of similar structures can be correlated with changes in measured or computed molecular features of

the molecules. Therefore, QSAR study remains as a very useful tool in the era of modern drug discovery to get better insight into structure activity relationships⁹⁻¹². QSAR models are analyzed with various statistical parameters to assess reliability and robustness and are useful for various purposes like lead optimization, risk assessment, toxicity prediction and regulatory decisions, including the prediction of activities of untested chemicals¹³.

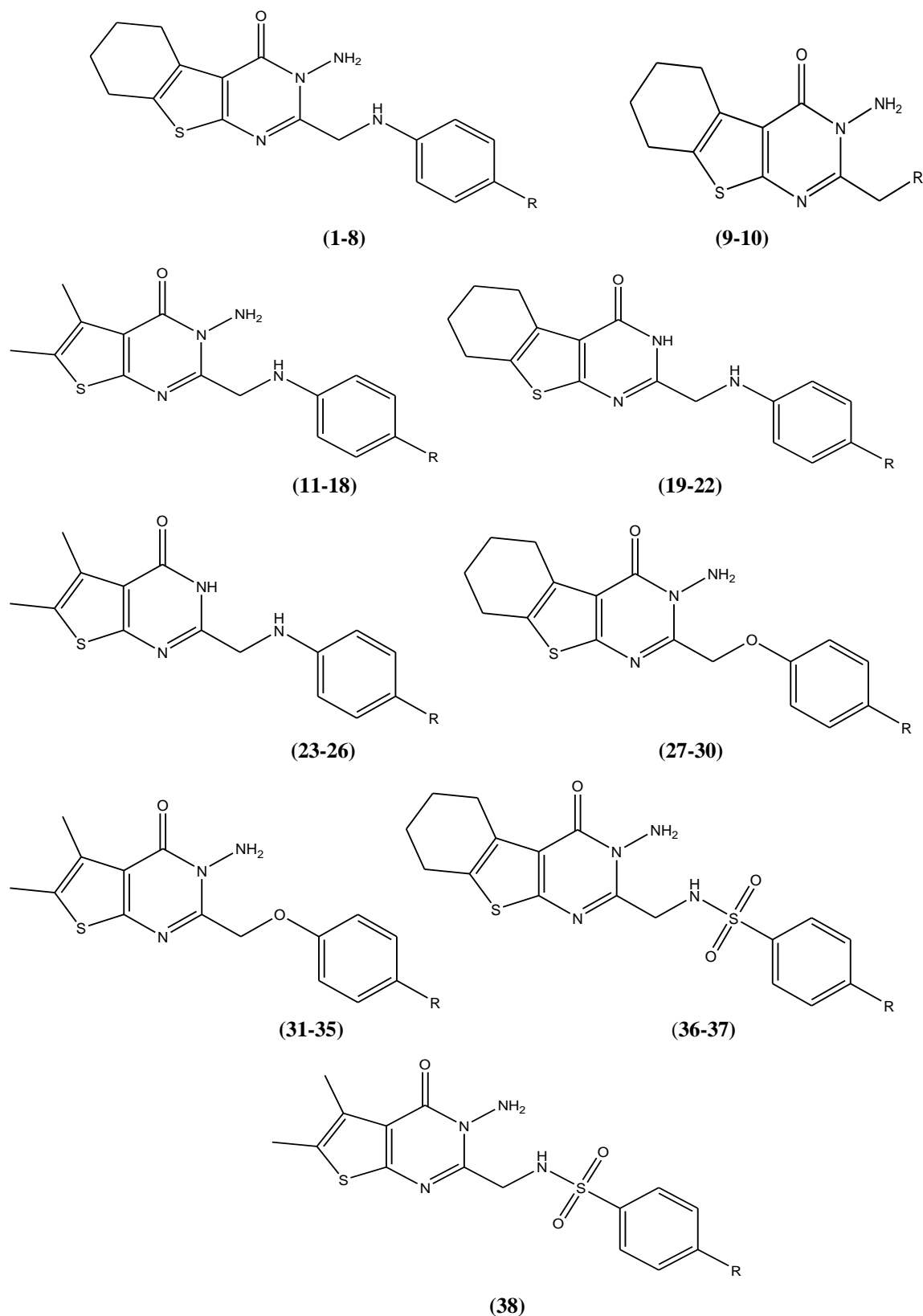
To get insight into structure-activity relationship, we have developed QSAR models for a series of thieno[2,3-*d*]pyrimidin-4(3H)-ones using Genetic Function Approximation (GFA) technique that automates the search for QSAR models by combining a genetic algorithm with statistical modeling tools. Replacing the regression analysis with the GFA algorithm allows the construction of models competitive with, or superior to, standard techniques and makes available additional information not provided by other techniques. GFA can build models using not only linear polynomials but also higher-order polynomials, splines, and Gaussians. By using spline based terms, GFA can perform a form of automatic outlier removal and classification. Therefore, GFA technique has been successfully applied for the generation of variety of QSAR models^{14,15}, such model provides structure-activity insights, which can be used for designing of new compounds and activity prediction prior to synthesis. The goal of this research is to rationalize the title compounds in terms of physicochemical and structural requirements for enhanced binding affinity to the H₁-Histamine receptor (H₁HR).

Data set

In present study, H₁-Histaminic inhibitory activity data of a series of thieno [2,3-*d*]pyrimidin-4(3H)-ones reported by Shirsath, V.S. *et al.*, was selected¹⁶. The IC₅₀(μM) values, were converted to negative logarithmic scale (pIC₅₀) to achieve normal distribution, used as dependent variable in

the QSAR study. Total set of 38 compounds were randomly divided into training set and test set of thirty one and seven compounds, respectively. Structures of all the compounds used for QSAR analysis and their H₁-Histaminic inhibitory activity (IC₅₀ in μM) are given in Table 1. Structures of all compounds used in this study were sketched by using Visualizer module of Discovery studio 2.1 software (Accelrys Inc., USA)¹⁷.

Table 1: Chemical structures and biological activity of total data set containing Thienopyrimidine derivatives



| Compound No. | R | IC ₅₀ (μM) | pIC ₅₀ |
|--------------|------------------------------------|-----------------------|-------------------|
| 1 | -H | 0.37 | 6.422 |
| 2 | -OCH ₃ | 0.12 | 6.906 |
| 3 | -OC ₂ H ₅ | 0.49 | 6.308 |
| 4 | -F | 0.29 | 6.537 |
| 5 | -CH(CH ₃) ₂ | 0.66 | 6.180 |
| 6* | -CH ₃ | 0.20 | 6.679 |
| 7* | -Br | 0.25 | 6.600 |
| 8* | -Cl | 0.21 | 6.673 |
| 9 | -benzylamino | 0.49 | 6.301 |
| 10 | -morpholino | 0.47 | 6.321 |
| 11 | -H | 0.41 | 6.378 |
| 12 | -OCH ₃ | 0.39 | 6.407 |
| 13 | -OC ₂ H ₅ | 0.58 | 6.235 |
| 14 | -F | 0.33 | 6.473 |
| 15 | -Br | 0.32 | 6.485 |
| 16 | -Cl | 0.30 | 6.515 |
| 17 | -CH(CH ₃) ₂ | 0.64 | 6.191 |
| 18* | -CH ₃ | 0.25 | 6.586 |
| 19 | -H | 0.42 | 6.375 |
| 20 | -CH ₃ | 0.29 | 6.531 |
| 21 | -F | 0.35 | 6.447 |
| 22 | -piperidino | 0.39 | 6.406 |
| 23 | -H | 0.51 | 6.290 |
| Compound No. | R | IC ₅₀ (μM) | pIC ₅₀ |
| 24 | -CH ₃ | 0.32 | 6.482 |
| 25 | -OC ₂ H ₅ | 0.65 | 6.181 |
| 26 | -Cl | 0.38 | 6.415 |
| 27 | -3-CH ₃ | 0.62 | 6.207 |
| 28 | -4-Cl | 0.59 | 6.229 |
| 29 | -3-CH ₃ ,4-Cl | 0.42 | 6.376 |
| 30* | -3,4-CH=CH-CH=CH- | 1.80 | 5.744 |
| 31 | -H | 0.67 | 6.173 |
| 32 | -4-Cl | 0.44 | 6.356 |
| 33 | -3-CH ₃ ,4-Cl | 0.63 | 6.200 |
| 34* | -4-CH ₃ | 0.79 | 6.102 |
| 35* | -3,4-CH=CH-CH=CH- | 1.05 | 5.978 |
| 36 | -H | 0.32 | 6.494 |
| 37 | -CH ₃ | 0.36 | 6.443 |
| 38 | -H | 0.46 | 6.337 |

Where, * Compounds in the test set

Regression analysis

GFA is genetic principle based technique of variable selection, which combines Holland's genetic algorithm with Friedman's multivariate adaptive regression (MARS) splines to generate population of equations that best fit the training set data¹⁸. Employing this technique in QSAR analysis, begins with generation of population of equations (set at 100 by default in the Discovery studio 2.1 software), rather than one single equation for correlation between the dependent (biological activity) and independent variables (descriptors). From this set of 100 equations, two best models are selected as "parents" and *genetic crossover* operations were performed at random and a new model created from each parents. A default value of number of crossing over was set at 5000. The goodness of each progeny equation is assessed by Friedman's lack of fit (LOF) score, which is given by following formula:

$$\text{LOF} = \text{LSE}/[1 - (c + dp)/m]^2$$

Where LSE is the least square error, c is the number of basis functions in the model, d is smoothing parameter, p is the numbers of descriptors and m is the number of observations in the training set. The smoothing parameter, which controls the scoring bias between equations of different sizes, was set at default value of 0.5. The length of equation was fixed to six terms and, the population size was established as 100, the equation term was set to quadratic polynomial with spline functionality. The statistical qualities of generated models were judged by various parameters such as regression coefficient (r), adjusted regression coefficient (r_{adj}), cross-validated regression coefficient (r_{cv}) and F-test values to select best equations, from 100 equations.

Selection of molecular descriptors

A common practice in building QSAR models is to select descriptors, considered as vital element of ligand based study, and should not be intercorrelated and show lesser degree of multi-collinearity. To develop QSAR models, descriptors were selected by analysis of correlation matrix and variance inflation factor¹⁹. In this study, sixty eight molecular descriptors representing electronic, spatial, structural, thermodynamic, geometric, topological and quantum mechanical properties were calculated using calculate molecular properties protocol of the PC based software, Discovery Studio 2.1. (Accelrys Inc, USA).

Table 2: List of descriptors used in the study

| Sr. No. | Descriptors | Type | Description |
|---------|--|--------------------|--|
| 1 | AlogP 98 | Thermodynamic | Determine the lipophilicity of the molecule. |
| 2 | Dipole_Z, Dipole_Y | Electronic | The dipole moment descriptor is a 3D electronic descriptor that indicates the strength and orientation behavior of a molecule in an electrostatic field. |
| 3 | Jurs_TASA, Jurs_RPCG | Geometric | Jurs descriptors combine shape and electronic information to characterize molecules. The descriptors calculated by mapping atomic partial charges on solvent accessible surface areas of individual atoms. |
| 4 | LUMO_Eigenvalue_VAMP, Dipole_Mag_VAMP | Quantum mechanical | Energy of the highest and lowest occupied molecular orbitals |

From the data of correlation matrix, highly correlated descriptors with value of 0.9 or above (implying highest multicollinearity) and with zero value were removed from the study. Remaining descriptors were used to develop QSAR models using GFA technique. Descriptors included

in developing QSAR models are listed and described in Table 2. Correlation matrix of descriptors used in the QSAR study is given in the Table 3.

Validation test

Validation is necessary in QSAR methodology to prove that the generated models are acceptable for its intended purpose. Statistical significance of the generated QSAR models was analyzed by variance inflation factor (VIF) analysis, cross validation or internal validation with training set and external validation with test data set.

To check the inter-correlation of descriptors, variance inflation factor (VIF) analysis was performed. VIF value is calculated from $1/1-r^2$, where r^2 is the multiple correlation coefficient of one descriptor's effect regressed on the remaining molecular descriptors. VIF value greater than 10, implying chance-correlation and hide the information of descriptors by inter-correlation and multicollinearity²⁰.

To determine the quality of model internally, cross-validation (CV) techniques are extensively employed and is analyzed by the value of correlation coefficient of the cross-validation procedure, that is r_{cv}^2 whose values greater than 0.5 indicates robustness and significance for a satisfactory QSAR model and it is calculated by following formula.

$$q^2 = 1 - \frac{\sum(Y_{\text{pred}} - Y_{\text{actual}})^2}{\sum(Y_{\text{actual}} - Y_{\text{mean}})^2}$$

To avoid chance correlation, an ultimate reliable validation procedure was carried out and is examined by means of external validation in terms of values of residuals, r_{pred}^2 and r_{m}^2 using test set compounds. The predictive correlation coefficient (r_{pred}^2) value is based on test set only and is defined by the following equation²¹.

$$r_{\text{pred}}^2 = (\text{SD} - \text{PRESS})/\text{SD}$$

where, SD is the sum of the squared deviations between the biological activities of the test set and mean activity of the training set molecules and PRESS is sum of the squared deviation between predicted and actual activity values for every molecule in the test set.

| Compound No. | LUMO_ | Dipole_Mag _VAMP | AlogP 98 | Dipole_Y | Dipole_Z | Jurs_TASA |
|--------------|---------------------|---------------------|-------------|----------|----------|-----------|
| | Eigenvalue_ VAMP | | | | | |
| 1 | -0.788 | 2.515 | 3.019 | -0.251 | -0.008 | 501.846 |
| 2 | -0.926 | 4.7 | 3.003 | -1.044 | 0.045 | 519.657 |
| 3 | -0.921 | 4.383 | 3.351 | -1.091 | 0.063 | 562.628 |
| 4 | -0.973 | 5.723 | 3.224 | 0.962 | -0.171 | 460.726 |
| 5 | -0.913 | 3.401 | 4.213 | 0.069 | 0.036 | 564.4 |
| 6 | -0.761 | 1.429 | 2.748 | 0.507 | 0.699 | 521.636 |
| 7 | -0.559 | 1.333 | 1.388 | -0.515 | 0.431 | 449.313 |
| 8 | -0.903 | 4.043 | 2.267 | 0.158 | 0 | 477.793 |
| 9 | -0.896 | 4.251 | 2.251 | 1.807 | 0 | 492.607 |
| 10 | -0.892 | 3.943 | 2.6 | 1.739 | 0.003 | 534.79 |
| 11 | -0.952 | 5.648 | 2.473 | 0.911 | 0.008 | 445.322 |
| 12 | -0.985 | 5.576 | 3.016 | 0.553 | 0.005 | 500.57 |
| 13 | -0.971 | 5.508 | 2.932 | 0.646 | 0.006 | 494.604 |
| 14 | -0.895 | 3.217 | 3.462 | 0.251 | 0.174 | 544.412 |
| 15 | -0.666 | 0.72 | 3.511 | -0.932 | 0.088 | 505.961 |
| 16 | -0.648 | 1.028 | 3.997 | -1.001 | 0.072 | 524.841 |
| 17 | -0.721 | 1.589 | 3.716 | -0.043 | 0.235 | 470.769 |
| 18 | -0.559 | 1.541 | 3.109 | -0.144 | -0.182 | 511.854 |
| 19 | -0.708 | 0.997 | 2.759 | -0.927 | 0.112 | 473.27 |
| 20 | -0.687 | 1.32 | 3.246 | -1.007 | 0.094 | 492.967 |
| 21 | -0.685 | 1.479 | 3.092 | 0.308 | -1.01 | 515.571 |
| 22 | -0.778 | 1.357 | 3.424 | -0.328 | 0.256 | 492.071 |
| 23 | -0.615 | 1.908 | 3.68 | -0.829 | -2.08 | 519.869 |
| 24 | -0.696 | 0.651 | 3.858 | -0.068 | -1.79 | 510.932 |
| 25 | -0.677 | 0.799 | 4.344 | -0.095 | -1.702 | 538.511 |
| 26 | -0.834 | 2.215 | 2.442 | -0.16 | -1.417 | 450.885 |
| 27 | -0.898 | 2.871 | 3.107 | 0.423 | -1.431 | 480.366 |
| 28 | -0.888 | 2.589 | 3.593 | 0.333 | -1.54 | 501.892 |
| 29 | -1.994 | 8.335 | 2.121 | -0.944 | 2.273 | 485.733 |
| 30 | -1.581 | 6.973 | 2.608 | 1.819 | 3.376 | 523.705 |
| 31 | -2.099 | 7.523 | 1.37 | -2.267 | 1.809 | 488.407 |

To better understand the external predictability of models, modified correlation coefficient (r_m^2) is determined by the following equation²².

$$r_m^2 = r^2 [1 - \sqrt{|r^2 - r_0^2|}]$$

where, r^2 is the squared correlation coefficient between observed and predicted values and r_0^2 is the squared correlation coefficient between observed and predicted values without intercept.

| | AlogP 98 | LUMO Eigen_VAMP | Dipole_Mag_VAMP | Dipole_Y | Dipole_Z | Jurs_RNCG | Jurs_TASA |
|-----------------|----------|-----------------|-----------------|----------|----------|-----------|-----------|
| AlogP 98 | 1 | | | | | | |
| LUMO Eigen_VAMP | 0.49 | 1 | | | | | |
| Dipole_Mag_VAMP | -0.49 | -0.75 | 1 | | | | |
| Dipole_Y | -0.009 | 0.10 | 0.17 | 1 | | | |
| Dipole_Z | -0.50 | -0.67 | 0.60 | 0.004 | 1 | | |
| Jurs_RNCG | -0.18 | 0.06 | -0.29 | -0.216 | -0.309 | 1 | |
| Jurs_TASA | 0.54 | 0.05 | -0.09 | -0.012 | -0.015 | -0.20 | 1 |

RESULT AND DISCUSSION

In the present study, 32 preselected descriptors from correlation matrix were correlated with training set using GFA technique. Initially, 100 QSAR equations with six descriptors were generated. For a statistically significant model, it is necessary that the descriptors evolved in the

equation should not be inter-correlated with each other, and the value was found very low in the selected models. Results of the best five models with six parameters as per the rule of 'per descriptors five compounds', which showed acceptable statistical characteristics are shown in Table 5.

Table 5: Selected QSAR equation and their regression statistics

| Equation No. | Equations |
|--------------|--|
| 1. | $= 5.9926 + 0.0854929 * \text{Dipole_Z} + 0.0172095 * \text{AlogP 98} * \text{AlogP 98} + 0.101165 * \langle \text{Dipole_Mag_VAMP} - 1.90968 \rangle - 0.0136121 * \langle \text{Jurs_TASA} - 531.927 \rangle + 0.121374 * \langle 1.28239 - \text{Dipole_Y} \rangle - 0.691162 * \langle -0.925744 - \text{LUMO_Eigenvalue_VAMP} \rangle$ |
| 2. | $= 5.8776 + 0.0923629 * \text{AlogP 98} + 0.0815622 * \text{Dipole_Z} + 0.0988983 * \langle \text{Dipole_Mag_VAMP} - 1.95464 \rangle - 0.0128245 * \langle \text{Jurs_TASA} - 532.858 \rangle + 0.122882 * \langle 1.28239 - \text{Dipole_Y} \rangle - 0.662946 * \langle -0.925744 - \text{LUMO_Eigenvalue_VAMP} \rangle$ |
| Equation No. | Equations |
| 3. | $= 5.98837 + 0.11172 * \text{AlogP 98} + 0.0935328 * \text{Dipole_Z} - 3.91047 * \text{Jurs_RPCG} * \text{Jurs_RPCG} + 0.0843938 * \langle \text{Dipole_Mag_VAMP} - 1.90968 \rangle - 0.0123058 * \langle \text{Jurs_TASA} - 529.42 \rangle + 0.111195 * \langle 0.850262 - \text{Dipole_Y} \rangle$ |
| 4. | $= 6.18475 + 0.125322 * \text{AlogP 98} + 0.103095 * \text{Dipole_Z} - 2.2012 * \text{Jurs_RPCG} + 0.0760798 * \langle \text{Dipole_Mag_VAMP} - 1.46901 \rangle - 0.0126567 * \langle \text{Jurs_TASA} - 528.754 \rangle + 0.108086 * \langle 0.925883 - \text{Dipole_Y} \rangle$ |
| 5. | $= 6.06119 + 0.0915044 * \text{Dipole_Z} + 0.0182142 * \text{AlogP 98} * \text{AlogP 98} + 0.0859676 * \langle \text{Dipole_Mag_VAMP} - 1.90968 \rangle - 5.11642 * \langle \text{Jurs_RPCG} - 0.299467 \rangle - 0.012167 * \langle \text{Jurs_TASA} - 529.926 \rangle + 0.115156 * \langle 0.744643 - \text{Dipole_Y} \rangle$ |

Table 6: Statistical parameters of the generated equations

| Equation No. | LOF | r ² | r ² adj | r ² cv | F Value |
|--------------|-------|----------------|--------------------|-------------------|---------|
| 1 | 0.030 | 0.771 | 0.714 | 0.611 | 13.53 |
| 2 | 0.031 | 0.759 | 0.699 | 0.588 | 12.62 |
| 3 | 0.032 | 0.756 | 0.695 | 0.560 | 12.40 |
| 4 | 0.032 | 0.755 | 0.693 | 0.554 | 12.33 |
| 5 | 0.032 | 0.753 | 0.692 | 0.575 | 12.25 |

Table 7: Summary of external validation parameters

| Equation no. | r^2_{pred} | r^2_{m} |
|--------------|---------------------|------------------|
| 1 | 0.726 | 0.561 |
| 2 | 0.712 | 0.575 |
| 3 | 0.714 | 0.511 |
| 4 | 0.706 | 0.510 |
| 5 | 0.714 | 0.508 |

To check the inter-correlation of descriptors, variance inflation factor analysis was performed. The VIF values of descriptors of best models included in the QSAR study was found to be 1.64 (Dipole_Z), 1.37 (AlogP 98), 2.38 (Dipole_Mag_VAMP), 5.75 (Jurs_TASA), 5.04 (LUMO_Eigenvalue_VAMP) and 2.12 (Dipole_Y). All the VIF values were found to be less than 10, indicates low

inter-correlation of the descriptors used in the selected models. The molecular descriptors selected after 5000 generations performed by GFA, rejecting highly correlated descriptors, used to generate QSAR models, and affected the biological activity are shown in the Histogram (Figure 1).

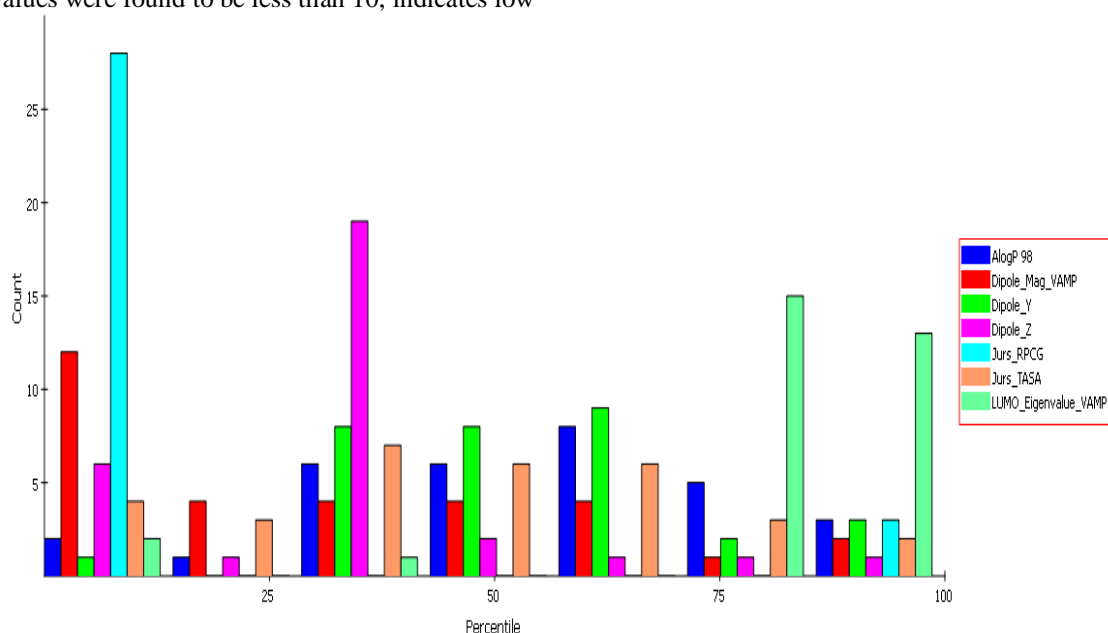


Figure 1: Histogram of molecular descriptors used in generation of models.

The models were also evaluated for their predictive power, i.e. internal and external cross-validation. Model 1 is considered to be the best based on both validation parameters. The results for the Equation 1 are summarized in Table 6 & 7. The regression correlation coefficient (r^2) of 0.77 shows good correlation with biological activity, cross-validation coefficient (r^2_{cv}) of 0.61 indicates good internal predictability, predicted correlation coefficient (r^2_{pred}) of 0.73 and mean correlation coefficient (r^2_{m}) value of 0.56 indicates reliability and significance of the model as external validity parameters. Each equation is assessed by Friedman's lack of fit (LOF) and F-value.

Values of observed activity and predicted activity of training set and test set compounds was found to be very close, as evident from the values of residual (Table 8 and 9), which indicates robustness of models and also the power to predict the activity of related compounds. Figure 2 and 3 depicts the plot of observed vs predicted activity for training and test set compounds, respectively as per equation 1.

QSAR study of the thienopyrimidine series showed thermodynamic, electronic and geometric descriptors, and were found to have significant influence on the inhibition activity of H_1 -receptor antagonists as exemplified from

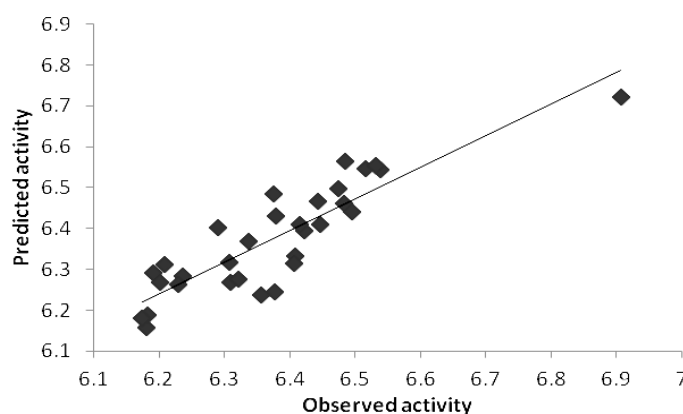
high value of their coefficients. AlogP 98 is thermodynamic descriptor which computes the lipophilicity of the molecule. All models (Table 5) shows that lipophilicity has positive contribution towards biological activity. Jurs descriptors are a group of molecular descriptors which combine shape and electronic information to characterize molecules. These descriptors are calculated by mapping atomic partial charges on solvent-accessible surface areas of individual atoms. Jurs_TASA is total atomic surface area, calculated by total atomic charged surface areas divided by the total molecular solvent-accessible surface area. A critical analysis of all equations showed negative contribution of these descriptors on biological activity. This means that the charge distribution within the molecules acts as the driving force for intermolecular interactions and the lesser the relative charge the larger the interactions. This suggests that molecules with less bulkier and more lipophilic substituents are more likely to show activity. The above fact is exemplified from compounds, **2** ($R = -OCH_3$), **6** ($R = -CH_3$), **7** ($R = -Br$) and **8** ($R = -Cl$) where lesser values of the Jurs descriptors resulted in increase in activity. Notably, the most potent compound **2** has a very less value of this descriptor signifying the effect of these geometric descriptors on biological activity.

Table 8: Observed and predicted pIC₅₀ values of training set compounds (as per equation 1)

| Compound No. | Observed Activity | Predicted Activity | Residuals |
|--------------|-------------------|--------------------|-----------|
| 1 | 6.423 | 6.395 | -0.028 |
| 2 | 6.907 | 6.721 | -0.186 |
| 3 | 6.308 | 6.316 | 0.008 |
| 4 | 6.538 | 6.545 | 0.007 |
| 5 | 6.18 | 6.157 | -0.023 |
| 6 | 6.31 | 6.268 | -0.042 |
| 7 | 6.321 | 6.277 | -0.044 |
| 8 | 6.379 | 6.43 | 0.051 |
| 9 | 6.408 | 6.333 | -0.075 |
| 10 | 6.236 | 6.283 | 0.047 |
| 11 | 6.474 | 6.497 | 0.023 |
| 12 | 6.485 | 6.565 | 0.08 |
| 13 | 6.516 | 6.547 | 0.031 |
| 14 | 6.192 | 6.291 | 0.099 |
| 15 | 6.376 | 6.484 | 0.108 |
| 16 | 6.532 | 6.555 | 0.023 |
| 17 | 6.447 | 6.409 | -0.038 |
| 18 | 6.407 | 6.315 | -0.092 |
| 19 | 6.291 | 6.402 | 0.111 |
| 20 | 6.483 | 6.462 | -0.021 |
| 21 | 6.182 | 6.187 | 0.005 |
| 22 | 6.416 | 6.411 | -0.005 |
| 23 | 6.208 | 6.312 | 0.104 |
| 24 | 6.229 | 6.263 | 0.034 |
| 25 | 6.377 | 6.245 | -0.127 |
| 26 | 6.174 | 6.18 | 0.006 |
| 27 | 6.357 | 6.237 | -0.12 |
| 28 | 6.201 | 6.267 | 0.066 |
| 29 | 6.495 | 6.441 | -0.054 |
| 30 | 6.444 | 6.466 | 0.022 |
| 31 | 6.337 | 6.368 | 0.031 |

Table 9: Observed and predicted pIC₅₀ values of test set compounds (as per equation 1)

| Compound No. | Observed activity | Predicted activity | Residuals |
|--------------|-------------------|--------------------|-----------|
| 1 | 6.679 | 6.514 | 0.165 |
| 2 | 6.600 | 6.644 | -0.043 |
| 3 | 6.673 | 6.624 | 0.049 |
| 4 | 6.586 | 6.424 | 0.162 |
| 5 | 5.744 | 6.027 | -0.282 |
| 6 | 6.102 | 6.241 | -0.138 |
| 7 | 5.978 | 6.281 | -0.302 |

**Figure 2:** Plot of observed Vs predicted pIC₅₀ values of training set compounds (as per equation 1)

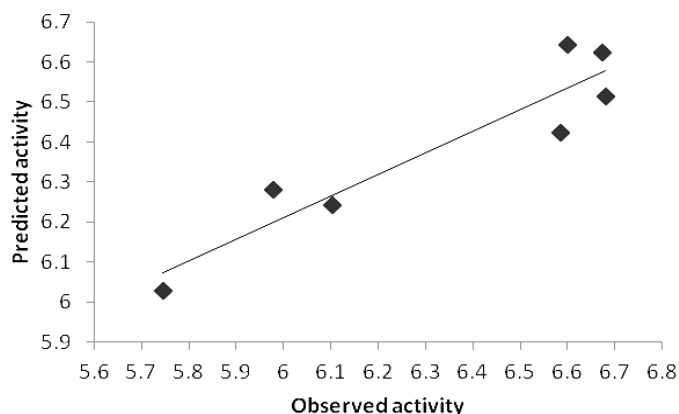


Figure 3: Plot of observed Vs predicted pIC_{50} values of test set compounds (as per equation 1)

The dipole moment descriptor is a 3D electronic descriptor that indicates the strength and orientation behavior of a molecule in an electrostatic field. Dipole_Y and Dipole_Z electronic descriptor which represents dipole moment of molecule in Y and Z dimension respectively, positive values of its coefficient indicates that as dipole moment increase in this Y and Z dimension respectively, activity increases. Moreover, its higher coefficient value was found among all descriptors. This indicates that dipole moment is the important descriptors that significantly influence the H_1 -Histaminic inhibitory activity.

Another set of descriptors, quantum mechanical viz., LUMO_Eigenvalue_VAMP and Dipole_Mag_VAMP, describes the energy of the highest and lowest occupied molecular orbitals showed significant contribution in biological activity being having high value of their coefficients.

In consistent with the above correlation, the 3D Point plot shown in Figure 4, describes that the less bulky and lipophilic-electronic substituents are likely to increase the biological activity of the similar molecules and as expected were found to be potent H_1 -Histaminic inhibitors.

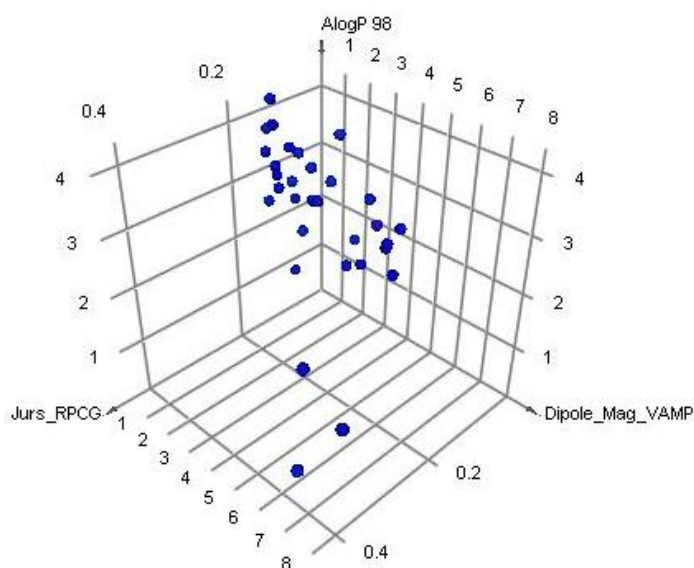


Figure 4: 3D Point plot of most influencing molecular descriptors used in generation of QSAR models.

CONCLUSION

The established QSAR models were found to be statistically significant as evidenced from their regression statistics that shows significant correlative and predictive ability in terms of good q^2 and r^2 values. Very low residuals were obtained in internal and external validation methods suggest that developed models are predictive. Moreover, the good r^2_{pred} and r^2_m values for an external test set, confirms the excellent predictive ability of the established QSAR model. Thermodynamic, electronic and geometric descriptors were found to be important descriptors which described H_1 -Histaminic inhibitory property of thieno[2,3-*d*]pyrimidin-4(3H)-ones series. Results indicated that potency can be

enhanced by emphasizing on more lipophilic, less bulkier substituents as well as on electrostatic potential of compounds as it is evident from their physicochemical properties. These results may provide valuable guidance for improving the biological activity of the analogs and continuing search for potent H_1 -Histaminic inhibitors prior to synthesis.

ACKNOWLEDGMENT

One of the authors gratefully acknowledge technical support, (for providing Discovery studio 2.1 Accelrys Inc. USA, molecular modeling software facilities), from L.M. College of Pharmacy, Gujarat University, Ahmedabad, Gujarat, India.

REFERENCES

1. Dale WA, Vogt W, Synthese des Imidazolyl-athylamins, *Chemische Berichte*, 1907, 40, 3691.
2. Matsumoto T, Nakamura M, Inhibition of calcium transients in cultured vascular smooth muscle cells by pertussis toxin, *Biochemical Biophysical Research Communication*, 1986, 140, 195-203.
3. Killackey JJ, Johnston MG, Movat HZ, Increased permeability of microcarrier-cultured endothelial monolayers in response to histamine and thrombin: A model for the in vitro study of increased vasopermeability. *American Journal of Pathology*, 1986, 122, 50-61.
4. Hill SJ, Ganellin CR, Timmerman H, Schwartz JC, Shankley NP, Young JM, Schunack W, Levi R, Haas HL, International Union of Pharmacology. XIII. Classification of histamine receptors, *Pharmacology Review*, 1997, 49(3), 253-278.
5. Barnes PJ, Chung KF, Page CP, Inflammatory mediators of asthma: An update. *Pharmacology Review*, 1988, 50, 515-596.
6. Jutel M, Watanabe T, Klunker S, Akdis M, Thomet OA, Malolepszy J, Zak-Nejmark T, Koga R, Kobayashi T, Blaser K, Akdis CA, Histamine regulates T-cell and antibody responses by differential expression of H₁ and H₂ receptors, *Nature*, 2001, 413, 420-425.
7. Billington CK, Penn RB, Signaling and regulation of G protein-coupled receptors in airway smooth muscle, *Respiratory Research*, 2003, 4, 2.
8. El-Enany MM, El-Shafie FS, Thienopyrimidines: synthesis, properties and biological properties, *Oriental Journal of Chemistry*, 1988, 5, 114-119.
9. Lin L, Lin WO, Jiang JH, Shen GL, Yu RQ, QSAR analysis of substituted bis[(acridine-4-carboxamide)propyl]methylamines using optimized block-wise variable combination by particle swarm optimization for partial least squares modeling, *European Journal of Pharmaceutical Science*, 2005, 25, 245-254.
10. Si HZ, Wang T, Zhang K, Hua ZD, Fan BT, QSAR study of 1,4-dihydropyridine calcium channel antagonists based on gene expression programming, *Journal of Bioorganic Medicinal Chemistry*, 2006, 14, 4834-4841.
11. Sivakumar PM, Geethababu SK, Doble M, Impact of topological and electronic descriptors in the QSAR of pyrazine containing thiazolines and thiazolidinones as antitubercular and antibacterial agents, *Chemical Biology Drug Research*, 2008, 71, 447-463.
12. Lü JX, Shen Q, Jiang JH, Shen GL, Yu RQ, QSAR analysis of cyclooxygenase inhibitor using particle swarm optimization and multiple linear regression, *Journal of Pharmaceutical Biomedical Analysis*, 2004, 35, 679-687.
13. Leonard JT, Roy K, On Selection of Training and Test Sets for the Development of Predictive QSAR models, *QSAR Combinatorial Science*, 2006, 25(3), 235-251.
14. Kulkarni VM, Karki R, Three-dimensional quantitative structure-activity relationship (3D-QSAR) of 3-aryloxazolidin-2-one antibacterials, *Journal of Biorganic Medicinal Chemistry*, 2001, 9, 3153-3160.
15. Roy K, Bhattacharya P, QSAR of adenosine A₃ receptor antagonist 1,2,4-triazolo[4,3-*a*]quinoxalin-1-one derivatives using chemometric tools, *Biorganic Medicinal Chemistry Letters*, 2005, 15, 3737-3743.
16. Shirsath VS, Design, Synthesis, and Biological activity of some Novel Heterocycles, PhD Thesis, The Gujarat University: Ahmedabad, India, 2000.
17. Discovery Studio 2.1. Accelrys Inc., 6985 Scranton Road, San Diego, CA, USA.
18. Rogers D, Hopfinger AJ, Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships, *Journal of Chemical Information and Computer Science*, 1994, 34, 854-866.
19. Cho DH, Lee SK, Kim BT, No KT, Quantitative Structure-Activity Relationship (QSAR) Study of New Fluorovinylxyacetamides, *Bull Korean Chemical Society*, 2001, 22, 388-394.
20. Jaiswal M, Khadikar PV, Scozzafava A, Supuran CT, Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides, *Bioorganic Medical Chemistry Letters*, 2004, 14, 3283-3290.
21. Kulkarni S, Kulkarni VM, Three-Dimensional Quantitative Structure-Activity Relationship of Interleukin 1- β Converting Enzyme Inhibitors: A Comparative Molecular Field Analysis Study, *Journal of Medicinal Chemistry*, 1999, 42, 373.
22. Roy P, Roy K, On Some Aspects of Variable Selection for Partial Least Squares Regression Models, *QSAR Combinatorial Science*, 2008, 27(3), 302-313.