Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. 76 (8) 1117–1127 (2011) UDC 544.022:546.141:547–316:577.354.3
JSCS–4189 *Original scientific paper*

A quantitative structure–activity relationship study of tetrabutylphosphonium bromide analogs as muscarinic acetylcholine receptors agonists

MEHDI NEKOEI[1]*, MAHMOUD SALIMI[2], MOHSEN DOLATABADI[3] and MAJID MOHAMMADHOSSEINI[1]

[1]*Department of Chemistry, Faculty of Science, Islamic Azad University, Shahrood Branch, Shahrood,* [2]*Department of Chemical Engineering, Faculty of Engineering, Islamic Azad University of Arak, Arak and* [3]*Department of Chemistry, Faculty of Science, University of Birjand, Birjand, Iran*

(Received 22 November 2010)

Abstract: Quantitative structure–activity relationship (QSAR) of tetrabutylphosphonium bromide (TBPB) analogs as muscarinic acetylcholine receptors (mAChRs) agonists was studied. A suitable set of molecular descriptors was calculated and stepwise multiple linear regression (SW-MLR) was employed to select those descriptors that resulted in the best fitted models. A MLR model with three selected descriptors was obtained. Furthermore, the MLR model was validated using the leave-one-out (LOO) and leave-group-out (LGO) cross-validation, and the *Y*-randomization test. This model, with high statistical significance (R^2_{train} = 0.982, F = 388.715, Q^2_{LOO} = 0.973, Q^2_{LGO} = 0.977 and R^2_{test} = 0.986) could predict the activity of the molecules with a percentage prediction error lower than 5 %.

Keywords: QSAR; muscarinic receptor; TBPB; multiple linear regression.

INTRODUCTION

Muscarinic acetylcholine receptors (mAChRs) are members of the GPCR family A that mediate the metabotropic actions of the neurotransmitter acetylcholine (ACh).[1–3] To date, five distinct subtypes of mAChRs (M1–M5) have been cloned and sequenced. M1, M3 and M5 activate phospholipase C and calcium through Gq, whereas M2 and M4 block the action of adenylyl cyclase through Gi/o.[1–3] mAChR-regulated cholinergic signaling plays a critical role in a wide variety of CNS and peripheral functions, including memory and attention mechanisms, motor control, nociception, regulation of sleep wake cycles, cardiovascular function, renal and gastrointestinal function and many others.[4–6] As a re-

* Corresponding author. E-mail: m_nekoei1356@yahoo.com
doi: 10.2298/JSC101122102S

sult, agents that can selectively modulate the activity of specific mAChRs have therapeutic potential in multiple pathological states.[1–6]

Some novel tetrabutylphosphonium bromide (TBPB) analogs were recently reported as highly selective M1 allosteric agonists, which displayed robust efficacy in several preclinical antipsychotic models, as well as significant effects on the processing of amyloid precursor protein (APP) towards the non-amyloidogenic pathway and decreased Aβ production.[7,8]

Although there are several experimental methods available for screening the biological activity of chemicals (*e.g.*, *in vivo* and *in vitro* assay tests), and all have been performed using receptors and other biological materials of human, rat, mouse, and calf origin at least,[9] they are costly, time-consuming, and can potentially produce toxic side products from the experimental methods used today. This has meant that the development of computational methods as an alternative tool for predicting the properties of chemicals has been a subject of intensive study. Among the computational methods, quantitative structure–activity relationships (QSAR) have found diverse applications for predicting the properties of compounds, including biological activity prediction,[10] physical property prediction[11] and toxicity prediction.[12] 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 activity of interest.

In the present work, stepwise multiple linear regressions were employed for variable selection and model development in a QSAR analysis of the activity of some novel TBPB analogs. Finally, the accuracy of the proposed model was illustrated using leave-one-out (LOO) and leave-group-out (LGO) cross-validations and the *Y*-randomization technique.

DATA AND METHODOLOGY

The data used in this QSAR study consisted of the half maximal effective concentration (EC_{50}), which refers to the concentration of a TBPB analog that induces a response halfway between the baseline and maximum, that were reported by Bridges *et al.*[7,8] The activity data (EC_{50} (nM)) was converted to the logarithmic scale pEC_{50} (–log EC_{50} (M)) and then used for the subsequent QSAR analyses as the response variables. The *z*-matrices (molecular models) were constructed with HyperChem 7.0 and molecular structures were optimized using the AM1 algorithm.[13] In order to calculate the theoretical descriptors, the Dragon package version 2.1 was used.[14] For this propose 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 1481 theoretical descriptors were calculated for each compound in the data set (32 compounds).

The theoretical descriptors were reduced by the following procedure: 1) the descriptors that were constant were eliminated (308 descriptors); 2) in addition, to decrease redundancy existing in the descriptors, the correlation of the descriptors with each other and with the pEC_{50} of the molecules were examined, and collinear descriptors ($R > 0.9$) were detected. Among the collinear descriptors, the one that had the highest correlation with the pEC_{50} values was retained and the others were removed from the data matrix (736 descriptors).

Multiple linear regression analysis

The multiple linear regression method (MLR) is one of the most used modeling methods in QSAR. MLR regressions, a linear technique that can determine the relative importance of descriptors, are usually used to generate QSAR models. The MLR method provides an equation linking the structural features to the pEC_{50} of the compounds:

$$\mathrm{p}EC_{50} = a_0 + a_i d_i + \cdot\ldots\cdot + a_n d_n \quad (1)$$

where the intercept (a_0) and the regression coefficients of the descriptors (a_i) are determined using the least-squares method. d_i has the usual definition, variable or descriptor in this case. The elements of this vector are equivalent numerical values of a 3D structure of the molecules or the structural descriptors. The program used for the MLR analysis was written in Matlab 6.5.[15]

Stepwise multiple regression

The stepwise multiple regression technique based on forward selection was used to select the most appropriate descriptors.[16] The variable considered for inclusion at any step was the one yielding the largest single degree of freedom *F*-ratio among the variables that were eligible for inclusion. The variable was included only if this value was larger than a fixed value F_{in}. Consequently, at each step, the j^{th} variable was added to a *k*-size model if:

$$F_j = \max_j \left(\frac{RSS_k - RSS_{k+j}}{S^2_{k+j}} \right) > F_{\mathrm{in}} \quad (2)$$

In the above inequality, *RSS* is the residual sum of squares and *S* is the mean square error. The subscript *k*+*j* refers to quantities computed when the j^{th} variable was added to the *k* variables that were already included in the model.

Cross-validation technique

The consistency and reliability of a method can be explored using the cross-validation technique.[17] Two different strategies, leave-one-out (LOO) and leave-group-out (LGO), can be employed in this method. In the LOO strategy, by deleting each time one object from the training set, a number of models will be produced. Obviously, the number of models produced by the LOO procedure is equal to the number of available examples *n* (*n* = 26). The prediction error sum of the squares (*PRESS*) is a standard index to measure the accuracy of a modeling method based on the cross-validation technique. Based on the *PRESS* and *SSY* (sum of the squares of the deviations of the experimental values from their mean) statistics, Q^2 can be easily calculated from Eq. (3):

$$Q^2_{\mathrm{LOO}} = \frac{PRESS}{SSY} = 1 - \frac{\sum_{i=1}^{n} \left(y_{\mathrm{exp}} - y_{\mathrm{pred}} \right)^2}{\sum_{i=1}^{n} \left(y_{\mathrm{exp}} - \bar{y} \right)^2} \quad (3)$$

In the case of the LGO, *G* represents a group of randomly selected data points which would be left out at the beginning and would be predicted by the model which was developed using the remaining data points. Thus, *G* molecules are considered as a prediction set. The value of Q^2_{LGO} can be calculated using Eq. (4):

$$Q^2_{\mathrm{LGO}} = \frac{PRESS}{SSY} = 1 - \frac{\sum_{i=1}^{\mathrm{test}} \left(y_{\mathrm{exp}} - y_{\mathrm{pred}} \right)^2}{\sum_{i=1}^{\mathrm{test}} \left(y_{\mathrm{exp}} - \bar{y}_{\mathrm{train}} \right)^2} \quad (4)$$

It is usual to choose 20 % of the total number of molecules to be left out. Therefore, in the present work, five data points were removed from the data set and the model was refitted; the predicted values for those points were then compared to their experimental values. Again, this was repeated until each data point had been omitted once. The higher the Q^2_{LOO} or Q^2_{LGO}, the higher is the predictive power of the model.

Y-*randomization test*

This technique ensures the robustness of a QSAR model.[18] The dependent variable vector (biological action) is randomly shuffled and a new QSAR model is developed using the original independent variable matrix. The new QSAR models (after several repetitions) are expected to have low R^2 and Q^2 values. If the opposite is the case, then an acceptable QSAR model cannot be obtained for the specific modeling method and data.

RESULTS AND DISCUSSION

For the selection of the most important descriptors, the stepwise multiple regression technique based on forward selection was used. According to a rule of thumb, at least five data points (compounds) should be included in the equation for every parameter (descriptor). On the other hand, the ratio of 5 training molecules for each descriptor must be included in the equation. In order to investigate the optimum number of descriptors to be used in a model for modeling pEC_{50}, the statistical parameters (R^2 and Q^2) for 1–5 parameter models were calculated. The results showed that the models with 4 and 5 descriptors did not significantly improve the statistics of the models, which determined that the optimum subset size had been achieved with a maximum of 3 descriptors.

Since colinearity between the variables degrades the performance of MLR-based QSAR models, before a multiparametric analysis was undertaken, the correlations between every one of the variables used in this study were examined. The correlation matrix itself showed how the employed descriptors were mutually correlated. The correlation matrix obtained in the present case is shown in Table I, from which it could be seen that the correlation coefficient value of each pair of descriptors was less than 0.60, which meant that the selected descriptors were independent.

TABLE I. The correlation coefficient matrix for the descriptors used in this study

	ATS8m	C-028	SIC2
ATS8m	1		
C-028	0.43	1	
SIC2	0.16	–0.58	1

Predictive power of the model

In order to build and test the models, a data set of 32 compounds was randomly separated into a training set of 26 compounds, which was used to build model and a test set of 6 compounds, which was applied to test the built model.

With the selected two descriptors, a linear model was built using the training set data, and the following equation was obtained:

$$\mathrm{p}EC_{50} = 15.15(\pm0.34) - 5.88(\pm0.94)\mathrm{ATS8m} - 2.18(\pm0.11)\mathrm{C\text{-}028} \quad (5)$$
$$N = 26,\ R^2_{\mathrm{train}} = 0.964,\ Q^2_{\mathrm{LOO}} = 0.953,\ Q^2_{\mathrm{LGO}} = 0.956,$$
$$F = 310.829,\ R^2_{\mathrm{test}} = 0.984$$

In this and the following equations, N is the number of compounds, R^2 is the squared correlation coefficient, Q^2 is the squared cross-validation coefficient 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 statistical quality of Eq. (5) is good, but to gain better quality results, the MLR model was used with three descriptors and the following equation was obtained:

$$\mathrm{p}EC_{50} = 24.91(\pm2.18) - 3.79(\pm0.83)\mathrm{ATS8m} -$$
$$- 2.54(\pm0.12)\mathrm{C\text{-}028} - 13.62(\pm3.02)\ \mathrm{SIC2} \quad (6)$$
$$N = 26,\ R^2_{\mathrm{train}} = 0.982,\ Q^2_{\mathrm{LOO}} = 0.973,\ Q^2_{\mathrm{LGO}} = 0.977,$$
$$F = 388.715,\ R^2_{\mathrm{test}} = 0.986$$

The prediction results and relative error percentages (*RE* / %) are given in Table II, from which it can be seen that the calculated values of pEC_{50} are in good agreement with those of the experimental values and also the variables used in this equation can predict the activity of the molecules with a prediction error percentage of lower than 5 %. The values of pEC_{50} for the compounds in the training and test sets calculated using the Eq (6) are plotted *versus* the corresponding experimental values in Fig. 1. A plot of the residuals for the calculated values of pEC_{50} in training and test sets *versus* the corresponding experimental values is illustrated in Fig. 2. As can be seen, the model did not show proportional and systematic error because the propagation of the residuals on both sides of zero are random.

The model obtained was validated using the leave-one-out (LOO) and leave--group-out (LGO) cross-validation processes. With the 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 is omitted once. For LGO cross-validation, 20 % of the data points are removed from the data set and the model is refitted; the predicted values for these points are then compared to their experimental values. Again, this is repeated until each data point has been omitted once. The cross-validation parameters are shown in Eq. (1). The crossvalidated correlation coefficient (Q^2) is 0.973 for the LOO and 0.977 for the LGO cross-validations. These confirm that the obtained regression model has a good internal- and external-predictive power.

However, the small data size may produce an overfitted model. In order to assess the robustness of the model, the *Y*-randomization test was applied. The de-

pendent variable vector pEC_{50} was randomly shuffled and a new QSAR model was developed using the original variable matrix. The new QSAR model is expected to show a low value for R^2_{train} and Q^2_{LOO}. Several random shuffles of the *y* vector were performed, for which the results are shown in Table III. The low R^2_{train} and Q^2_{LOO} values show that the good results in the original model were not due to a chance correlation or a structural dependency of the training set.

TABLE II. Chemical structures and the corresponding observed and predicted pEC_{50} values by MLR method

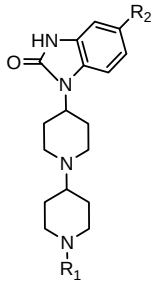
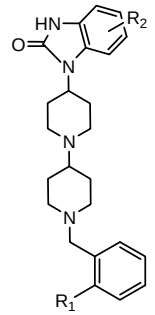
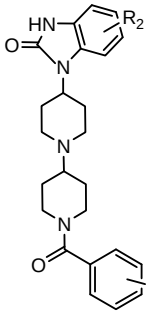
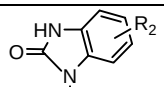
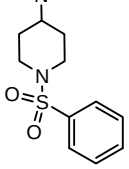
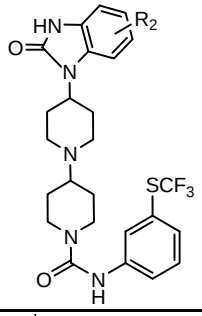
Compound No.	General structure	R_1	R_2	p$EC_{50(exp)}$	$MLR_{(pred)}$	RE[a] / %
1[b]		2-MeBn	H	6.54	6.69	2.29
2[b]		CO_2Et	H	8.68	8.72	0.46
3[b]		2-MeBn	Cl	5.99	5.92	–1.17
4[b]		Bn	Cl	6.45	6.37	–1.24
5[b]		2-CF_3Bn	H	6.39	6.48	1.41
6[c]		2-CF_3Bn	Cl	5.82	5.71	–1.89
7[c]		2-ClBn	H	6.59	6.55	–0.61
8[b]		2-ClBn	Cl	5.74	5.90	2.79
9[b]		2-NO_2Bn	H	6.92	6.73	–2.75
10[b]		2-NO_2Bn	Cl	5.96	5.92	–0.67
11[c]		2-CNBn	H	6.31	6.58	4.28
12[b]		2-CNBn	Cl	5.80	5.78	–0.34
13[b]		CH_3	4-F	8.97	9.01	0.45
14[c]		CH_3	5-F	9.07	9.05	–0.22
15[b]		CH_3	6-F	9.12	9.17	0.55
16[b]		CF_3	4-F	8.77	8.83	0.68
17[b]		CF_3	5-F	8.98	8.87	–1.22
18[b]		2-CH_3	H	9.06	9.18	1.32
19[c]		2-CH_3	Cl	8.57	8.50	–0.82
20[b]		3-CH_3	H	9.29	9.08	–2.26
21[b]		3-CH_3	Cl	8.68	8.45	–2.65
22[b]		2-CF_3	H	9.04	9.00	–0.44
23[b]		2-CF_3	Cl	7.95	8.32	4.65
24[b]		3-Cl	H	9.03	8.81	–2.44
25[c]		3-Cl	Cl	8.15	8.37	2.70

TABLE II. Continued

Compound No.	General structure	R_1	R_2	p$EC_{50(exp)}$	$MLR_{(pred)}$	RE / %[c]
26[b]		–	4-F	8.59	8.55	–0.47
27[b]		–	5-F	8.64	8.59	–0.58
28[b]		–	6-F	8.48	8.74	3.07
29[b]		–	5-Cl	8.39	8.53	1.67
		–	5-Br	8.36	8.34	–0.24
30[b]						
31[b]		–	6-F	8.02	8.21	2.37
32[b]		–	5-Cl	8.44	8.07	–4.38

[a]Relative error percentage; [b]training set; [c]test set

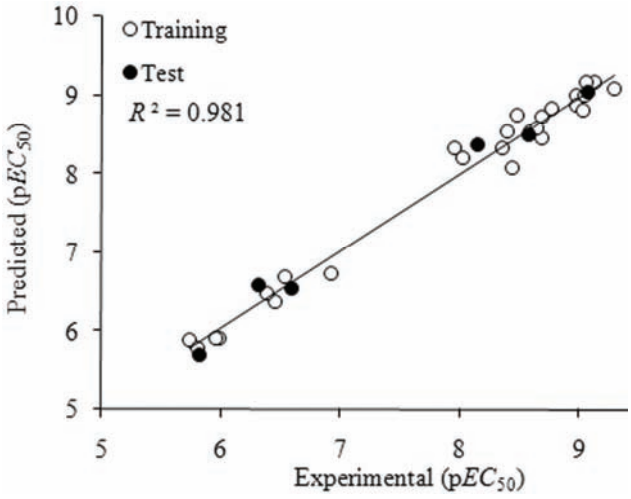


Fig. 1. The predicted *versus* the experimental pEC_{50} by MLR.

Besides demonstrating statistical significance, QSAR models should also provide useful chemical insights for drug design. For this reason, an acceptable interpretation of the QSAR results is provided below. By interpreting the descrip-

tors contained in the model, it is possible to gain some insight into factors which are related to the activity of the mAChRs selective compounds.

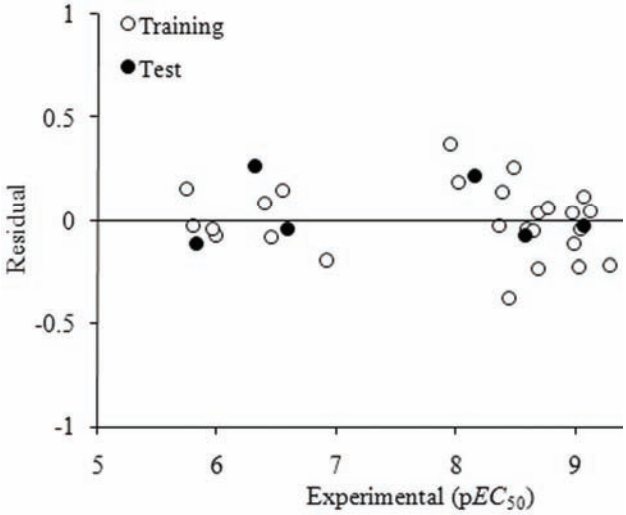


Fig. 2. The residual *versus* the experimental pEC_{50} by MLR.

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

Iteration	R^2_{train}	Q^2_{LOO}
1	0.156	0.004
2	0.087	0.059
3	0.399	0.219
4	0.108	0.009
5	0.108	0.010
6	0.279	0.064
7	0.347	0.171
8	0.149	0.002
9	0.171	0.002
10	0.077	0.058

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. The *MF* value indicates the relative importance of a descriptor in comparison with the other descriptors in the model. Its sign exhibits the variation direction in the values of the activities as a result of an increase (or reduction) of the value of this descriptor.

ATS8m is one of the 2D autocorrelation descriptors which appeared in the model. The 2D autocorrelation descriptor was successfully employed by Fernandez *et al.*[19,20] In these descriptors, the molecule atoms represent a set of discrete points in space, and the atomic property and function are evaluated at these

points. The symbol for each of the autocorrelation descriptors is followed by two indices *d* and *w*, whereby *d* stands for the lag, and *w* stands for the weight. Thus, for example, ATS8m means: the Broto–Moreau autocorrelation descriptor of lag 8 that is weighted by atomic mass. The lag is defined as the topological distance *d* between pairs of atoms. The topological distance between a pair of atoms (*i*,*j*) is given in the ij^{th} entry in the topological level matrix. The lag can have any value from the set {0,1,2,3,4,5,6,7,8}. The weight can be *m* (relative atomic mass), *p* (polarizability), *e* (Sanderson electronegativity) and *v* (Van der Waals volume). The relative mass is defined as the ratio of the atomic mass of an atom to that of carbon. Similarly, the other three weights *p*, *e* and *v* are scaled by the corresponding values for carbon. The physico–chemical property (weights) for ATS8m is atomic mass, which shows the mass of the atoms or molecules play the main role in this descriptor. The ATS8m mean effect (*MF* = 0.083) has a positive sign, which indicates that the pEC_{50} value is directly related to this descriptor. Hence, it was concluded that by increasing the molecular mass, the value of this descriptor increased, causing an increase of its pEC_{50} value.

C-028 is the second descriptor, appearing in the model. It is one of the atom-centered fragment descriptors that describes each atom by its own atom type and the bond types and atom types of its first neighbors.[21] The C-028 descriptor displays R-CR-X. This atom centered fragment descriptor is defined for each ring atom that has three neighbors. In this case, R-CR-X can be defined as a central carbon atom (C) on an aromatic ring that has one carbon neighbor (R) and one heteroatom neighbor (X) on the same aromatic ring and the third neighbor outside this ring is a carbon (R). The C-028 mean effect (*MF* = 0.346) has a positive sign, which indicates that the pEC_{50} value is directly related to this descriptor. Hence, it was concluded that by increasing the number of heteroatom (with R-CR-X format) in molecules the value of this descriptor increased, causing an increase of its pEC_{50} value.

The last descriptor appearing in the model is the second-order neighborhood structural information content (SIC2). It is calculated based on a hydrogen-depleted molecular graph and represents a measure of the structural complexity per vertex.[21] The positive sign of the descriptor mean effect (*MF* = 0.571) confirms that complex molecular structures with a diverse set of atoms in addition to carbon, such as nitrogen, oxygen, and halogens, which can establish covalent bonds in the lattice, have a high pEC_{50} value.

Summarizing, it is concluded that the mass, the number of heteroatoms, and the number of halogens play the main roles in determining the activity of mAChRs selective compounds.

CONCLUSIONS

Quantitative structure–activity relationships were applied on TBPB analogs as muscarinic acetylcholine receptors using the multiple parameter linear regression method. The separation of the data into two independent sets (training and test) showed that the obtained MLR model can predict external data with great accuracy. The proposed method, due to its high predictive ability, is a useful aid to costly and time consuming experiments for determining the activity of TBPB analogs. Selection of three variables, *i.e.*, ATS8m, C-028 and SIC2, by the stepwise multiple regression technique indicates the complexity of the activity mechanism. Synthesis of the new proposed molecules and their biological evaluation will show if this procedure can be used as a general rational drug discovery tool.

ИЗВОД

СТУДИЈА КВАНТИТАТИВНОГ ОДНОСА СТРУКТУРА–АКТИВНОСТ ЗА АНАЛОГЕ ТЕТРАБУТИЛФОСФОНИЈУМ-БРОМИДА КАО АГОНИСТА МУСКАРИНСКИХ АЦЕТИЛХОЛИНСКИХ РЕЦЕПТОРА

MEHDI NEKOEI[1], MAHMOUD SALIMI[2], MOHSEN DOLATABADI[3] и MAJID MOHAMMADHOSSEINI[1]

[1]*Department of Chemistry, Faculty of Science, Islamic Azad University, Shahrood Branch, Shahrood,* [2]*Department of Chemical Engineering, Faculty of Engineering, Islamic Azad University of Arak, Arak и* [3]*Department of Chemistry, Faculty of Science, University of Birjand, Birjand, Iran*

Проучаван је квантитативни однос структура–активност (QSAR) за аналоге тетрабутилфосфонијум-бромида (TBPB) као агониста мускаринских ацетилхолинских рецептора (mAChRs). Одређен је погодан скуп молекулских дескриптора и примењена поступна вишеструка линеарна регресија (SW-MLR). На тај начин су изабрани они дескриптори који дају најбоље слагање. Добијен је MLR модел са три дескриптора, који је даље тестиран помоћу LOO и LGO проступака, као и *Y*-насумичне расподеле. Овај модел, са добрим статистичким показатељима (R^2_{train} = 0,982, F = 388,715, Q^2_{LOO} = 0,973, Q^2_{LGO} =0,977, R^2_{test} = 0,986), омогућује да се предвиди активност молекула са грешком мањом од 5 %.

(Примљено 22. новембра 2010)

REFERENCES

1. T. I. Bonner, N. J. Buckley, A. C. Young, M. R. Brann, *Science* **237** (1987) 527
2. C. C. Felder, F. P. Bymaster, J. Ward, N. J. De-Lapp, *J. Med. Chem.* **43** (2000) 4333
3. F. P. Bymaster, D. L. McKinzie, C. C. Felder, J. Wess, *Neurochem. Res.* **28** (2003) 437
4. R. M. Eglen, A. Choppin, M. P. Dillon, S. Hedge, *Curr. Opin. Chem. Biol.* **3** (1999) 426
5. N. J. M. Birdsall, N. M. Nathanson, R. D. Schwarz, *Trends Pharm. Sci.* **22** (2001) 215
6. R. A. Eglen, A. Choppin, N. Watson, *Trends Pharm. Sci.* **22** (2001) 409
7. T. M. Bridges, A. E. Brady, J. P. Kennedy, R. N. Daniels, N. R. Miller, K. Kim, M. L. Breininger, P. R. Gentry, J. T. Brogan, C. K. Jones, P. J. Conn, C. W. Lindsley, *Bioorg. Med. Chem. Lett.* **18** (2008) 5439
8. N. R. Miller, R. N. Daniels, T. M. Bridges, A. E. Brady, P. J. Conn, C. W. Lindsley, *Bioorg. Med. Chem. Lett.* **18** (2008) 5443

9. D. L. Hill, *The Biochemistry and Physiology of Tetrahymena*, First Academic Press, New York, 1972
10. O. Deeb, B. Hemmateenejad, *Chem. Biol. Drug Des.* **70** (2007) 19
11. R. P. Verma, A. Kurup, C. Hansch, *Bioorg. Med. Chem.* **13** (2005) 237
12. V. K. Agrawal, P. V. Khadikar, *Bioorg. Med. Chem.* **9** (2001) 3035
13. *HyperChem Release 7*, HyperCube, Inc., http://www.hyper.com
14. R. Todeschini, Milano Chemometrics and QSPR Group, http://michem.disat.unimib.it/ /chm/ (accessed in August, 2011)
15. *Matlab 6.5*, Mathworks, 1984–2002
16. M. A. Efroymson, in *Mathematical Methods for Digital Computers*, A. Ralston, H. S Wilf, Eds.,Wiley, New York, 1960
17. D. W. Osten, *J. Chemom.* **2** (1988) 39
18. S. Wold, L. Eriksson, in *Chemometrics Methods in Molecular Design*, H. Van de Waterbeemd, Ed., VCH, Weinheim, 1995
19. M. Fernández, A. Tundidor-Camba, J. Caballero, *J. Chem. Inf. Model.* **45** (2005) 1884
20. J. Caballero, A. Tundidor-Camba, M. Fernández, *QSAR Comb. Sci.* **26** (2007) 27
21. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, 2000.