QSAR study of the 5HT 1A receptor affinities of arylpiperazines using a genetic algorithm–artificial neural network model

1 AUTHOR:

Aziz Habibi-Yangjeh
University of Mohaghegh Ardabili
92 PUBLICATIONS   **1,006** CITATIONS

QSAR study of the 5-HT$_{1A}$ receptor affinities of arylpiperazines using a genetic algorithm–artificial neural network model

Aziz Habibi-Yangjeh

Abstract Genetic algorithm–multiparameter linear regression (GA–MLR) and genetic algorithm–artificial neural network (GA–ANN) models have been used for prediction of the 5-HT$_{1A}$ receptor affinities (pK_i) of 66 arylpiperazines. A large number of theoretical descriptors were calculated for each compound. The genetic algorithm (GA) was used for selection of the variables that resulted in the best fit to the MLR and ANN models. The models were generated using seven descriptors as variables. For evaluation of the predictive power of the models, pK_i values of 13 compounds in the prediction set were calculated. Mean percentage deviation (MPD) for the GA–MLR and GA–ANN models were 0.344 and 0.065, respectively. Comparison of the results obtained by use of the models reveals the GA–ANN model is superior to the GA–MLR model.

Keywords QSAR · 5-HT$_{1A}$ receptor affinities · Arylpiperazines · Genetic algorithm · Artificial neural network

Introduction

Substantial research effort has been expended on the 5-HT$_1$ receptor family, and a number of reviews have covered the potential therapeutic applications of agonists and antagonists of these receptors. The various serotonin (5-HT$_1$) receptors are strongly implicated in neuropsychiatric diseases such as anxiety, depression, and schizophrenia [1–3].

A. Habibi-Yangjeh (✉)
Department of Chemistry, Faculty of Science,
University of Mohaghegh Ardabili, P.O. Box 179, Ardabil, Iran
e-mail: ahabibi@uma.ac.ir

They are members of the superfamily of G protein-coupled receptors with seven transmembrane α-helices [4, 5]. Their biological effects are dependent on the signaling pathways generated by different conformations of the receptors and ligands [6, 7].

Although several experimental methods are available for screening chemicals for biological activity (e.g., in-vivo and in-vitro assay tests), all of which have been carried out using receptors and other biological materials of human, rat, mouse, and calf origin, at least [8], they are costly, time-consuming, and can potentially cause toxic side products from the experimental methods used today. This has meant that the development of computational methods as alternative tools for predicting the properties of chemicals has been a subject of intensive study. Among computational methods, quantitative structure–activity relationships (QSAR) have found diverse applications for predicting compounds' properties, including biological activity [9, 10], physical properties [11, 12], and toxicity [13, 14]. QSPR/QSAR models are essentially calibration models in which the independent variables are molecular descriptors that describe the structure of molecules and the dependent variable is the property and/or activity of interest. In QSAR studies, techniques which can be applied for model construction, for example multiple linear regression (MLR) and artificial neural networks (ANN), have been used for inspection of linear and nonlinear relationships between the activity of interest and molecular descriptors. Artificial neural networks (ANNs) have become popular in QSPR/QSAR models, because of their success when complex non-linear relationships exist amongst data [15, 16]. An ANN is formed from artificial neurons, connected with coefficients (weights), which constitute the neural structure and are organized in layers. The layers of neurons between the input and output layers

are called hidden layers. Neural networks do not need explicit formulation of the mathematical or physical relationships of the problem being handled. These give ANNs an advantage over traditional fitting methods for some chemical applications. For these reasons in recent years ANNs have been applied to a wide variety of chemical problems [17–28]. Application of these techniques usually requires selection of variables to build well-fitting models. Nowadays, genetic algorithms (GA) are well-known as interesting and more widely used methods for variable selection [29–31]. GA are stochastic methods used to solve optimization problems defined by fitness criteria, applying the evolution hypothesis of Darwin and different genetic functions, i.e. crossover and mutation.

Recently, QSAR models have been used to predict the biological properties of different arylpiperazines [32, 33]. Weber and da Silva have developed partial least-squares regression (PLS) models for prediction of the biological properties (pK_i) of 66 arylpiperazines [33]. These QSAR studies have the advantage that chemometric methods were used for classification and the results are discussed in detail. However, the suitability of these models for prediction of biological properties is rather poor (for example, the best correlation coefficients (R^2) of the models are 0.832 and 0.884 for training and test sets).

In order to predict accurately the biological properties of these compounds, in this work, genetic algorithm–multiparameter linear regression (GA–MLR) and genetic algorithm–artificial neural network (GA–ANN) models were used to generate QSAR models between the descriptors and pK_i of 66 arylpiperazines with diverse chemical structures and the results were compared with each other, those from previous work [33], and the experimental values.

Results and discussion

For selection of the most important descriptors the genetic algorithm technique was used. To select the optimum number of descriptors the effects of the number of the descriptors were investigated for one to ten descriptors.

The R^2 value can be generally increased by adding the additional predictor variables to the model, even if the added variable does not contribute to the reduction of the unexplained variance of the dependent variable. Therefore, the use of R^2 requires special attention. For this reason, it is better to use another statistical parameter, known as adjusted R^2 (R_{adj}^2), where R_{adj}^2 is defined by Eq. 1.

$$R_{adj}^2 = 1 - \left(1 - R^2\right)\left(\frac{n-1}{n-p-1}\right) \qquad (1)$$

R_{adj}^2 is interpreted similarly to the R^2 value, considering the number of degrees of freedom, also. It is adjusted by
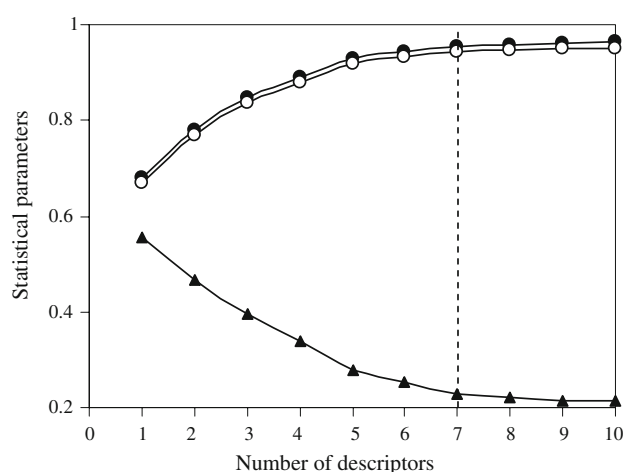

Fig. 1 Effects of the number of descriptors on R^2 (*filled circles*), R_{adj}^2 (*open circles*), and s (*filled triangles*) for the regression model

dividing the residual sum of squares and total sum of squares by their respective degrees of freedom. The R_{adj}^2 value diminishes if addition of a variable to the equation does not reduce the unexplained variance [34]. Subsequently, R_{adj}^2 is used to compare models with different numbers of predictor variables.

Another statistical parameter is the standard error of the estimate (s), that measures the dispersion of the observed values about the regression line. When the s value is low, the reliability of the prediction is higher.

Figure 1 shows the plots of R^2, R_{adj}^2, and s for the training set as a function of the number of descriptors for the 1–10 parameter models. R^2 and R_{adj}^2 increase with increasing number of descriptors. However, values of s decreased with increasing number of descriptors. As can be seen, the models with 8, 9, and 10 descriptors did not significantly improve the statistics of a model; it was found that the optimum subset size had been achieved with a maximum of seven descriptors.

The selected variables and the correlation matrix for these descriptors are shown in Table 1. From Table 1, it can be seen that the correlation coefficient value for each pair descriptors was less than 0.70, which meant that the selected descriptors were independent.

Table 1 Correlation coefficient matrix for the selected descriptors

	nCIR	BEHm5	GATS5e	G1m	G3m	L1s	R6v+
nCIR	1	0.698	−0.066	−0.504	−0.353	0.187	−0.231
BEHm5		1	−0.339	−0.497	−0.382	0.166	−0.415
GATS5e			1	−0.053	0.118	−0.325	0.432
G1m				1	0.553	0.082	0.330
G3m					1	−0.022	0.260
L1s						1	−0.367
R6v+							1

By interpreting the descriptors contained in the model, it is possible to gain useful chemical insights into the biological activity. For this reason, an acceptable interpretation of the QSAR results is provided below.

One of the constitutional descriptors appearing in the model is nCIR. This descriptor represents the number of circuits in molecules.

BEHm5 (the highest eigenvalue n.5 of the Burden matrix/weighted by atomic mass) is the second descriptor, appearing in the model. It is one of the BCUT descriptors. The BCUT (Burden, CAS, University of Texas eigenvalues) descriptors are the eigenvalues of a modified connectivity matrix known as the Burden matrix [35].

The third descriptor is GATS5e (Geary autocorrelation—lag 5/weighted by atomic Sanderson electronegativities), which is one of the 2D autocorrelation descriptors. In this descriptor, the Geary coefficient is a distance-type function, that function is any physicochemical property calculated for each atom of the molecule, such as atomic mass, polarizability, etc. Therefore, the molecule atoms represent the set of discrete points in space and the atomic property the function evaluated at those points. The physicochemical property in this case is atomic polarizability.

The descriptors G1m, G3m, and L1s belong to the WHIM descriptors, which are based on the statistical indices calculated from projection of atoms along principal axes. The algorithm consists of performing a principal-components analysis on the centered Cartesian coordinates of a molecule by using a weighted covariance matrix obtained from different weighting schemes for the atoms. Directional WHIM symmetry descriptors are related to the number of central symmetric atoms (along the mth component), the number of unsymmetric atoms, and the total number of atoms of the molecule. The weighting schemes that are used for computing the weighted covariance matrix are the atomic mass for the G1m and G3m descriptors and the atomic electrotopological states for the L1s descriptor.

The final descriptor is R6v+, which is the R maximal autocorrelation of lag 6/weighted by atomic van der Waals volumes. It is one of the geometry topology and atomic weight assembly (GETAWAY) descriptors that were presented by Consonni et al. [36]. They encode geometrical information given by the influence matrix, topological information given by the molecular graph, and chemical information obtained from selected atomic properties.

In summary, it is concluded that number of circuits in molecules, atomic polarizability, atomic mass, atomic electrotopological states, and atomic van der Waals volumes play the main roles in the 5-HT$_{1A}$ receptor affinities of the arylpiperazines.

Genetic algorithm–multiparameter linear regression

We used a GA for selection of the most relevant descriptors. Multiparameter linear correlation of pK_i values for 40 various arylpiperazines in the training set was obtained using seven descriptors selected by the GA. The calculated values of pK_i for the compounds in training, validation, and prediction sets using the GA–MLR model were plotted versus the experimental values of pK_i (Fig. 2).

Genetic algorithm–artificial neural network

To process non-linear relationships existing between the activity and the descriptors, the ANN modeling method combined with GA for feature selection was employed. The input vectors were the set of descriptors which were selected by the GA, and therefore, the number of nodes in the input layer was dependent on the number of selected descriptors. In the GA–MLR model it is assumed that the descriptors are independent of each other and truly additive with regard to the property under study. ANNs are particularly well-suited to QSAR/QSPR models because of their ability to extract non-linear information present in the data matrix. For this reason the next step in this work was generation of the ANN model. There are no rigorous theoretical principles for choosing the proper network topology; so different structures were tested in order to obtain the optimal hidden neurons and training cycles [22–25]. Before training the network, the number of nodes in the hidden layer was optimized. In order to optimize the number of nodes in the hidden layer, several training sessions were conducted with different numbers of hidden nodes (from 1 to 10). The root mean square error of training (RMSET) and validation (RMSEV) sets were
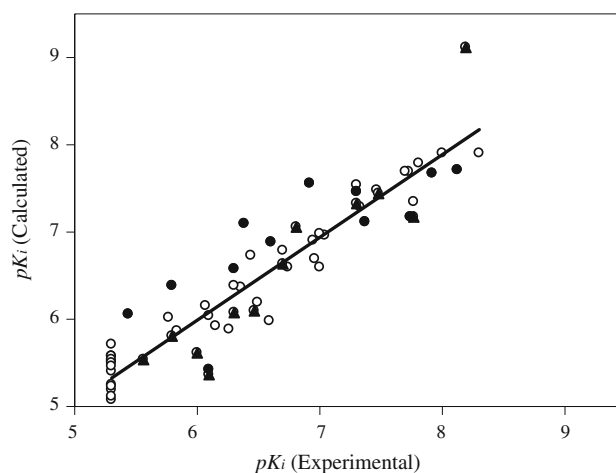

Fig. 2 Plot of the calculated values of pK_i from the GA–MLR model versus the experimental values for the training (*open circles*), validation (*filled circles*), and prediction (*filled triangles*) sets

obtained after various iterations for different numbers of neurons in the hidden layer and the minimum value of RMSEV was recorded as the optimum value. Plots of RMSET and RMSEV versus the number of nodes in the hidden layer are shown in Fig. 3. It is clear that seven nodes in the hidden layer is the optimum value.

This network consists of seven inputs, the same descriptors in the GA-MLR model, and one output for pK_i, so an ANN with architecture 7-7-1 was generated. It is noteworthy that training of the network was stopped when the RMSEV started to increase, i.e. when overtraining begins. Overtraining causes the ANN to lose its prediction power [16]. Therefore, during training of the network, it is desirable that iterations are stopped when overtraining begins. To control the overtraining of the network during the training procedure, the values of RMSET and RMSEV were calculated and recorded to monitor the extent of learning in the different iterations. Results showed that overfitting was not seen when the optimum architecture was used (Fig. 4).
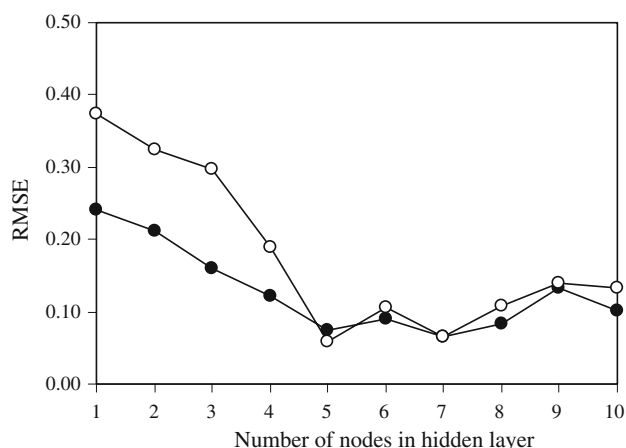

Fig. 3 Plot of RMSE for the training (*filled circles*) and validation (*open circles*) sets versus the number of nodes in the hidden layer
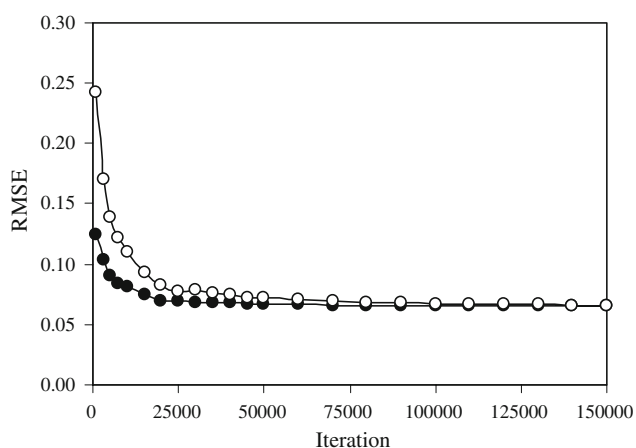

Fig. 4 Plot of RMSE for the training (*filled circles*) and validation (*open circles*) sets versus the number of iterations

The generated ANN was then trained using the training and validation sets for optimization of the weights and biases. For evaluation of the predictive power of the generated ANN, an optimized network was applied for prediction of the pK_i values in the prediction set which were not used in the modeling procedure (Table 2). The calculated values of pK_i for the compounds in the training, validation and prediction sets using the ANN model have been plotted versus the experimental values of pK_i in Fig. 5.

As expected, the calculated values of pK_i are in good agreement with the experimental values. The correlation equation for all the calculated values of pK_i from the ANN model and the experimental values is as given in Eq. 2.

$$pK_i(\text{cal}) = 0.998 pK_i(\text{exp}) + 0.009 \qquad (2)$$

$(R = 0.997; \text{MPD} = 0.065; \text{RMSE} = 0.772; F = 12720.76)$.

Similarly, the correlation equation for pK_i (cal) versus pK_i (exp) values in the prediction set is given by Eq. 3.

$$pK_i(\text{cal}) = 0.953 pK_i(\text{exp}) + 0.309 \qquad (3)$$

$(R = 0.998; \text{MPD} = 0.056; \text{RMSE} = 0.520; F = 3355.73)$.

Table 3 compares the results obtained using the GA–MLR and GA–ANN models. The MPD and RMSE of the models for total, training, validation, and prediction sets show the potential of the ANN model for prediction of pK_i values of various arylpiperazines.

As a result, it was found that a properly selected and trained neural network could fairly represent the dependence of the 5-HT$_{1A}$ receptor affinities of arylpiperazines on the descriptors. The optimized neural network could then simulate the complicated nonlinear relationship between the pK_i value and the descriptors. MPD and RMSE of 0.446 and 5.031 for the prediction set by the GA–MLR model should be compared with the values of 0.056 and 0.520 by the GA–ANN model. As can be seen, the ability of the proposed model to predict pK_i is better than the QSAR models proposed recently [33]. It can be seen from Table 3 that although parameters appearing in the GA–MLR model are used as inputs for the generated GA–ANN model, the statistics have shown a large improvement. These improvements are because pK_i values of arylpiperazines show non-linear correlations with the selected descriptors.

Data and methodology

The data set of 5-HT$_{1A}$ receptor affinities of arylpiperazines used for the QSAR analyses was selected from literature [33]. The z-matrices (molecular models) were constructed with HyperChem 7.0 and molecular structures were optimized using the AM1 algorithm [37]. In order to

calculate the theoretical descriptors, the Dragon package version 2.1 was used [38]. For this process the output of the HyperChem software for each compound was fed into the Dragon program and the descriptors were calculated. As a result, a total of 1,481 theoretical descriptors were calculated for each compound in the data set (66 compounds).

Table 2 Experimental and calculated values of pK_i for various arylpiperazines for the training, validation, and prediction sets for the GA-MLR and GA-ANN models[a]
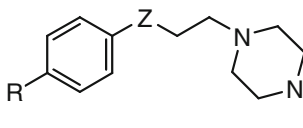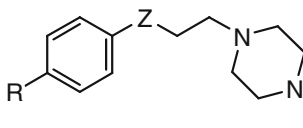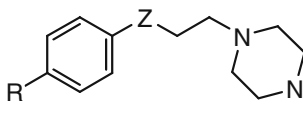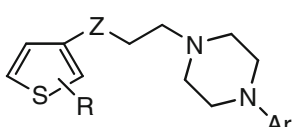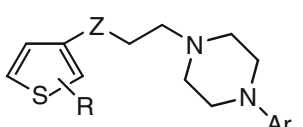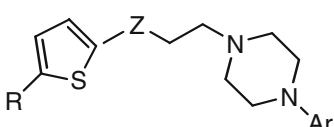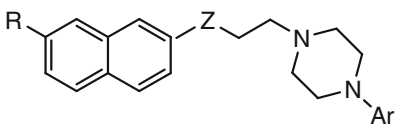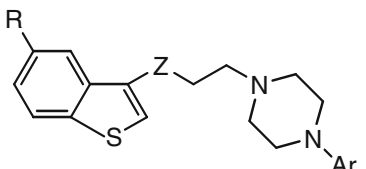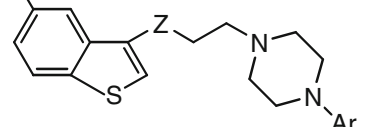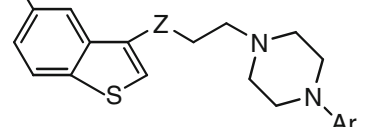
Compound	General structure	R	Z	Ar	$pK_{i \text{ (exp)}}$	Pred (I)	Pred (II)
1		H	CHOH	2-Methoxyphenyl	7.32	7.28	7.30
2		H	CHO-4-CF3C6H4	2-Methoxyphenyl	6.35	6.37	6.35
3		H	CNOH	2-Methoxyphenyl	7.76	7.34	7.84
4[b]		H	CO	4-Chlorophenyl	6.1	5.42	5.99
5		H	CHO-4-CH3C6H4	4-Chlorophenyl	5.84	5.87	5.82
6		H	CHO-3,4-OCH2OC6H3	4-Chlorophenyl	6.26	5.88	6.26
7		H	CO	4-Methoxyphenyl	5.3	5.08	5.31
8		H	CHOH	4-Methoxyphenyl	5.3	5.57	5.27
9		H	CO	2-Chlorophenyl	6.74	6.59	6.77
10		H	CHOH	2-Chlorophenyl	6.94	6.90	6.95
11		H	CHO-4-CF3C6H4	2-Chlorophenyl	5.3	5.71	5.20
12[c]		H	CO	4-Fluorophenyl	6.1	5.37	6.10
13		H	CHO-4-CF3C6H4	4-Fluorophenyl	5.3	5.25	5.32
14[b]		H	CO	2-Pyridyl	7.3	7.46	7.21
15[c]		H	CHOH	2-Pyridyl	6.81	7.05	6.92
16		H	CO	4-Nitrophenyl	5.3	5.51	5.38
17[b]		H	CHOH	4-Nitrophenyl	5.3	4.84	5.40
18[c]		H	CHO-4-CF3C6H4	4-Nitrophenyl	5.3	4.63	5.30
19		H	CO	2-Methoxyphenyl	7.3	7.53	7.25
20		H	CHOH	4-Chlorophenyl	6.1	6.03	6.05
21		H	CHO-4-CF3C6H4	4-Methoxyphenyl	5.3	5.47	5.37
22[b]		H	CO	2-Pyrimidyl	6.92	7.56	6.93
23[b]		H	CHO-4-CF3C6H4	2-Pyrimidyl	5.8	6.39	5.83
24[c]		H	CHO-4-CF3C6H4	2-Pyridyl	5.8	5.81	5.80
25		Phenyl	CHOH	2-Methoxyphenyl	6.07	6.16	5.95
26[b]		Phenyl	CO	2-Methoxyphenyl	5.44	6.05	5.56
27		Phenyl	CHO-4-CF3C6H4	2-Methoxyphenyl	5.3	5.19	5.29
28		Methoxy	CO	2-Methoxyphenyl	5.76	6.01	5.79
29		Methoxy	CHOH	2-Methoxyphenyl	6.49	6.19	6.45
30[c]		Methoxy	CHO-4-CF3C6H4	2-Methoxyphenyl	6	5.61	5.99
31[c]		H	CHOH	2-Methoxyphenyl	7.3	7.33	7.31
32		H	CHO-4-CF3C6H4	2-Methoxyphenyl	6.59	5.99	6.61
33[c]		H	CNOH	2-Methoxyphenyl	8.19	9.12	8.26
34		H	CO	4-Chlorophenyl	6.15	5.93	6.14
35		H	CO	2-Chlorophenyl	6.7	6.78	6.69
36[c]		H	CHOH	2-Chlorophenyl	6.7	6.63	6.66
37		H	CO	1-Naphthyl	7.46	7.48	7.44
38		H	CO	2-Methoxyphenyl	7.8	7.79	7.76
39[c]		H	CHOH	4-Chlorophenyl	5.56	5.53	5.56
40		2,5-Dimethyl	CO	2-Methoxyphenyl	8.3	7.91	8.48
41[b]		2,5-Dimethyl	CO	2-Hydroxyphenyl	8.12	7.71	8.03
42		2,5-Dimethyl	CHOH	2-Hydroxyphenyl	7.04	6.96	7.02
43		2,5-Dimethyl	CO	1-Naphthyl	7	6.99	6.98
44[b]		2,5-Dimethyl	CHOH	2-Methoxyphenyl	7.92	7.68	7.85

Table 2 continued

Compound	General structure	R	Z	Ar	$pK_{i\ (\mathrm{exp})}$	Pred (I)	Pred (II)
45		H	CO	2-Methoxyphenyl	8	7.90	7.90
46		H	CHOH	2-Methoxyphenyl	7.72	7.69	7.81
47		H	CO	4-Chlorophenyl	5.3	5.57	5.39
48		H	CHOH	4-Chlorophenyl	5.3	5.54	5.35
49[c]		5-Methyl	CO	2-Methoxyphenyl	7.76	7.18	7.76
50[c]		5-Methyl	CHOH	2-Methoxyphenyl	7.47	7.44	7.57
51[c]		5-Nitro	CO	2-Methoxyphenyl	6.47	6.10	6.49
52[b]		H	CHOH	2-Methoxyphenyl	6.38	7.10	6.38
53		H	CO	4-Chlorophenyl	5.3	5.23	5.40
54		H	CHOH	4-Chlorophenyl	5.3	5.11	5.26
55[b]		H	CO	2-Methoxyphenyl	6.6	6.89	6.60
56[b]		H	CO	2-Methoxyphenyl	7.36	7.12	7.40
57		H	CHOH	2-Methoxyphenyl	7.7	7.69	7.66
58		H	CO	4-Chlorophenyl	5.3	5.50	5.20
59		H	CHOH	4-Chlorophenyl	5.3	5.40	5.17
60		H	CO	2-Hydroxyphenyl	6.96	6.69	6.90
61[b]		H	CHOH	2-Hydroxyphenyl	7.74	7.17	7.75
62[b]		H	CO	4-Chloro-2-methoxyphenyl	6.3	6.58	6.33
63		H	CHOH	4-Chloro-2-methoxyphenyl	6.44	6.73	6.50
64[c]		H	CHOH	4-Fluoro-2-methoxyphenyl	6.3	6.07	6.19
65		H	CO	1-Naphthyl	7	6.59	7.00
66		H	CO	4-Fluoro-2-methoxyphenyl	6.3	6.38	6.42

[a] I and II denote the values obtained by using the GA–MLR and GA–ANN models, respectively

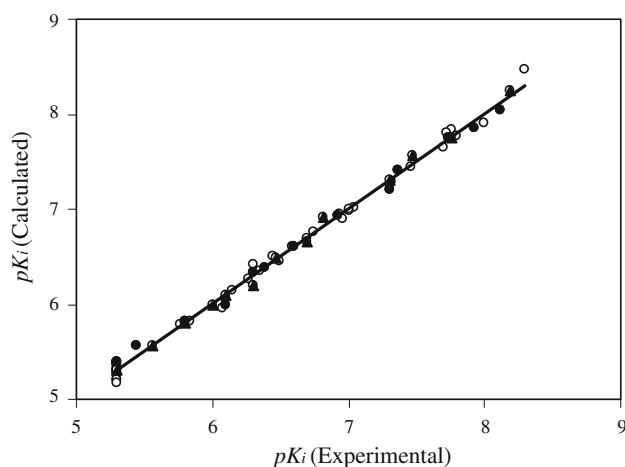[b] The validation set

[c] The prediction set

Fig. 5 Plot of the calculated values of pK_i from the GA–ANN model versus the experimental values of pK_i for the training (*open circles*), validation (*filled circles*), and prediction (*filled triangles*) sets

The theoretical descriptors were reduced by the following procedure:

1. descriptors that were constant were eliminated (436 descriptors); and

2. to reduce the redundancy existing in the descriptors, the correlation of descriptors with each other and with pK_i of the molecules were examined, and collinear descriptors ($R > 0.9$) were detected. Among the collinear descriptors, that with the highest correlation with pK_i values was retained and the others were removed from the data matrix (629 descriptors).

Genetic algorithm

To select the most relevant descriptors, evolution of the population was simulated [39–43]. Each individual of the population defined by a chromosome of binary values represented a subset of descriptors. The number of genes on each chromosome was equal to the number of descriptors. The population of the first generation was selected randomly. A gene took a value of 1 if its corresponding descriptor was included in the subset; otherwise, it took a value of zero. The number of genes with a value of 1 was kept relatively low to maintain a small subset of descriptors [43], that is, the probability of generating 0 for a gene was set greater (at least 60%) than the value unity.

Table 3 Comparison of statistical parameters obtained by using the GA–MLR and GA–ANN models for pK_i values of arylpiperazines[a]

Model	RMSE$_{tot}$	RMSE$_{train}$	RMSE$_{valid}$	RMSE$_{pred}$	MPD$_{tot}$	MPD$_{train}$	MPD$_{valid}$	MPD$_{pred}$
GA–MLR	4.124	2.882	7.037	5.031	0.344	0.228	0.490	0.446
GA–ANN	0.772	0.832	0.841	0.520	0.065	0.067	0.069	0.056

[a] The subscript "train" refers to the training set, "valid" refers to the validation set, "pred" refers to the prediction set, and "tot" refers to the total data set

RMSE is the root mean square error and MPD is the mean percentage deviation

The operators used here were crossover and mutation. The probability of the application of these operators was varied linearly with generation renewal (0–0.1% for mutation and 60–90% for crossover). The population size was varied between 50 and 250 for different GA runs. For a typical run, the evolution of the generation was stopped when 90% of the generations took the same fitness [29]. The GA program was written in Matlab 6.5 [44].

Artificial neural network

A feed forward artificial neural network with a back-propagation of error algorithm was used to process the non-linear relationship between the selected descriptors and the biological activity (pK_i). The number of input nodes in the ANN was equal to the number of descriptors. The ANN models were confined to a single hidden layer, because a network with more than one hidden layer would be harder to train. A three-layer network with a sigmoidal transfer function was designed. The initial weights were randomly selected between 0 and 1. Optimization of the weights and biases was carried out according to the resilient back-propagation algorithm. The data set was randomly divided into three groups: a training set, a validation set, and a prediction set consisting of 40, 13, and 13 molecules. The training and validation sets were used for generation of the model and the prediction set was used for evaluation of the generated model. The training, validation, and prediction performances of models are evaluated by the mean percentage deviation (MPD) and root mean square error (RMSE), which are defined by Eqs. 4 and 5.

$$MPD = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\left(P_i^{exp} - P_i^{cal}\right)}{P_i^{exp}}\right| \qquad (4)$$

$$RMSE = \sqrt{\sum_{i=1}^{N}\frac{\left(P_i^{exp} - P_i^{cal}\right)^2}{N}} \qquad (5)$$

where P_i^{exp} and P_i^{cal} are experimental and calculated values of pK_i with the models and N denotes the number of data points. Individual percent deviation (IPD) is defined by Eq. 6.

$$IPD = 100 \times \left(\frac{P_i^{cal} - P_i^{exp}}{P_i^{exp}}\right). \qquad (6)$$

Processing of the data was carried using Matlab 6.5 [45]. The neural networks were implemented using Neural Network Toolbox Ver. 4.0 for Matlab [46].

Conclusions

Quantitative structure–activity relationships were sought for the 5-HT$_{1A}$ receptor affinities of 66 different arylpiperazines by using the GA–MLR and GA–ANN methods. Comparison of the MPD values (and other statistical parameters in Table 3) for the training, validation, and prediction sets for the GA–MLR and GA–ANN models demonstrate superiority of the GA–ANN model over the GA–MLR model. Mean percentage deviation of 0.446 for the prediction set by the GA–MLR model should be compared with the value of 0.056 for the GA–ANN model. Because the improvement of the results obtained by using the non-linear model (GA–ANN) is substantial, it can be concluded there is a non-linear relationship between the pK_i values and the characteristic descriptors of the arylpiperazines.

References

1. Hamon M (1997) Handbook of experimental pharmacology. Springer, Berlin
2. Zifa E, Fillion G (1992) Pharmacol Rev 44:401
3. Saudou F, Hen R (1994) Med Chem Res 4:16
4. Hoyer D, Fozard JR, Mylecharane EJ, Clarke DE, Martin GR, Humphrey PPA (1994) Pharmacol Rev 46:153
5. Verges D, Calas A (2000) J Chem Neuroanat 18:41
6. Perez DM, Hwa J, Gaivin R, Mathur M, Brown F, Graham RM (1996) Mol Pharmacol 49:112
7. Sidhu A, Niznik HB (2000) Int J Dev Neurosci 18:669
8. Hill DL (1972) The biochemistry and physiology of tetrahymena, 1st edn edn. Academic Press, New York
9. Seierstad M, Agrafiotis DK (2006) Chem Biol Drug Des 67:284
10. Wang Y, Wang XW, Cheng YY (2006) Chem Biol Drug Des 68:166
11. Liu J, Yang L (2006) Bioorg Med Chem 14:2225
12. Verma RP, Kurup A, Hansch C (2005) Bioorg Med Chem 13:237
13. Toropov AA, Benfenati E (2006) Bioorg Med Chem Lett 16:1941
14. Khadikar PV, Phadnis A, Shrivastava A (2002) Bioorg Med Chem 10:1181
15. Despagne F, Massart DL (1998) Analyst 123:157

16. Zupan J, Gasteiger J (1999) Neural networks in chemistry and drug design. Wiley-VCH, Germany
17. Meiler J, Meusinger R, Will M (2000) J Chem Inf Comput Sci 40:1169
18. Habibi-Yangjeh A, Pourbasheer E, Danandeh-Jenagharad M (2008) Bull Korean Chem Soc 29:833
19. Habibi-Yangjeh A, Pourbasheer E, Danandeh-Jenagharad M (2008) Monatsh Chem. doi: 10.1007/s00706-008-0951-z
20. Habibi-Yangjeh A, Pourbasheer E, Danandeh-Jenagharad M (2008) Monatsh Chem. doi:10.1007/s00706-008-0049-7
21. Habibi-Yangjeh A, Nooshyar M (2005) Phys Chem Liq 43:239
22. Habibi-Yangjeh A, Nooshyar M (2005) Bull Korean Chem Soc 26:139
23. Habibi-Yangjeh A, Danandeh-Jenagharad M, Nooshyar M (2005) Bull Korean Chem Soc 26:2007
24. Habibi-Yangjeh A, Danandeh-Jenagharad M, Nooshyar M (2006) J Mol Model 12:338
25. Tabaraki R, Khayamian T, Ensafi AA (2006) J Mol Graph Model 25:46
26. Habibi-Yangjeh A (2007) Phys Chem Liq 45:471
27. Habibi-Yangjeh A, Danandeh-Jenagharad M (2007) Indian J Chem 46B:478
28. Habibi-Yangjeh A, Esmailian M (2007) Bull Korean Chem Soc 28:1477
29. Depczynski U, Frost VJ, Molt K (2000) Anal Chim Acta 420:217
30. Alsberg BK, Marchand-Geneste N, King RD (2000) Chemom Intell Lab Syst 54:75
31. Jouanrimbaud D, Massart DL, Leardi R, Denoord OE (1995) Anal Chem 67:4295
32. Lopez-Rodriguez ML, Morcillo MJ, Fernandez E, Rosado ML, Pardo L, Schaper KJ (2001) J Med Chem 44:198
33. Weber KC, da Silva ABF (2008) Eur J Med Chem 43:364
34. Hansch C, Taylor J, Sammes P (1990) Comprehensive medicinal chemistry: the rational design, mechanistic study and therapeutic application of chemical compounds, vol 6. Pergamon, New York, pp 1–19
35. Burden FR (1989) J Chem Inf Comput Sci 29:225
36. Consonni V, Todeschini R, Pavan M, Gramatica P (2002) J Chem Inf Comput Sci 42:693
37. HyperChem Release 7, HyperCube, Inc., http://www.hyper.com
38. Todeschini R, Milano Chemometrics and QSPR Group, http://www.disat.unimib.it/chm
39. Cho SJ, Hermsmeier MA (2002) J Chem Inf Comput Sci 42:927
40. Baumann K, Albert H, Von Korff M (2002) J Chemometr 16:339
41. Lu Q, Shen G, Yu R (2002) J Comput Chem 23:1357
42. Ahmad S, Gromiha MM (2003) J Comput Chem 24:1313
43. Deeb O, Hemmateenejad B, Jaber A, Garduno-Juarez R, Miri R (2007) Chemosphere 67:2122
44. The Mathworks Inc (2002) Genetic algorithm and direct search toolbox user's guide, Massachusetts
45. Matlab 6.5. Mathworks, 1984–2002
46. The Mathworks Inc (2002) Neural network toolbox user's guide, Massachusetts