informa
healthcare

ORIGINAL ARTICLE

A QSAR study on 2-(4-methylpiperazin-1-yl)quinoxalines as human histamine H$_4$ receptor ligands

Brij K. Sharma[1], Pradeep Pilania[1], Prithvi Singh[1], and Yenamandra S. Prabhakar[2]

[1]*Department of Chemistry, S. K. Government College, Sikar 332 001, Rajasthan, India, and* [2]*Medicinal and Process Chemistry Division, Central Drug Research Institute, CSRI, Lucknow 226 001, Uttar Pradesh, India*

Abstract
The histamine H$_4$ receptor binding affinity of 2-(4-methylpiperazin-1-yl)quinoxaline derivatives has been quantitatively analyzed in terms of Dragon descriptors. The derived QSAR models have provided rationales to explain the activity of titled derivatives. The descriptors identified in CP-MLR analysis have highlighted the role of path/walk 4-Randic shape index (PW4), mean square distance (MSD) index, topological charges (GGI9, JGI2, and JGI7), atomic properties in respective lags of 2D-autocorrelations (MATS7e, GATS7e, and MATS8p), and Burden matrix (BELm1) to explain the binding affinity. Certain structural fragments (C-002 and C-027) have also shown prevalence to optimize the H$_4$R binding affinity of titled compounds. The PLS analysis has also confirmed the dominance of information content of CP-MLR-identified descriptors for modelling the activity.

Keywords: QSAR, 2-(4-methylpiperazin-1-yl)quinoxalines, histamine H$_4$ receptor (H$_4$R), combinatorial protocol in multiple linear regression (CP-MLR), partial least square (PLS) analysis

Introduction

Histamine plays a role in many physiological processes such as inflammation and vasodilatation, gastric acid secretion, cognitive processes, regulation of food intake, sleep, and wakefulness. It is a member of G-protein-coupled receptors (GPCRs) family. At present, there are four histamine receptor subtypes, namely, H$_1$R, H$_2$R, H$_3$R, and H$_4$R. The antagonists of H$_1$ and H$_2$ receptors are widely used for the treatment of allergic disorders such as hay fever[1] and gastric ulcers,[2] respectively. The H$_3$R antagonists are being assessed for clinical efficiency in attention-deficit hyperactivity disorder (ADHD), dementia, and narcolepsy.[3] The most recently identified histamine receptor subtype[4–8] H$_4$R has a diverse pharmacological profile than H$_1$, H$_2$, and H$_3$R.[9] It is primarily expressed in haematopoietic and immune cells such as eosinophils, mast cells, and macrophages, as well as in peripheral tissues such as spleen, thymus, and bone marrow,[10] and plays a role in immunological and inflammatory processes. For the treatment of various chronic inflammatory diseases, such as inflammatory bowel disease, asthma, and rheumatoid arthritis,[11–13] and possible role in the proliferation of colon carcinoma cells, in the modulation of angiogenesis and in mediating pruritis,[14,15] the H$_4$R is being considered as a potential drug target.

Recently, a series of *N*-methylpiperazinylquinoxaline derivatives, as a novel class of potent H$_4$R ligands, has been reported by de Esch and colleagues.[16] Apart from this, Kiss et al.[17] have reported structure-based virtual screening of >8.7 million diverse chemical entities from different database using a ligand-supported homology model of the human histamine H$_4$ receptor (hH$_4$R) and identified several novel scaffolds as selective H$_4$ ligands. Several of these hits shared aryl, heteroaryl, and functionalized unsaturated chains as part of their structures. These results furthermore highlight the *N*-methylpiperazinylquinoxalines as potential H$_4$R ligands. Moreover, in view of the importance of anti-inflammatory agents in the clinical management of several disorders, a quantitative structure–activity relationship is attempted on the binding affinity of these

RIGHTS LINK()

quinoxaline derivatives. The present study is aimed at rationalizing the substituent variations of these analogues to provide insight for the future endeavours.

Materials and methods

Dataset and molecular descriptors

The 45 reported *N*-methylpiperazinylquinoxaline derivatives [16] have been considered as the dataset for present QSAR study. The binding affinity of these compounds were measured by displacement of [³H]histamine binding using membranes of HEK cells transiently expressing the human H$_4$R. The structural variations and binding affinity (as p*K$_i$*) are mentioned in Table 1. The descriptors, accounting molecular features, for all the reported compounds have been calculated from Dragon software.[18] The input for this software was the energy-minimized 3D structures of the compounds drawn in ChemDraw software.[19] A total number of 458 descriptors corresponding to 0D-, 1D-, and 2D-classes of Dragon software have been considered in the present modelling study. Before model development procedure, all those descriptors that are intercorrelated beyond 0.90 and showing a correlation of <0.1 with the biological endpoints (descriptor versus activity, *r* < 0.1) were excluded. This procedure has reduced the total descriptors from 458 to 92 as relevant ones to explain the biological actions of titled compounds. In the dataset, the initial assessment of activity with the all descriptors has suggested the compound **4** as potential outlier. An outlier to a QSAR can indicate the limits of applicability of QSAR models.

In these derivatives, two-ring heterocyclic scaffolds are directly connected to the *N*-methylpiperazine moiety. These analogues were synthesized on the basis of flexible alignment model of VUF6884 (a tricyclic clozapine analogue) and JNJ7777120 (an indole-*N*-methylpiperazine derivative).[20] The model suggested that both the compounds had overlapping binding mode for the H$_4$ receptor. The reported quinoxaline derivatives are having the hybrid scaffold based on the structures of clozapine analogue and indole anologue. The alignment model suggested that substitution of quinoxaline with an additional aromatic ring system could occupy the aromatic pocket, which is also occupied by one of the aromatic rings of a tricyclic clozapine analogue when it binds to the H$_4$R. Compound **4** may be treated as an outlier as it may not fit to an aromatic pocket and/or may furnish an additional interaction with an aromatic pocket. In view of this, the modelling study has been carried out without this analogue. Inclusion of this analogue to the dataset leads to statistically insignificant results.

Development and validation of model

The combinatorial protocol in multiple linear regression (CP-MLR)[21–25] and partial least-squares (PLS)[26–28] procedures have been used in the present study for developing QSAR models. The CP-MLR is a "filter"-based variable selection procedure, which employs a combinatorial strategy with MLR to result in selected subset regressions for the extraction of diverse structure–activity models, each having unique combination of descriptors from the generated dataset of the compounds under study. The embedded filters make the variable selection process efficient and lead to unique solution. Fear of "chance correlations" exists where large descriptor pools are used in multilinear QSAR/QSPR studies.[29,30] In view of this, to find out any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been subjected to randomization test[31,32] by repeated randomization (100 simulation runs) of the biological responses. The datasets with randomized response vector have been reassessed by multiple regression analysis. The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to unscrambled response data were counted. This has been used as a measure to express the percent chance correlation of the model under scrutiny.

Validation of the derived model is necessary to test its prediction and generalization within the study domain. The dataset of the present study is randomly divided into training set for model development and test set for external prediction or validation. The compounds of test set were randomly picked using an in-house written randomization program.

For each model, derived by involving *n* data points, a number of statistical parameters such as *r* (the multiple correlation coefficient), *s* (the standard deviation), *F* (the *F* ratio between the variances of calculated and observed activities), and Q^2_{LOO} (the cross-validated index from leave-one-out procedure) have been obtained to access its overall statistical significance. In case of internal validation, Q^2_{LOO} is used as a criterion of both robustness and predictive ability of the model. A value greater than 0.5 of Q^2 index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5.
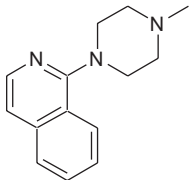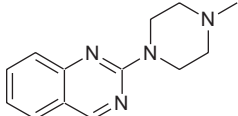
Applicability domain

The usefulness of a model is based on its accurate prediction ability for new congeners. A model is valid only within its training domain and new compounds must be assessed as belonging to the domain before the model is applied. The applicability domain (AD) is evaluated by the leverage values for each compound.[33] A Williams plot (the plot of standardized residuals versus leverage values (*h*)) is constructed, which can be used for a simple graphical detection of both the response outliers (*Y* outliers) and structurally influential chemicals (*X* outliers) in the model. In this plot, the AD is established inside a squared area within ±*x* standard deviations and a leverage threshold *h**, which is generally fixed at $3(k + 1)/n$ (*n* is the number of training set compounds

Table 1. Structures and observed experimental binding affinity of *N*-methylpiperazinylquinoxaline derivatives.[a]

Cpd.	R_1	R_2	pK_i

Cpd.	R_1	R_2	pK_i
1[b]	H	H	6.05
2	H	CH_3	6.70
3	H	C_6H_5	4.99
4[c]	H	$CH_2C_6H_5$	—
5[b]	H	SPh	6.44
6	H	$NHCH_2Ph$	5.13
7	H	OPh	6.49
8	H	OCH_2Ph	6.53
9	H	$O(CH_2)_2Ph$	6.32
10[b]	H	$O(CH_2)_4Ph$	5.53
11	H	OCH_2-2-pyridyl	6.33
12	H	OCH_2-3-pyridyl	5.53
13	H	OCH_2-4-pyridyl	5.36
14	H	OCH_2-(4-OCH_3-Ph)	6.15
15	H	OCH_2-(4-CH_3-Ph)	5.86
16	H	OCH_2-(4-Cl-Ph)	5.64
17[b]	H	OCH_2-(3-Cl-Ph)	6.57
18[b]	H	O(3-pyridyl)	5.89
19[b]	H	O(4-Cl-Ph)	5.63
20[b]	H	O(3,4-Cl-Ph)	5.77
21	H	O(4-F-Ph)	5.80
22	H	O(3-CH_3-Ph)	6.24
23	H	O(4-CH_3-Ph)	5.66
24	H	O-(4-OCH_3-Ph)	5.63
25	H	O-(3-*N*,*N*-dimethylaniline)	5.81
26	H	O-cyclohexyl	4.88
27	H	$OCH_2CH(CH_3)_2$	5.24
28	H	OCH_2CH_3	6.64
29	H	$O(CH_2)_3$-*N*,*N*-dimethylamine	5.40
30	H	OCH_3	6.47
31	H	OH	7.21
32	6-Cl	OCH_3	7.58
33[b]	6-Cl	OH	7.93
34	6,7-di-Cl	OH	8.25
35[b]	6,7-di-Cl	$(CH_2)_2Ph$	5.40
36	6,7-di-Cl	CH_3	7.20
37[b]	6,7-di-Cl	OPh	5.93
38[b]	6,7-di-Cl	OCH_3	7.24
39	H	CF_3	5.60
40[b]	H	Cl	6.64
41[b]	6-Cl	H	7.04
42			5.16

| 43 | | | 6.23 |

Table 1. continued on next page

Table 1. Continued.

Cpd.	R$_1$	R$_2$	pK_i
44			4.69
45[b]			5.12

[a]Taken from reference [16].
[b]Compounds in test set.
[c]Not part of the dataset.

and k is the number of model parameters), whereas $x = 2$ or 3. If the compounds have a high leverage value ($h > h^*$), then the prediction is not trustworthy. On the other hand, when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training set compounds.

Results and discussion

QSAR results

In multidescriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. For the purpose of modelling study, 14 compounds have been included in the test set for the validation of the models derived from 30 training set compounds. A total number of 92 significant descriptors from 0D-, 1D-, and 2D-classes have been subjected to CP-MLR analysis with default "filters" set in it. Statistical models in two and three descriptor(s) have been derived successively to achieve the best relationship correlating H$_4$R binding affinity. A total number of 6 and 12 models in two and three descriptors, respectively, were obtained through CP-MLR. These models (with 92 descriptors) were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this, the optimum r-bar value of the preceding level model has been used as the new threshold of filter-3 for the next generation. The selected models in two and three descriptors are given below.

$$p K_i = 69.455 - 30.940(4.665)\text{BELm1} - 5.352(1.382)\text{GGI9} \quad (1)$$

$n = 30$, $r = 0.804$, $s = 0.513$, $F = 24.639$, $Q^2_{\text{LOO}} = 0.574$, $Q^2_{\text{L5O}} = 0.532$, $r^2_{\text{randY}}(\text{sd}) = 0.255(0.148)$, $r^2_{\text{Test}} = 0.607$

$$p K_i = 59.179 + 48.132(19.392)\text{PW4} - 30.416(5.330)\text{BELm1} \quad (2)$$

$n = 30$, $r = 0.743$, $s = 0.578$, $F = 16.616$, $Q^2_{\text{LOO}} = 0.443$, $Q^2_{\text{L5O}} = 0.361$, $r^2_{\text{randY}}(\text{sd}) = 0.241(0.126)$, $r^2_{\text{Test}} = 0.513$

$$p K_i = 61.334 - 26.893(4.349)\text{BELm1} - 4.723(1.240)\text{GGI9} + 2.534(0.866)\text{MATS7e} \quad (3)$$

$n = 30$, $r = 0.857$, $s = 0.454$, $F = 23.883$, $Q^2_{\text{LOO}} = 0.674$, $Q^2_{\text{L5O}} = 0.677$, $r^2_{\text{randY}}(\text{sd}) = 0.314(0.141)$, $r^2_{\text{Test}} = 0.559$

$$p K_i = 69.945 - 31.162(4.237)\text{BELm1} - 5.207(1.256)\text{GGI9} - 0.241(0.093)\text{C-002} \quad (4)$$

$n = 30$, $r = 0.848$, $s = 0.466$, $F = 22.175$, $Q^2_{\text{LOO}} = 0.644$, $Q^2_{\text{L5O}} = 0.580$, $r^2_{\text{randY}}(\text{sd}) = 0.282(0.108)$, $r^2_{\text{Test}} = 0.589$

In above and all follow-up regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The $r^2_{\text{randY}}(\text{sd})$ is the mean random-squared multiple correlation coefficient of the regressions in the activity (Y) randomization study with its standard deviation from 100 simulations. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models.

The descriptor BELm1 belongs to BCUT class of Dragon descriptors. The BCUT descriptors are the first eight highest and the lowest absolute eigenvalues, BEHwk and BELwk, respectively, for the modified Burden adjacency matrix. Here, w refers to the atomic property and k to the eigenvalue rank. The ordered sequence of the highest and the lowest eigenvalues reflect upon the relevant aspects of molecular structure, useful for similarity searching. The negative contribution of descriptor BELm1 to the activity advocates that a higher value of this descriptor is detrimental to the activity.

The other participated descriptors, GGI9, PW4, MATS7e, and C-002, belong to Galvez topological charge indices (GALVEZ), topological (TOPO), 2D-autocorrelations (2D-AUTO), and atom-centred fragment (ACF) classes of Dragon descriptors, respectively. GALVEZ descriptors are the Galvez topological charge indices and have their origin in the first 10 eigenvalues of the polynomial of corrected adjacency matrix of the compounds. All the GALVEZ class descriptors belong to two categories. Of this, one category corresponds to the topological

charge index of order n (GGIn) and the other to the mean topological charge index of order n (JGIn), where "n" represents the order of eigen value. The negative influence of descriptor GGI9 (topological charge index of ninth order) from this class to the activity suggested that a lower value of ninth order charge index would be beneficiary to the activity. The TOPO class descriptors are based on a graphical representation of the molecule and are numerical quantifiers of molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labelling. They can be sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching, and cyclicity and can also encode chemical information concerning atom type and bond multiplicity. The descriptor PW4, participated in above models, is representative of TOPO class. The descriptor PW4, path/walk ratio 4, is the Randic's molecular shape descriptor and its value increases with increased branching in the vertices. The sign of regression coefficient of this descriptor shows positive influence on the activity. A higher value of this descriptor would be in favour of the activity.

The 2D-autocorrelations are molecular descriptors, which describe how a considered property is distributed along a topological molecular structure. The 2D-AUTO descriptors have their origin in autocorrelation of topological structure of Broto-Moreau (ATS), of Moran (MATS) and of Geary (GATS). The computation of these descriptors involves the summations of different autocorrelation functions corresponding to the different fragment lengths and leads to different autocorrelation vectors corresponding to the lengths of the structural fragments. Also a weighing component in terms of a physicochemical property has been embedded in this descriptor. As a result, these descriptors address the topology of the structure or parts thereof in association with a selected physicochemical property. In these descriptors' nomenclature, the penultimate character, a number, indicates the number of consecutively connected edges considered in its computation and is called as the autocorrelation vector of lag n (corresponding to the number of edges in the unit fragment). The very last character of the descriptor's nomenclature indicates the physicochemical property considered in the weighing component—m for mass or v for volume or e for Sanderson electronegativity or p for polarizability—for its computation. It is evident from the sign of regression coefficient of the participating descriptor that descriptor MATS7e from this class has contributed positively to the activity. Thus a higher positive value of descriptor MATS7e (Moran autocorrelation of lag 7 weighed by atomic Sanderson electronegativities) will be in favour of activity. ACF descriptors are simple molecular descriptors defined as the number of specific atom types in a molecule and their calculation is based on the knowledge of the molecular composition and atom connectivities. The ACF class descriptor C-002, representing a CH_2R_2 fragment in a molecular structure,

has shown negative influence on the activity suggesting absence of such type of fragment for improved activity.

The three descriptor models could estimate nearly 72% in observed activity of the compounds. Considering the number of observation in the dataset, models with up to four descriptors were explored. It has resulted in 19 four-parameter models with test set $r^2 > 0.50$. These models have shared 18 descriptors among them. All these 18 descriptors along with their brief meaning, average regression coefficients, and total incidence are listed in Table 2, which will serve as a measure of their estimate across these models. Following are some four-descriptor models for the activity.

$$pK_i = 58.866 - 25.630(3.918)BELm1 - 4.752(1.110)GGI9 + 2.816(0.781)MATS7e - 0.379(0.139)C\text{-}027 \tag{5}$$

$$n = 30, \quad r = 0.891, \quad s = 0.406, \quad F = 24.255, \quad Q^2_{LOO} = 0.748,$$
$$Q^2_{L5O} = 0.744, r^2_{randY}(sd) = 0.350(0.095), r^2_{Test} = 0.539$$

$$pK_i = 68.543 - 30.138(3.763)BELm1 - 5.358(1.112)GGI9 - 0.804(0.280)GATS7e - 0.232(0.082)C\text{-}002 \tag{6}$$

$$n = 30, \quad r = 0.888, \quad s = 0.412, \quad F = 23.337, \quad Q^2_{LOO} = 0.738,$$
$$Q^2_{L5O} = 0.738, r^2_{randY}(sd) = 0.335(0.125), r^2_{Test} = 0.531$$

$$pK_i = 58.674 + 25.525(6.462)MSD + 55.605(14.318)PW4 - 33.799(3.979)BELm1 - 0.216(0.085)C\text{-}002 \tag{7}$$

$$n = 30, \quad r = 0.882, \quad s = 0.423, \quad F = 21.795, \quad Q^2_{LOO} = 0.598,$$
$$Q^2_{L5O} = 0.632, r^2_{randY}(sd) = 0.375(0.126), r^2_{Test} = 0.558$$

$$pK_i = 61.502 - 27.830(4.002)BELm1 + 20.729(5.469)JGI2 - 94.202(25.592)JGI7 + 2.324(0.771)MATS8p \tag{8}$$

$$n = 30, \quad r = 0.881, \quad s = 0.424, \quad F = 21.629, \quad Q^2_{LOO} = 0.658,$$
$$Q^2_{L5O} = 0.665, r^2_{randY}(sd) = 0.353(0.125), r^2_{Test} = 0.543$$

These models have accounted for nearly 77% variance in the observed activities. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The values greater than 0.5 of Q^2 index is in accordance to a reasonable robust QSAR model. The pK_i values of training set compounds calculated using Eqs. (5)–(8) and predicted from LOO procedure have been included in Table 3. The models (5)–(8) are validated with an external test set of 14 compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values and the same is reported in Table 4. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

The newly appeared descriptors in above models are C-027 (an ACF class descriptor), GATS7e and MATS8p (2D-AUTO descriptors), MSD (TOPO class descriptor), JGI2 and JGI7 (both from GALVEZ class). Above equations reveal that a higher value of Balaban mean square distance (MSD) index, second order mean topological charge index (JGI2), and Moran autocorrelation of lag 8 weighed by atomic polarizabilities (MATS8p) are

Table 2. Physical meaning, average regression coefficients and the total incidences, and MLR-like coefficients from PLS model of descriptors identified from four-parameter CP-MLR models for the binding affinity of *N*-methylpiperazinylquinoxaline derivatives.

S. No.	Class	Descriptors' symbol, and meaning	Avg. reg. coeff.(incidence)[a]	MLR-like coeff.(fc)order[b]
1	CONST	nF, number of fluorine atoms	−0.303(2)	−0.065(−0.027)10
2	TOPO	MSD, mean square distance index (Balaban)	27.145(4)	−0.031(−0.013)12
3		X2A, average connectivity index chi-2	−54.948(1)	0.004(0.002)17
4		PW4, path/walk 4-Randic shape index	57.145(2)	0.008(0.004)16
5		PW5, path/walk 5-Randic shape index	78.401(2)	0.0217(0.009)13
6		CIC5, complementary information content of fifth order neighbourhood symmetry	−2.385(1)	−0.252(−0.106)4
7	BCUT	BELm1, lowest eigen value n.1 of Burden matrix weighed by atomic masses	−30.081(19)	−0.767(−0.323)1
8	GALVEZ	GGI9, ninth order Galvez topological charge index	−4.997(13)	−0.256(−0.108)3
9		JGI2, second order mean topological charge index	17.523(2)	0.127(0.054)7
10		JGI7, seventh order mean topological charge index	−84.510(2)	−0.161(−0.068)5
11	2D-AUTO	MATS4v, Moran autocorrelation of lag 4 weighed by atomic van der Waals volumes	−3.763(2)	−0.085(−0.036)9
12		MATS6e, Moran autocorrelation of lag 6 weighed by atomic Sanderson electronegativities	1.763(1)	−0.004(−0.002)18
13		MATS7e, Moran autocorrelation of lag 7 weighed by atomic Sanderson electronegativities	2.502(2)	0.091(0.038)8
14		GATS7e, Geary autocorrelation of lag 7 weighed by atomic Sanderson electronegativities	2.324(1)	−0.015(−0.006)14
15		MATS8p, Moran autocorrelation of lag 8 weighed by atomic polarizabilities	−0.790(4)	−0.060(−0.025)11
16	ACF	C-002, CH$_2$R$_2$	−0.231(5)	−0.133(−0.056)6
17		C-027, R-CH-X	−0.347(8)	−0.284(−0.120)2
18		H-046, H attached to C0(SP^3), no heteroatom (X) attached to next carbon	−0.127(5)	−0.011(−0.005)15

[a]The average regression coefficient of the descriptor corresponding to all models and the total number of its incidences; the arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.
[b]MLR like regression coefficient of four-component PLS model; (fc) is fraction contribution of the regression coefficient to the activity; order indicates the order of their significance in the PLS model; the constant term of PLS model is 6.024; number of compounds are 30. PLS regression and validation statistics: $r=0.955$, $s=0.265$, $F=65.153$, $Q^2_{LOO}=0.872$, $Q^2_{L5O}=0.883$, $r^2_{Test}=0.666$.

advantageous to enhance the activity. A higher value of seventh order mean topological charge index (JGI7) and Geary autocorrelation of lag 7 weighed by atomic Sanderson electronegativities (GATS7e) are detrimental to the activity. Counts for certain structural fragment, R-CH-X (descriptor C-027) strongly recommend the absence of such structural features favourable to activity. Thus the descriptors identified for rationalizing the activity give avenues to modulate the structure to a desirable biological endpoint.

In model Eqs. (1)–(8), the participated descriptors have indicated that the steric effect explained through molecular size, shape, and spacer (BELm1, PW4, MSD), the electronic effect (MATS7e, GATS7e, GGI9, JGI2, JGI7), the polarizability, accounting for both polarity and hydrophobicity (MATS8p), and certain structural fragments (C-002, C-027) played significant role in explaining the binding affinity of *N*-methylpiperazinylquinoxaline derivatives. However, the electronic and steric effects have appeared to impart dominant character in rationalizing the binding activity of titled compounds.

A partial least square (PLS) analysis has been carried out on these 18 CP-MLR identified descriptors (Table 2) to facilitate the development of a "single window" structure–activity model. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit SD) to give each one of them equal weight in the analysis. In the PLS cross-validation, four components are found to be the optimum for these 18 descriptors and they explained 91.20% variance in the activity ($r^2=0.912$, $Q^2_{LOO}=0.872$, $s=0.265$, $F=65.153$, $r^2_{Test}=0.666$). The MLR-like PLS coefficients of these 18 descriptors are given in Table 2. For the sake of comparison, the plot showing goodness of fit between observed and calculated activities (through PLS analysis) for the training and test set compounds is also given in Figure 1. Figure 2 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity (Table 2).

The PLS analysis has also suggested BELm1 (a BCUT class descriptor) as the most determining descriptor for modelling the activity of the compounds (descriptor S. No. 7 in Table 2; Figure 2). The other nine significant descriptors in decreasing order of significance are C-027, GGI9, CIC5, JGI7, C-002, JGI2, MATS7e, MATS4v, and nF (descriptors S. Nos. 17, 8, 6, 10, 16, 9, 13, 11, and 1 in Table 2; Figure 2). Of these descriptors, C-027, GGI9, JGI7, C-002, JGI2, MATS7e are part of Eqs. (1)–(8) and convey same inference in the PLS model as well.

Table 3.   Observed and modeled binding affinity of *N*-methylpiperazinylquinoxaline derivatives included in the training set.

Cpd.[a]	Obsd.	Eq. (5)		Eq. (6)		Eq. (7)		Eq. (8)		PLS	
		Calc.	LOO	Calc.	LOO	Calc.	LOO	Calc.	LOO	Calc.	LOO
2	6.70	6.36	6.33	6.34	6.31	6.48	6.46	6.45	6.39	6.41	6.37
3	4.99	4.76	4.65	4.66	4.50	4.18	3.52	4.23	3.85	4.49	4.25
6	5.13	5.88	5.91	5.82	5.86	5.89	5.92	5.56	5.60	5.70	5.73
7	6.49	5.96	5.92	6.19	6.18	6.08	6.04	5.85	5.70	6.25	6.21
8	6.53	6.05	6.02	6.12	6.11	6.09	6.07	5.77	5.70	6.06	6.03
9	6.32	6.25	6.24	6.01	5.99	6.00	5.97	6.14	6.11	6.20	6.19
11	6.33	5.80	5.73	6.20	6.20	6.12	6.11	5.89	5.85	6.10	6.08
12	5.53	5.60	5.65	6.26	6.29	6.12	6.15	5.69	5.70	5.64	5.65
13	5.36	5.46	5.53	6.22	6.25	6.12	6.16	5.79	5.83	5.47	5.48
14	6.15	5.81	5.78	5.83	5.81	6.02	6.02	6.03	6.01	5.89	5.86
15	5.86	5.55	5.52	5.48	5.42	5.80	5.80	5.92	5.92	5.68	5.67
16	5.64	5.82	5.83	5.88	5.90	5.87	5.89	6.05	6.09	5.99	6.01
21	5.8	5.91	5.93	5.86	5.87	5.77	5.77	5.86	5.87	5.99	6.01
22	6.24	5.77	5.73	5.95	5.93	5.91	5.88	5.71	5.68	6.25	6.25
23	5.66	5.51	5.49	5.65	5.65	5.74	5.74	5.83	5.85	5.81	5.82
24	5.63	5.59	5.58	5.54	5.53	6.01	6.03	5.85	5.87	5.93	5.94
25	5.81	5.68	5.64	5.49	5.41	5.69	5.68	5.79	5.79	5.76	5.75
26	4.88	5.84	5.91	4.91	5.26	4.93	5.72	5.80	6.02	5.07	5.10
27	5.24	5.79	5.85	6.03	6.07	5.64	5.71	5.18	5.16	5.15	5.11
28	6.64	6.73	6.74	6.90	6.94	6.87	6.91	6.43	6.42	6.75	6.76
29	5.4	5.96	6.04	5.58	5.60	5.47	5.49	5.81	5.84	5.29	5.25
30	6.47	6.83	6.88	6.89	6.95	6.97	7.08	6.96	7.08	6.78	6.83
31	7.21	6.64	6.58	6.76	6.71	6.92	6.88	7.04	7.00	7.05	7.02
32	7.58	7.62	7.63	7.54	7.53	7.46	7.43	7.43	7.38	7.50	7.48
34	8.25	8.21	8.19	7.88	7.74	7.71	7.53	8.06	7.96	8.21	8.19
36	7.2	7.26	7.26	7.12	7.11	7.10	7.08	7.10	7.05	7.25	7.27
39	5.6	5.73	5.84	5.59	5.54	5.51	5.50	5.81	5.93	5.58	5.58
42	5.16	5.17	5.17	5.15	5.15	5.32	5.41	5.43	5.52	5.55	5.78
43	6.23	6.30	6.32	5.95	5.90	5.73	5.50	6.12	6.10	6.16	6.13
44	4.69	4.89	4.99	4.92	5.02	5.20	5.45	5.11	5.33	4.75	4.79

[a]As in Table 1.
[b]On molar basis.

Table 4.   Observed, modelled, and residual[a] binding affinity of *N*-methylpiperazinylquinoxaline derivatives included in the test set.

Cpd.[b]	Obsd.	Eq. (5)		Eq. (6)		Eq. (7)		Eq. (8)		PLS	
		Calc.	Resi.	Calc.	Resi.	Calc.	Resi.	Calc.	Resi.	Calc.	Resi.
1	6.05	5.96	0.09	6.55	−0.50	6.74	−0.69	6.81	−0.76	6.25	−0.20
5	6.44	5.95	0.49	6.22	0.22	6.28	0.16	5.90	0.54	6.29	0.15
10	5.53	5.85	−0.32	5.00	0.53	5.87	−0.34	6.06	−0.53	5.66	−0.13
17	6.57	5.87	0.70	5.76	0.81	6.02	0.55	5.94	0.63	6.28	0.29
18	5.89	5.34	0.55	6.26	−0.37	6.11	−0.22	5.96	−0.07	5.85	0.04
19	5.63	5.79	−0.16	5.79	−0.16	5.77	−0.14	5.93	−0.30	6.05	−0.42
20	5.77	6.03	−0.26	5.75	0.02	5.70	0.07	6.09	−0.32	6.43	−0.66
33	7.93	7.41	0.52	7.41	0.52	7.38	0.55	7.85	0.08	7.75	0.18
35	5.40	5.75	−0.35	5.01	0.39	5.21	0.19	5.59	−0.19	5.47	−0.07
37	5.93	7.03	−1.10	6.92	−0.99	6.78	−0.85	6.70	−0.77	7.01	−1.08
38	7.24	8.31	−1.07	7.99	−0.75	7.78	−0.54	7.89	−0.65	7.88	−0.64
40	6.64	6.63	0.01	6.86	−0.22	7.09	−0.45	6.91	−0.27	6.96	−0.32
41	7.04	7.12	−0.08	7.38	−0.34	7.28	−0.24	7.64	−0.60	7.11	−0.07
45	5.12	5.27	−0.15	5.87	−0.75	6.23	−1.11	5.98	−0.86	5.75	−0.63

[a]Difference of observed and calculated pK_i value.
[b,c]See footnote under Table 3.

Among remaining ones, CIC5 belongs to TOPO class. The sign of regression coefficient of descriptor CIC5 (complementary information content of fifth order neighbourhood symmetry) suggests its detrimental nature to activity. The negative regression coefficient of the Moran autocorrelation of lag 4 weighed by atomic van der Waals volumes (descriptor MATS4v) advocates that a higher positive value of it is detrimental to the
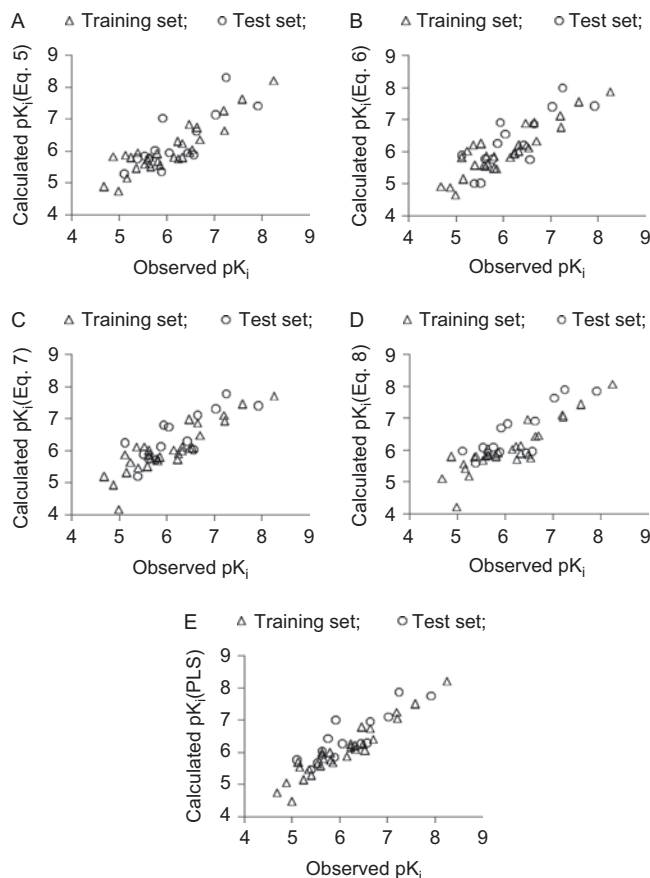
activity. The CONST class descriptor nF recommends a lower number or absence of fluorine atoms in a molecular structure for elevated activity. In comparison with these 10 descriptors, the remaining ones appear in lower order of significance to influence the activity of the compounds (Table 2; Figure 2). It is also observed that PLS model from the dataset devoid of 18 descriptors (Table 2) is inferior in explaining the activity of the analogues.

Applicability domain

On analyzing the model AD in the Williams plot (Figure 3) of the model based on the whole dataset (Table 5), it has appeared that none of the compounds were identified as an obvious outlier for the H$_4$R binding affinity if the limit of normal values for the Y outliers (response outliers) was set as 2.5 (standard deviation) units. None of the compounds was found to have leverage (h) values greater than the threshold leverages (h^*). For both the training set and test set, the suggested model matches the high-quality parameters with good fitting power and the capability of assessing external data. Furthermore, almost all of the compounds was within the AD of the proposed model and were evaluated correctly.

Conclusions

This study has provided a rational approach for the development of new N-methylpiperazinylquinoxaline derivatives as H$_4$ receptor ligands. The descriptors identified in CP-MLR analysis have highlighted the role of path/walk 4-Randic shape index (PW4), MSD, topological charges (GGI9, JGI2, and JGI7), atomic properties in respective lags of 2D-autocorrelations (MATS7e, GATS7e, and MATS8p), and Burden matrix (BELm1) to the activity. Certain

Figure 1. Plot of observed versus calculated pK_i values for the training (Δ) and test set ($\circ$) compounds. A, B, C, and D correspond, respectively, to four-parameter models (5), (6), (7), and (8) identified through CP-MLR. E corresponds to the MLR-like PLS equation from 18 CP-MLR-identified descriptors.
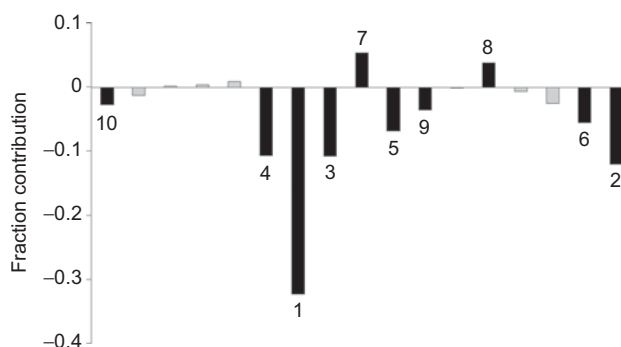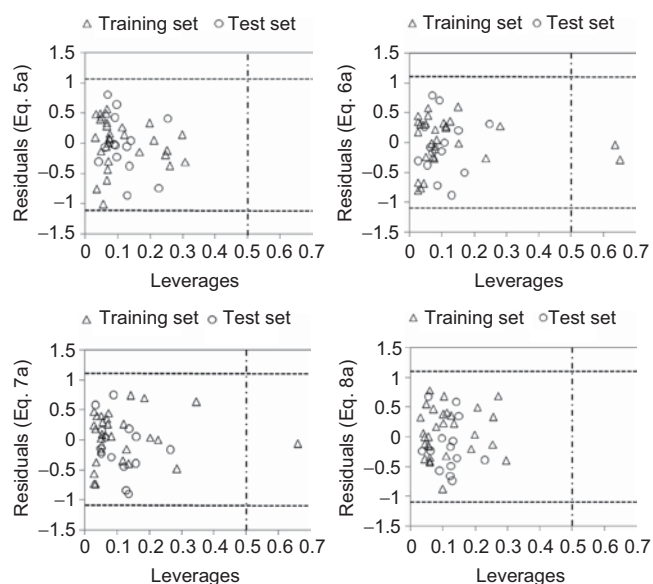

Figure 2. Plot of fraction contribution of MLR-like PLS coefficients (normalized) of the 18 descriptors (Table 2) to the activity. The 10 most significant descriptors are identified by black-shaded lines. The numerals on bars are the order of the descriptors.

Figure 3. Williams plot for the training set and external prediction set for H$_4$R binding affinity of N-methylpiperazinylquinoxaline derivatives.

Table 5. Resultant models for the whole dataset ($n = 44$) in descriptors of training set models.

Model	r	s	F	Q^2_{LOO}	Q^2_{L5O}	Eq.
$pK_i = 56.188 - 24.333(3.726)BELm1$ $-5.094(0.907)GGI9 + 1.977(0.661)MATS7e$ $-0.310(0.118)C\text{-}027$	0.860	0.437	27.774	0.675	0.675	5a
$pK_i = 63.082 - 27.568(3.513)BELm1$ $-4.565(0.928)GGI9 - 0.672(0.277)GATS7e$ $-0.172(0.076)C\text{-}002$	0.854	0.446	26.159	0.656	0.660	6a
$pK_i = 56.111 + 19.364(4.707)MSD$ $+55.586(13.834)PW4 - 31.839(3.484)BELm1$ $-0.217(0.074)C\text{-}002$	0.854	0.445	26.339	0.646	0.582	7a
$pK_i = 56.984 - 25.527(3.655)BELm1$ $+17.335(4.633)JGI2 - 82.597(19.769)JGI7$ $+2.159(0.653)MATS8p$	0.853	0.447	26.073	0.644	0.617	8a

structural fragments have also shown prevalence to optimize the H_4R binding affinity of titled compounds. The PLS analysis has confirmed the dominance of information content of CP-MLR identified descriptors for modelling the activity when compared with those of leftover ones.

Acknowledgements

Declaration of interest

References

1. Hill SJ. Distribution, properties, and functional characteristics of three classes of histamine receptor. Pharmacol Rev 1990; 42:45–83.

2. Black JW, Duncan WA, Durant CJ, Ganellin CR, Parsons EM. Definition and antagonism of histamine H2-receptors. Nature 1972;236:385–390.

3. Celanire S, Wijtmans M, Talaga P, Leurs R, de Esch IJP. Histamine H3 receptor antagonists reach out for the clinic. Drug Discov Today 2005;23/24:1613–1627.

4. Liu C, Ma X-J, Jiang X, Wilson SJ, Hofstra CL, Blevitt K, Li X, Chai W, Carruthers N, Lovenberg TW. Cloning and pharmacological characterization of a fourth histamine receptor (H4) expressed in bone marrow. Mol Pharmacol 2001;59:420–426.

5. Morse KL, Behan J, Laz TM, West RE Jr, Greenfender SA, Anthes JC, Umland S, Wan Y, Hipkin RW, Gonsiorek W, Shin N, Gustafson EL, Qiao X, Wang S, Hedrick JA, Green J, Bayne M, Monsma FJ Jr. Cloning and characterization of a novel human histamine receptor. J Pharmacol Exp Ther 2001;296:1058–1066.

6. Nguyen T, Shapiro DA, George SR, Setola V, Lee DK, Cheng R, Rauser L, Lee SP, Lynch KR, Roth BL, O'Dowd BF. Discovery of a novel member of the histamine receptor family. Mol Pharmacol 2001;59:427–433.

7. Oda T, Morikawa N, Saito Y, Masuho Y, Matsumoto S. Molecular cloning and characterization of a novel type of histamine receptor preferentially expressed in leukocytes. J Biol Chem 2000;275:36781–36786.

8. Zhu Y, Michalovich D, Wu H-L, Tan KB, Dytko GM, Mannan IJ, Boyce R, Alston J, Tierney LA, Li X, Herrity NC, Vawter L, Sarau HM, Ames RS, Davenport CM, Hieble JP, Wilson S, Bergsma DJ, Fitzgerald LR. Cloning, expression, and pharmacological characterization of a novel human histamine receptor. Mol Pharmacol 2001;59:434–444.

9. de Esch IJP, Thurmond RLJA, Jongejan A, Leurs R. The histamine H4 receptor as a new therapeutic target for inflammation. Trends Pharmacol Sci 2005;26:462–469.

10. Gantner F, Sakai K, Tusche MW, Cruikshank WW, Center DM, Bacon KB. Histamine H4 and H2 receptors control histamine induced interleukin-16 release from human CD8+ T cells. J Pharmacol Exp Ther 2002;303:300–307.

11. Varga C, Horvath K, Berko A, Thurmond RL, Dunford PJ, Whittle BJ. Inhibitory effects of histamine H4 receptor antagonists on experimental colitis in the rat. Eur J Pharmacol 2005;522:130–138.

12. Thurmond RL, Desai PJ, Dunford PJ, Fung-Leung WP, Hofstra CL, Jiang W, Nguyen S, Riley JP, Sun S, Williams KN, Edwards JP, Karlsson L. A potent and selective histamine H4 receptor antagonist with anti-inflammatory properties. J Pharmacol Exp Ther 2004;309:404–413.

13. Ikawa Y, Suzuki M, Shiono S, Ohki E, Moriya H, Negishi E, Ueno K. Histamine H4 receptor expression in human synovial cells obtained from patients suffering from rheumatoid arthritis. Biol Pharm Bull 2005;10:2016–2018.

14. Bell JK, McQueen DS, Rees JL. Involvement of histamine H4 and H1 receptors in scratching induced by histamine receptor agonists in BalbC mice. Br J Pharmacol 2004;142:374–380.

15. Cianchi F, Cortesini C, Schiavone N, Perna F, Magnelli L, Fanti E, Bani D, Messerini L, Fabbroni V, Perigli G, Capaccioli S, Masini E. The role of cyclooxygenase-2 in mediating the effects of histamine on cell proliferation and vascular endothelial growth factor production in colorectal cancer. Clin Cancer Res 2005;19(Pt 1):6807–6815.

16. Smits RA, Lim HD, Hanzer A, Zuiderveld OP, Guaita E, Adami M, Coruzzi G, Leurs R, de Esch IJP. Fragment based design of new H4 receptor-ligands with anti-inflammatory properties in vivo. J Med Chem 2008;51:2457–2467.

17. Kiss R, Kiss B, Könczöl Á, Szalai F, Jelinek I, László V, Noszál B, Falus A, Keser GM. Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening. J Med Chem 2008;51:3145–3153.

18. Todeschini R, Consonni V. Dragon software (version 1.11-2001). Milano, Italy.

19. Chemdraw Ultra 6.0 and Chem3D Ultra. Cambridge Soft Corporation, Cambridge, USA.

20. Smits RA, Lim HD, Stegink B, Bakker RA, de Esch IJP, Leurs R. Characterization of the histamine H4 receptor binding site: Part I. Synthesis and pharmacological evaluation of dibenzodiazepine derivatives. J Med Chem 2006;49:4512–4516.

21. Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: analysis of Selwood et al. data set—a case study. QSAR Comb Sci 2003;22:583–595.

22. Sharma BK, Pilania P, Singh P. Modeling of cyclooxygnase-2 and 5-lipooxygenase inhibitory activity of apoptosis inducing agents

potentially useful in prostate cancer chemotherapy: the derivatives of diarylpyrazole. J Enz Inhib Med Chem 2009;24:607–615.

23. Sharma S, Sharma BK, Pilania P, Singh P, Prabhakar YS. Modeling of the growth hormone secretagogue receptor antagonistic activity using chemometric tools. J Enz Inhib Med Chem 2009;24:1024–1033.

24. Sharma BK, Pilania P, Singh P, Prabhakar YS. Combinatorial protocol in multiple linear regression/partial least-squares directed rationale for the caspase-3 inhibition activity of isoquinoline-1,3,4-trione derivatives. SAR QSAR Environ Res 2010;21:169–185.

25. Sharma BK, Pilania P, Sarbhai K, Singh P, Prabhakar YS. Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. Mol Divers 2010;14:371–384.

26. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 1978;20:397–405.

27. Kettaneh N, Berglund A, Wold S. PCA and PLS with very large data sets. Comput Stat Data Anal 2005;48:69–85.

28. Stahle L, Wold S. Multivariate data analysis and experimental design. In: Ellis GP, West WB. Eds., Biomedical research. Progress in medicinal chemistry. Elsevier Science Publishers, BV, Amsterdam. 1988;25:291–338.

29. Topliss JG, Edwards RP. Chance factors in studies of quantitative structure–activity relationships. J Med Chem 1979;22:1238–1244.

30. Katritzky AR, Dobchev DA, Slavov S, Karelson M. Legitimate utilization of large descriptor pools for QSAR/QSPR models. J Chem Inf Model 2008;48:2207–2213.

31. So S-S, Karplus M. Three-dimensional quantitative structure–activity relationship from molecular similarity matrices and genetic neural networks. 1. Method and validation. J Med Chem 1997;40:4347–4359.

32. Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. CP-MLR/PLS directed structure–activity modeling of the HIV-1 RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. QSAR Comb Sci 2004;23:234–244.

33. Gramatica P. Principles of QSAR models validation: internal and external. QSAR Comb Sci 2007;26:694–701.