A quantitative structure–activity relationship study on histamine receptor antagonists using the genetic algorithm–multi--parameter linear regression method

MARYAM ADIMI[1], MAHMOUD SALIMI[2], MEHDI NEKOEI[3*],
ESLAM POURBASHEER[4] and ABOLGHASEM BEHESHTI[4]

[1]Chemical Engineering Department, Faculty of Engineering, Islamic Azad University – Farahan Branch, Farmahin, Iran, [2]Chemical Engineering Department , Faculty of Engineering, Islamic Azad University – Arak Branch, Arak, Iran, [3]Department of Chemistry, Shahrood branch, Islamic Azad University, Shahrood, Iran and [4]Center of Excellence in Electrochemistry, Faculty of Chemistry, University of Tehran, P. O. Box 14155-6455, Tehran, Iran

Abstract: A quantitative structure activity relationship (QSAR) model has been generated for predicting the antagonist potency of biphenyl derivatives as human histamine (H3) receptors. The molecular structures of the compounds were numerically represented by various kinds of molecular descriptors. The whole data set was divided into training and test sets. A genetic algorithm based multiple linear regression was used to select the most statistically effective descriptors. The final QSAR model ($N = 24$, $R^2 = 0.916$, $F = 51.771$, $Q^2_{LOO} = 0.872$, $Q^2_{LGO} = 0.847$, $Q^2_{BOOT} = 0.857$) was fully validated employing the leave-one-out (LOO) cross-validation approach, Fischer statistics (F), the *Y*-randomization test, and predictions based on the test data set. The test set presented an external prediction power of $R^2_{test} = 0.855$. In conclusion, the generated QSAR model could be used as a valuable tool for designing similar groups of new antagonists of histamine (H3) receptors.

Keywords: QSAR; genetic algorithm; multiple linear regression; biphenyl derivatives; histamine (H3) receptors.

INTRODUCTION

Alzheimer's disease is the most common form of neurodegenerative dementia.[1] It accounts for approximately 50–60 % of the overall cases of dementia among individuals over the age of 65 years.[2] Unfortunately, the therapeutic options for Alzheimer's disease are limited. No true disease-modifying therapies are known and at best, physicians can only hope to alleviate with symptomatic

treatments the cognitive deficits that characterize this disease. One of the most frequently prescribed anti-Alzheimer's drugs are the histamine H3 receptor antagonists. Histamine is a biogenic amine that influences a wide range of pathophysiological processes through the activation of different G protein-coupled receptors (GPCRs).[3] The histamine H3 receptor was identified in 1983 and was initially described as an autoreceptor, mainly expressed in the central nervous system (CNS), regulating histamine biosynthesis and release from histaminergic neurons.[4] The high density of H3 receptors in different CNS areas and their influence on the release of a large variety of neurotransmitters encouraged wide pharmacological investigations on their physiological role and the quest for potential therapeutic applications of H3-antagonists in the treatment of various CNS diseases.[5,6]

Recently, some classes of H3-antagonists have been described that are endowed with additional pharmacological properties that may synergistically potentate their therapeutic efficacy in the treatment of disorders related to neurotransmitter deficits, such as Alzheimer's disease. Although there are several experimental methods available for screening the estrogenic activity of chemicals (*e.g.*, *in vivo* and *in vitro* assay tests), which have been performed using receptors and other biological material of human, rat, mouse, and calf origin at least,[7] they are costly, time-consuming, and can potentially produce toxic side products from the experimental methods used today. An efficient way to obtain a complete data set without the necessity of performing expensive laboratory experiments is the application of quantitative structure activity relationship (QSAR) techniques. A QSAR model describes a mathematical relationship between structural attribute(s) and activity of a set of chemicals. The potential promise of using QSAR models for the screening of chemical databases or virtual libraries before their synthesis appears equally attractive for chemical manufacturers, pharmaceutical companies, and government agencies, particularly in times of shrinking resources.[8] The main aim of the present work was to establish a new QSAR model for predicting the antagonist potency of biphenyl derivatives as histamine (H3) receptors by a genetic algorithm–multiple linear regression (GA–MLR) technique. In addition, several possible approaches for QSAR model validation are described, including the leave-one-out (LOO), leave-group-out (LGO), cross-validation, external test set and the *Y*-randomization test approaches.

<div align="center">MATERIALS AND METHODS</div>

Data set

The 30 studied biphenyl derivatives and the corresponding antagonist potency (pK_B) at human histamine H3-receptors (hH3) were collected from the literature.[9] The chemical structures and activity data (pK_B) for the complete set of compounds are presented in Table I. The data set was randomly divided into two subsets: the training set containing 24 compounds (80 %) and the test set containing 6 compounds (20 %). The training set was used to build a regression model, while the test set was used to evaluate the predictive ability of the obtained model.

TABLE I. Experimental and predicted antagonist potency of biphenyl derivatives at H3 receptors (pK_B) by the GA–MLR technique

$$R_1-(CH_2)_n-\text{[biphenyl]}-(CH_2)_m-R_2$$
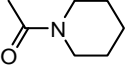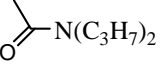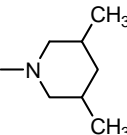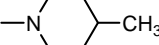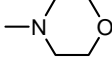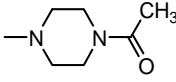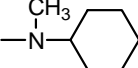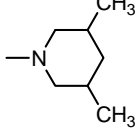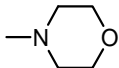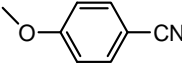
No.	n	m	R_1	R_2	Exp.	Predicted
1	1	2		—N(piperidine)	9.28	9.33
2	1	1	—N(C$_3$H$_7$)$_2$	—C(=O)—N(piperidine)	6.96	7.24
3	1	1		—C(=O)—N(C$_3$H$_7$)$_2$	8.16	8.02
4	1	2		—N(piperidine)	8.82	8.67
5	1	1			8.77	8.38
6	1	1		—N(CH$_3$)$_2$	8.17	8.11
7	1	1		—N(H)—(C$_5$H$_{11}$)	6.51	6.52
8	1	1		—N(3,5-dimethylpiperidine)	6.19	6.44
9	1	1		—N(4-methylpiperidine)	7.68	7.55
10	1	1		—N(morpholine)	7.67	7.87
11	1	1		—N(piperazine)—C(=O)CH$_3$	6.7	6.69
12	1	1	—N(piperidine)	—N(CH$_3$)(cyclohexyl)	9.19	8.76
13	1	1		—N(3,5-dimethylpiperidine)	8.48	8.16
14	1	1		—N(morpholine)	7.77	7.83
15	1	1		—O—(4-cyanophenyl)	7.27	7.22

TABLE I. Continued

$$R_1-(CH_2)_n-\underset{}{\text{[biphenyl]}}-(CH_2)_m-R_2$$

No.	n	m	R$_1$	R$_2$	Exp.	Predicted
16	1	1	—N(piperidine)	—N(CH$_3$)$_2$	8.92	9.25
17	1	1		—N(C$_2$H$_5$)$_2$	8.9	9.13
18	1	1		—N(C$_3$H$_7$)$_2$	8.18	8.41
19	1	1		—N(CH$_3$)(CH$_2$C$_6$H$_5$)	7.98	8.19
20	1	1		—N(piperazine)C(O)CH$_3$	7.86	7.38
21	1	1		—NH(cyclohexyl)	8.03	8.12
22	1	1		—N(piperidin-2-one)	7.46	7.76
23	2	2		—N(piperidine)	8.5	8.59
24	2	2		—N(4-methylpiperidine)	8.17	7.95
25[a]	1	1	—N(C$_2$H$_5$)$_2$	—N(C$_2$H$_5$)$_2$	8.96	8.32
26[a]	1	1	—N(C$_3$H$_7$)$_2$	—N(C$_3$H$_7$)$_2$	7.8	7.26
27[a]	1	1		—NH(cyclohexyl)	6.54	6.86
28[a]	1	1		—N(CH$_3$)(cyclohexyl)	7.27	7.41
29[a]	1	1	—N(piperidine)	—N(4-methylpiperidine)	8.64	8.57
30[a]	2	2	—N(CH$_3$)$_2$	—N(CH$_3$)$_2$	8.06	7.91

[a]Test set

Descriptors calculation

To obtain a QSAR model, compounds are often represented by molecular descriptors. The numerical representation (often-called molecular descriptor) of the chemical structure is the most important factor affecting the quality of a QSAR model. The calculation process of the molecular descriptors employed in the present investigation is described as below: all molecules were entered into HyperChem[10] and pre-optimized using MM+ molecular mecha-

nics force field. A more precise optimization has been realized with the semi-empirical AM1 method. The molecular structures were optimized using the Polak–Ribiere algorithm until the root mean square gradient was 0.01. Then, 1481 theoretical molecular descriptors were calculated in the DRAGON program,[11] including a) 0D-constitutional descriptors, b) 1D-functional groups, atom centered fragments, c) 2D-topological descriptors, walk and path counts, connectivity index, information index, various auto-correlations from the molecular graph, edge adjacency indices, descriptors of the Burden eigenvalues, topological charge indices, eigenvalues-based indices, d) 3D-Randic molecular profiles, geometrical descriptors, Weighted Holistic Invariant Molecular descriptors (WHIMs), Geometry, Topology and Atom-Weights AssemblY (GETAWAY) descriptors, e) charge descriptors and f) molecular properties (calculated from models, together with some empirical descriptors). The list and meaning of the molecular descriptors can be found from the DRAGON package, and the calculation procedure is explained in detail in the literature.[12] The theoretical descriptors were reduced by the following procedures: first, descriptors that were constant were eliminated. Second, to decrease the redundancy existing in the descriptors, the correlations of the descriptors with each other and with the pK_B values of the molecules were additionally examined, whereby collinear descriptors ($R > 0.9$) were detected. Among the collinear descriptors, the one having the highest correlation with pK_B was retained, and the others were removed from the data matrix.

Genetic algorithm (GA)

The genetic algorithm (GA) developed by Holland *et al.*[13] is a stochastic optimization technique that mimics selection in nature and has proved itself to be a very effective tool in QSAR studies with many merits.[14–17] The distinctive aspect of a GA is that it investigates many possible solutions simultaneously, each of which explores different regions in parameter space.[18] The first step is to create a population of *N* individuals. Each individual encodes the same number of randomly chosen descriptors. The fitness of each individual in this generation is determined. In the second step, a fraction of the children of the next generation is produced by crossover (crossover children) and the rest by mutation (mutation children) from the parents based on their scaled fitness scores. A new offspring contains characteristics from two or one of its parents. In this study, GA and MLR were combined to build a QSAR model. The fitness score used herein was the leave-one-out (LOO) cross-validated correlation coefficient (Q^2).

Model validation and applicability domain

A good fit alone does not guarantee that the model is useful for prediction purposes. Some kind of validation is necessary to test how stable the model is and how well it predicts. In the current work, several statistic terms, such as correlation coefficient (R^2), leave-one-out (LOO) cross-validated Q^2, root mean squared error (*RMSE*), were used to assess the internal predictive ability of the proposed models.

Another validation technique is the bootstrap technique.[19] By this technique, validation is performed by randomly generating training sets with sample repetitions and then evaluating the predicted responses of the samples not included in the training set. The bootstrapping was repeated 5000 times for each validated model.

In addition, *Y*-scrambling techniques were employed to exclude the possibility of chance correlation and to check for reliability and robustness by permutation testing. Multi-collinearity between the selected descriptors was detected by calculating their variation inflation factors (*VIF*), which can be calculated as follows:

$$VIF = \frac{1}{1 - r^2} \qquad (1)$$

where *r* is the correlation coefficient of the multiple regression between one variable and the others in the model. If *VIF* equals 1.0, no intercorrelation exists for each variable; if *VIF* falls into the range 1.0–5.0, the related model is acceptable; and if *VIF* is larger than 10.0, the related model is unstable and recheck is necessary.[20] To examine the relative importance as well as the contribution of each descriptor in the model, the value of the mean effect (*MF*) was calculated for each descriptor. This calculation was performed using the equation:

$$MF_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j^m \beta_j \sum_i^n d_{ij}} \tag{2}$$

where MF_j represents the mean effect for the considered descriptor *j*, β_j is the coefficient of the descriptor *j*, d_{ij} stands for the value of the target descriptors for each molecule and *m* is the descriptors number in the model. The *MF* value indicates the relative importance of a descriptor, compared with the other descriptors in the model. Its sign exhibits the variation direction in the values of the activities resulting from an increase (or a reduction) of this descriptor value.

The Williams plot, a plot of the standardized residuals *vs*. the leverage, was exploited to visualize the applicability domain.[21] The leverage indicates the distance of a compound from the centroid of *X*. The leverage of a compound in the original variable space is defined as:[22]

$$h_i = x_i^T \left(X^T X \right)^{-1} x_i \tag{3}$$

where x_i is the descriptor vector of the considered compound and *X* is the descriptor matrix derived from the descriptor values of the training set. The warning leverage (h^*) is defined as:[23]

$$h^* = 3(p+1)/n \tag{4}$$

where *n* is the number of training compounds and *p* is the number of predictor variables. A compound with $h_i > h^*$ seriously influences the regression performance, but it does not appear to be an outlier because its standardized residual may be small, even though it has been excluded from the applicability domain. Moreover, a value of 3 for a standardized residual is commonly used as a cut-off value for accepting predictions, because points that lie ±3 standardized residual from the mean cover 99 % of the normally distributed data.[24] Thus, the leverage and the standardized residual were combined for the characterization of the applicability domain.

RESULTS AND DISCUSSION

The total data set was divided into a training set of 24 compounds to develop the models and a test set of 6 compounds. In order to select the training and test sets, the range of the activity values of both the training set and test set should be covered from the lowest to the highest. The two sets are indicated in Table I. After splitting the data set into a training set and a test set, the next step was to select the main factors that were the most important for the antagonist potency of the biphenyl derivatives. The genetic algorithm was performed to select descriptors correlated to the activity based on the training set samples and various models with various numbers of descriptors were obtained. To select the optimum number of descriptors, the influence of the number of the descriptors was investi-

gated for one to six descriptors. The influence of the number of descriptors on the coefficients of determination (R^2), the adjusted R^2 (R^2_{adj}),[25] the standard deviation (*s*), and the coefficient of the leave one out cross-validation (Q^2) for the training set are shown in Fig. 1.
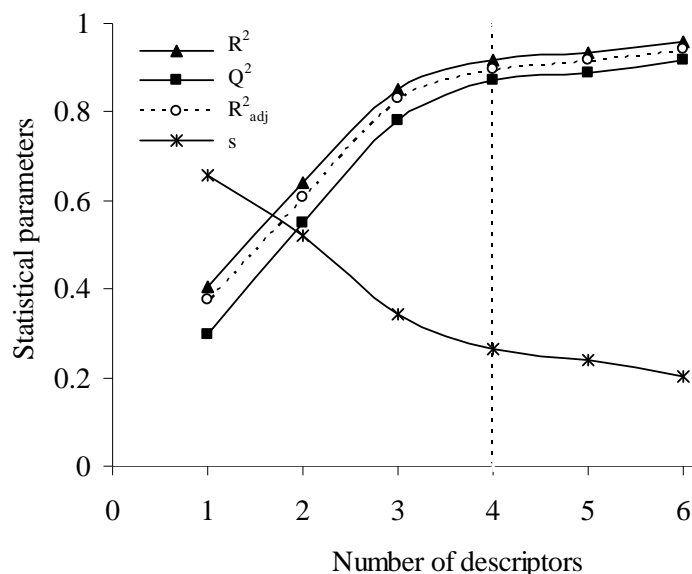

Fig. 1. Influence of the number of descriptors on the statistical parameters.

As can be seen, the models with 5 and 6 descriptors did not improve significantly the statistics of the model; it was determined that the optimum subset size had been achieved with a maximum 4 descriptors. The four most significant descriptors according to the GA–MLR algorithm are given in Table II. The genetic algorithm–MLR analysis led to the derivation of one model, with four variables, which is described by the following equation:

$$pK_B = -11.023\ (\pm1.602) + 1.042\ (\pm0.131)\ \text{IDDE} + 4.085\ (\pm0.813)\ \text{MATS3v} +$$
$$+ 0.688\ (\pm0.089)\ \text{GATS2e} + 24.364\ (\pm2.687)\ \text{RARS} \tag{5}$$
$$N_{train} = 24,\ R^2_{train} = 0.916,\ F = 51.771,\ RMSE_{train} = 0.0482,\ Q^2_{LOO} = 0.872$$
$$Q^2_{LGO} = 0.847,\ Q^2_{BOOT} = 0.857,\ R^2_{test} = 0.855,\ RMSE_{test} = 0.1537$$

In this equation, N is the number of compounds, R^2 is the squared correlation coefficient, Q^2_{LOO}, Q^2_{LGO} and Q^2_{BOOT} are the squared cross-validation coefficients for leave one out, leave group out and bootstrapping, respectively, *RMSE* is the root mean square error and F is the Fisher F statistic. The figures in parentheses are the standard deviations.

Then, the built model was used to predict the test set data. The prediction results are given in Table I. The predicted values of pK_B for the compounds in

the training and test sets using Eq. (1) are plotted *vs.* the experimental values in Fig. 2. As can be seen, the predicted values of pK_B are in good agreement with those of the experimental values.

TABLE II. The linear model based on the four parameters selected by genetic algorithm method (*VIF* – variation inflation factor)

| Descriptor	Chemical meaning	Coefficient	Mean effect	*VIF*
Constant	Intercept	–11.023	–	–
IDDE	Mean information content on the distance degree equality topological descriptors	1.042	0.202	1.227
MATS3v	Moran autocorrelation - lag 3 / weighted by atomic van der Waals volumes 2D autocorrelations	4.085	–0.020	1.301
GATS2e	Geary autocorrelation – lag 2 / weighted by atomic Sanderson electronegativities 2D autocorrelations	0.688	0.122	1.527
RARS	R matrix average row sum GETAWAY descriptors	24.364	0.696	1.307
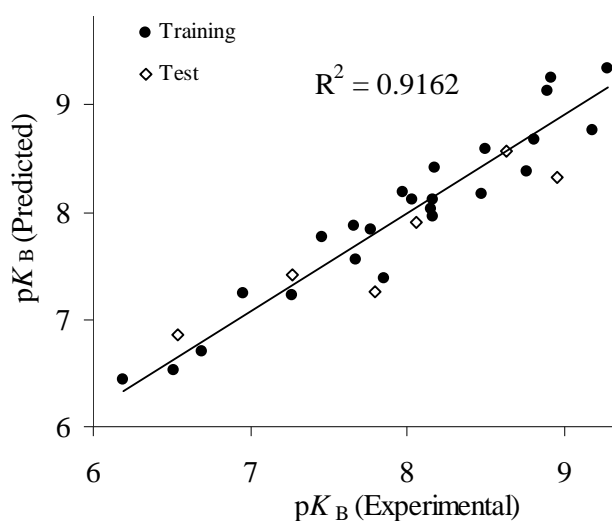
Fig. 2. Predicted *vs.* experimental antagonist potency (pK_B) by the GA–MLR methodology.

The chemical meaning and the mean effect (*MF*) of selected descriptors are given in Table II. In addition, the multi-collinearity between the above six descriptors was detected by calculating their variation inflation factors (*VIF*), which are also given in Table II. As can be seen from this table, most variables have *VIF* values of less than 5, indicating that the obtained model has obvious statistical significance. The real usefulness of QSAR models is not just their ability to reproduce known data, verified by their fitting power (R^2), but more importantly,

their possibility of predictive application. For this reason, the model obtained was validated using Leave-One-Out (LOO) and Leave-Group-Out (LGO) cross-validation process. For LOO cross-validation, a data point is removed from the set and the model is recalculated. The predicted activity for that point is then compared to its actual value. This is repeated until each data point has been omitted once. For LGO, 20 % of the data points are removed from the dataset and the model is refitted; the predicted values for those points are then compared to their experimental values. Again, this is repeated until each data point has been omitted once. The internal predictive ability of the model was also verified using the bootstrap Q^2_{BOOT} procedure, as is strongly recommended for QSAR modeling. The robustness of the proposed model and its predictive ability was guaranteed by the high value of Q^2_{BOOT} based on the bootstrapping being repeated 5000 times.

The cross-validation parameters are shown by Eq. (1). The cross-validation results confirmed that the obtained regression model has a good internal and external predictive power. In addition, in order to assess the robustness of the model, the Y-randomization test was applied in this contribution.[8] The dependent variable vector (pK_B) was randomly shuffled and a new QSAR model was developed using the original independent variable matrix. The new QSAR models (after several repetitions) are expected to have low R^2 and Q^2_{LOO} values. If the opposite is the case, then an acceptable QSAR model cannot be obtained for the specific modeling method and data. The results of Y-randomization tests are shown in Table III. The applicability domain of this model was also evaluated by leverage analysis expressed as a Williams plot (Fig. 3), in which the standardized residuals and the leverage values (h) were plotted. From this figure, it is obvious that there are no chemicals with a leverage higher than the warning h^* value of 0.625. Additionally, for all the compounds in the training and test sets, their standardized residuals were smaller than three standard deviation units (3δ), which means a coverage of more than 99 % of the training compounds. Thus, there are no outliers for the developed QSAR model (GA–MLR). Consequently, this GA––MLR approach currently constitutes the most accurate method for predicting the antagonist potency of biphenyl derivatives as histamine (H3) receptors.

Molecular descriptor interpretations

By interpreting the descriptors contained in the model, it is possible to gain some insight into the factors that are related to the antagonist potency of biphenyl derivatives. For this reason, an acceptable interpretation of the QSAR results is provided below.

The molecular descriptors selected by the genetic algorithm are listed in Table II. IDDE (mean information content on the distance degree equality) is one of the topological descriptors that appear in the model. The mean information content derives from the pruning partition of acyclic graphs. The mean information content on atomic composition is the mean value of the total information

content. As is apparent from Table II, the IDDE mean effect has a positive sign, which indicate that antagonist potency of histamine (H3) receptors is directly related to this descriptor. Therefore, increasing the value of this descriptor leads to an increase in antagonist potency (pK_B value).

TABLE III. R^2_{train} and Q^2_{LOO} values after several *Y*-randomization tests

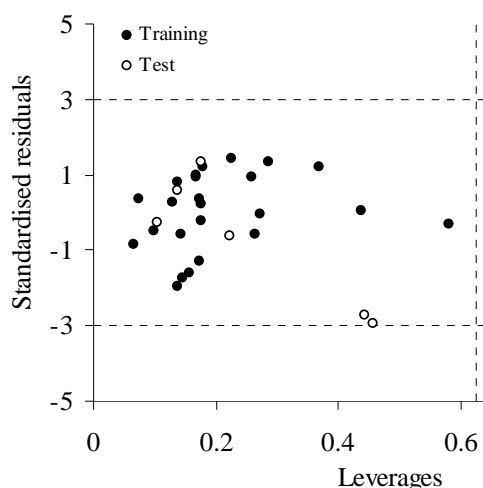| Iteration	R^2_{train}	Q^2_{LOO}
1	0.269	0.019
2	0.137	0.067
3	0.117	0.124
4	0.040	0.392
5	0.265	0.003
6	0.070	0.199
7	0.271	0.040
8	0.188	0.002
9	0.200	0.005
10	0.160	0.004

Fig. 3. Williams plot of the GA–MLR model. The training and test set samples are labeled differently. The dashed lines are the 3δ limit and the warning value of hat ($h^* = 0.625$) Compounds with high *h* values are numbered as in Table I.

MATS3v (Moran autocorrelation – lag 3 / weighted by atomic van der Waals volumes) is the second descriptor appearing in the model. It is one of the 2D autocorrelations descriptors. In this descriptor, the Geary coefficient is a distance-type function, the function is any physicochemical property calculated for each atom of the molecule, such as atomic mass, polarizability, *etc*.[12] Thus, the molecule atoms represent a set of discrete points in space and the atomic property is the function evaluated at these points. The physicochemical property in this case is the atomic van der Waals volumes. The MATS3v mean effect displays a negative sign, which indicates that the pK_B value is inversely related to this descriptor. Hence, it was concluded that by increasing the molecular volume the value of this descriptor increased, causing a reduction in the pK_B value.

The third descriptor is GATS2e (Geary autocorrelation – lag 2 / weighted by atomic Sanderson electronegativities), which is one of the 2D autocorrelations descriptors. The physico-chemical property in this case is atomic Sanderson electronegativities. The MATS3v mean effect displays a positive sign, which indicates that the pK_B value is directly related to this descriptor. Hence, by increasing the atomic Sanderson electronegativities, the value of this descriptor increases, causing an increasing in the pK_B value.

The final descriptor is RARS (R matrix average row sum) that belong to the GETAWAY descriptors. The GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) descriptors have been recently proposed as chemical structure descriptors derived from a new representation of molecular structure, the Molecular Influence Matrix (MIM).[12] The RARS descriptor has the highest mean effect value with a positive sign. It can be concluded that RARS displays a great effect in the model and the value of this descriptor is directly related to antagonist potency of histamine (H3) receptors.

CONCLUSIONS

In this work, an accurate and validated QSAR model using the GA–MLR method was developed for predicting the antagonist potency of biphenyl derivatives as histamine (H3) receptors. The proposed model has good stability, robustness and predictive ability, which were verified by internal validation (cross-validation by LOO, LGO, Bootstrap, and *Y*-scrambling) and external validation. The results indicate that the constructed QSAR model is more accurate and could be useful to understand the QSAR of biphenyl compounds as histamine (H3) receptors. The proposed model could identify and provide some insight into which structural features are related to the antagonist potency of biphenyl derivatives.

И З В О Д

QSAR СТУДИЈА АНТАГОНИСТА ХИСТАМИН (H3) РЕЦЕПТОРА ПРИМЕНОМ ГЕНЕТИЧКИН АЛГОРИТАМА И ВИШЕПАРАМЕТАРСКЕ ЛИНЕАРНЕ РЕГРЕСИЈЕ

MARYAM ADIMI[1], MAHMOUD SALIMI[2] и MEHDI NEKOEI[3]

[1]Chemical Engineering Department , Faculty of Engineering, Islamic Azad University – Farahan Branch, Farmahin, Iran, [2]Chemical Engineering Department, Faculty of Engineering, Islamic Azad University – Arak Branch, Arak, Iran u [3]Department of Chemistry, Shahrood branch, Islamic Azad University, Shahrood, Iran

Направљен је квантитативни модел за релацију између структуре и активитета (QSAR) за предвиђање антагонистичког учинка деривате бифенила на хумане хистамин (H3) рецепторе. Молекулска структура једињења нумерички је репрезентована различитим молекулским дескрипторима. Скуп података подељен је на тренинг сет и тест сет. Примењени су генетички алгоритми засновани на вишеструкој линеарној регресији и помоћу њих су изабрани статистички најефектнији дескриптори. Коначни QSAR модел ($N = 24$, $R^2 = 0{,}916$, $F = 51{,}771$, $Q^2_{LOO} = 0{,}872$, $Q^2_{LGO} = 0{,}847$ и $Q^2_{BOOT} = 0{,}857$) је у потпуности проверен стандардним статистичким поступцима. Тест сет поседује спољашњу моћ предвиђања од $R^2_{test} =$

= 0,855. У закључку, генерисани QSAR модел се може употребити као средство за дизајнирање нових антагониста хистаминских (Н3) рецептора.

REFERENCES

1. I. Dorronsoro, A. Castro; A. Martinez, *Expert. Opin. Ther. Patents* **13** (2003) 1725
2. D. Knopman, *Clin. Neuropharmacol*. **26** (2003) 314
3. M. E. Parsons, C. R. Ganellin, *J. Pharmacol.* **147** (2006) S127
4. J.-M. Arrang, M. Garbarg, J.-C. Schwartz, *Nature (London)* **302** (1983) 832
5. R. Leurs, R. A. Bakker, H. Timmerman, I. J. P. de Esch, *Nat. Rev. Drug Discov.* **4** (2005) 107
6. M. Wijtmans, R. Leurs, I. de Esch, *Expert Opin. Invest. Drugs* **16** (2007) 967
7. D. L. Hill, *The Biochemistry and Physiology of Tetrahymena*, 1st ed., Academic Press, New York, USA, 1972
8. A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Sci.* **22** (2003) 69
9. G. Morini, M. Comini, M. Rivara, S. Rivara, F. Bordi, P. V. Plazzi, L. Flammini, F. Saccani, S. Bertoni, V. Ballabeni, E. Barocelli, *Bioorgan. Med. Chem.* **16** (2008) 9911
10. HyperChem Release 7, HyperCube, Inc., http://www.hyper.com
11. R. Todeschini, Milano Chemometrics and QSPR Group, http://www.disat.unimib.it/chm
12. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, Germany, 2000
13. J. H. Holland, in *Adaptation in natural and artificial systems*, University of Michigan, Ann Arbor, MI, 1975
14. D. Rogers, A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **34** (1994) 854
15. A. Habibi-Yangjeh, E. Pourbasheer, M. Danandeh-Jenagharad, *Bull. Korean. Chem. Soc.* **29** (2008) 833
16. A. Habibi-Yangjeh, E. Pourbasheer, M. Danandeh-Jenagharad, *Monatsh. Chem.* **139** (2008) 1423
17. A. Habibi-Yangjeh, E. Pourbasheer, M. Danandeh-Jenagharad, *Monatsh. Chem.* **140** (2009) 15
18. H. M. Cartwright, *Applications of artificial intelligence in chemistry*. Oxford University Press, Oxford, 1993
19. R. Wehrens, H. Putter, L. M. C. Buydens, *Chemom. Intell. Lab. Systems* **54** (2002) 35
20. V. K. Agrawal, P. V. Khadikar, *Bioorg. Med. Chem.* **9** (2001) 3035
21. OECD, Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships [(Q)SAR] Models, Organisation for Economic Co-operation and Development, Paris, France, 2007
22. T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. T. D. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. J. M. van de Sandt, W. Tong, G. Veith, C. Yang, *Altern. Lab. Anim.* **33** (2005) 155
23. L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, P. Gramatica, *Environ. Health. Persp.* **111** (2003) 1361
24. J. S. Jaworska, J. N. Nikolova, T. Aldenberg, *Altern. Lab. Anim.* **33** (2005) 445
25. C. Hansch, J. Taylor, P. Sammes, *Comprehensive Medicinal Chemistry: The Rational Design, Mechanistic Study & Therapeutic Application of Chemical Compounds*, Vol. 6, Pergamon: New York, 1990, p. 1.