Quantitative Structure-Activity Relationship Studies of 4-Imidazolyl- 1,4-dihydropyridines as Calcium Channel Blockers

Farzin Hadizadeh [1*], Saadat Vahdani [2], Mehrnaz Jafarpour [3]

[1] Biotechnology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran
[2] Department of Chemistry, Islamic Azad University-North Tehran Branch, Tehran, Iran
[3] School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

ABSTRACT

Objective(s): The structure- activity relationship of a series of 36 molecules, showing L-type calcium channel blocking was studied using a QSAR (quantitative structure–activity relationship) method.

Materials and Methods: Structures were optimized by the semi-empirical AM1 quantum-chemical method which was also used to find structure-calcium channel blocking activity trends. Several types of descriptors, including electrotopological, structural and thermodynamics were used to derive a quantitative relationship between L-type calcium channel blocking activity and structural properties. The developed QSAR model contributed to a mechanistic understanding of the investigated biological effects.

Results: Multiple linear regressions (MLR) was employed to model the relationships between molecular descriptors and biological activities of molecules using stepwise method and genetic algorithm as variable selection tools. The accuracy of the proposed MLR model was illustrated using cross-validation, and Y-randomisation -as the evaluation techniques.

Conclusion: The predictive ability of the model was found to be satisfactory and could be used for designing a similar group of 1,4- dihydropyridines , based on a pyridine structure core which can block calcium channels.

Introduction

Voltage-gated calcium channels are transmembrane proteins which allow selective Ca^{2+} permeation in excitable cells, upon membrane depolarization. Voltage-gated calcium channels are heteromeric proteins consisting of the pore forming a1 subunit, disulfide-linked transmembrane complex of a2 and d subunits, intracellular b subunit and a subunit characteristic for skeletal muscle Ca^{2+} channels (1). Variability of regularity subunits distinguishes the tissue-specific calcium channel types L, N, T, P, Q and R (2). L-type Ca^{2+} channels are sensitive to numerous agonist and antagonist drugs that modulate the Ca^{2+} flow. Dihydropyridines (DHP) include both blocker and activators of L-type Ca^{2+} channels (3). Since their introduction as calcium channel blockers by Fleckenstein (4), these compounds have achieved special significance in the therapy of hypertension, angina pectoris and cardiovascular disease (5). Among the classes of calcium channel blockers, DHP derivatives are widely used. A quantitative structure–activity relationship (QSAR) study indicated that the potency of nifedipine analogues was dependent upon lipophilicity and electronic term and separate terms for each position on the aromatic ring (6). Making changes in the substitution pattern at C-3, C-4 and C-5 positions of nifedipine alter its potency (7), tissue selectivity (8, 9) and conformation of the 1,4-dihydropyridine ring (10). Our previous studies suggested that heterocyclic substituent like 1-substituted - alkylthioimidazol - 5-yl as bioisosteric replacement of nitrophenyl group at C-4 , enabled these compounds to have potent calcium antagonist activity (11-14). QSAR analysis is an effective method in the field of designing rational drugs and discovering the mechanism of drug actions. The fundamental hypothesis of the QSAR methodology is that the biological activity is a function of the molecular structure.

*Corresponding author: Farzin Hadizadeh. Biotechnology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. email: hadizadehf@mums.ac.ir

This method is used to find empirical relationships in a set of compounds (the instructional set) that are known to have interesting properties. Here, calcium channel blocking activity was the biological effect investigated. Such an approach to study the SAR consists of three basic stages. These are forming the instructional (investigational) set of compounds and selecting the descriptors. In addition, it is useful in areas like designing virtual compound libraries and optimizing computational-chemical of compounds. QSAR studies can express the biological activities of compounds as a function of their various structural parameters and also describes how the variation in biological activity depends on changes in the chemical structure (15). Recently, a QSAR study of biological activity has been published by our research team (16-18). If such a relationship can be derived from the structure-activity data, the model equation allows medicinal chemists conclude with an agreeable degree of confidence which properties are determing in the mechanism of drug action. The success of a QSAR study depends on choosing robust statistical methods for producing the predictive model and also the relevant structural parameters for expressing the essential features within those chemical structures. Nowadays, genetic algorithms (GA) are well known as interesting and widely used methods for variable selection (19-25). GA are stochastic methods used to solve the optimization problems defined by the fitness criteria, applying the evolutionary hypothesis of Darwin and also different genetic functions i.e. crossover and mutation. In the present work, we have used a genetic algorithm for the variable selection and developed an MLR model for the QSAR analysis of the 1, 4- dihydropyridines compounds. In a QSAR study, the model must be validated for its predictive value before it can be used to predict the response of additional chemicals. Validating QSAR with external data (i.e. data not used in the model development), although demanding, is the best method for validation. Finally, the accuracy of the proposed model was illustrated using leave one out, cross-validations and Y-randomisation techniques.

Materials and Methods
Data set
In this study, the data set of 1,4- dihydropyridines constitutes a group of small organic compounds based on a core pyridine structure which can both block and enhance calcium currents. (10-12). The inhibitory activity values are expressed as the half maximal inhibitory concentration (IC_{50}). The chemical structures and activity data for the complete set of compounds are presented in Table 1. The activity data [IC_{50} (μM)] was converted to the logarithmic scale pIC_{50} [-log IC_{50} (M)] and then used for the subsequent QSAR analyses as the response variables (26). The data set was randomly divided into two subsets: the training set containing 29 compounds (80%) and the test set containing 7 compounds (20%). The training set was used to build a regression model and the test set was used to evaluate the predictive ability of the obtained model.
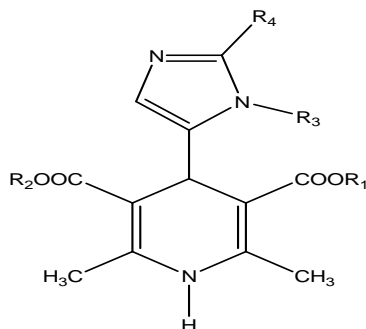
Structure entry and optimization
All of the molecules were drawn into the HyperChem software (Version 7.0 Hypercube, Alberta, Canada) and pre-optimized using the MM+ molecular mechanics force field. Then, a more precise optimization was performed with the semi-empirical AM1 method in HyperChem (27). The molecular structures were optimized using the Polak–Ribiere algorithm until the root mean square gradient reached 0.01.

Molecular descriptor generation
The Dragon packages (28) were used for calculating the molecular descriptors. The molecular structures were saved by the HIN extension and entered in the DRAGON software for the calculation of the 18 different types of theoretical descriptors for each molecule. They included (a) 0D-constitutional (atom and group counts); (b) 1D-functional groups, 1D-atom centered fragments; (c) 2D-topological, 2DBCUTs, 2D-walk and path counts, 2D-autocorrelations, 2D-connectivity indices, 2D-information indices, 2D-topological charge indices, and 2D-eigenvalue-based indices; and (d) 3D-Randic molecular profiles from the geometry matrix, 3D-geometrical, 3D-WHIM, and 3D-GETAWAY descriptors. These descriptors could represent a variety of aspects of the compounds and have been successfully used in various QSAR and quantitative structure-property relationship (QSPR) researches. Any descriptors with a constant or almost constant value for all the molecules were eliminated. Also, any pairs of variables with a correlation coefficient greater than 0.90 were classified as inter-correlated and only one of them were considered in developing the model. A total of 557 descriptors were considered for further investigations after discarding the descriptors with constant values and the ones that were inter-correlated.

Genetic algorithm
Genetic algorithms (GAs) are governed by biological evolution rules (29). These are stochastic optimization methods that have been inspired by evolutionary principles. The distinctive aspect of a GA is that it investigates many possible solutions, simultaneously, each of which explores a different region in the parameter of space (30). To select the most relevant descriptors, the evolution of the population was simulated (31, 32). The first generation population was randomly selected; each individual member in the population was defined by a chromosome of binary values and represented a subset of descriptors. The number of the genes at each chromosome was equal to the number of the descriptors. A gene was given the value of 1, if its corresponding descriptor was included in the subset; otherwise, it was given the value of zero. The number of genes with the value of 1 was kept relatively low to have a small subset of descriptors

Table1. Chemical structures and the corresponding observed and predicted pIC_{50} values as measured by the MLR method

No.	R_1	R_2	R_3	R_4	pIC_{50} Exp.	pIC_{50} Pred.	Ref.
1	CH_3	CH_3	CH_3	SO_2CH_3	0.86027	0.953305	10
2	C2H5	C_2H_5	CH_3	SO_2CH_3	0.789347	0.83378	10
3	n-Pr	n-Pr	CH_3	SO_2CH_3	0.574877	0.636895	10
4	n-Bu	n-Bu	CH_3	SO_2CH_3	0.667086	0.67542	10
5	n-pentyl	n-pentyl	CH_3	SO_2CH_3	0.875205	0.806819	10
6	Isobutyl	isobutyl	CH_3	SO_2CH_3	0.723134	0.670808	10
7	t-butyl	t-butyl	CH_3	SO_2CH_3	0.85661	0.825192	10
8	CH_3	CH_3	$4-FC_6H_4CH_2$	SCH_3	1.095585	1.014314	12
9	CH_3	CH_3	$C_6H_5CH_2$	SCH_3	0.461588	0.884273	11
10	C_2H_5	C_2H_5	$4-FC_6H_4CH_2$	SCH_3	1.020177	0.932849	12
11	CH_3	CH_3	$C_6H_5CH_2$	SC_2H_5	0.518944	0.543192	11
12	CH_3	CH_3	$4-FC_6H_4CH_2$	SC_2H_5	1.046083	0.563275	12
13	C_2H_5	C_2H_5	$C_6H_5CH_2$	SC_2H_5	0.41345	0.399285	11
14	C_2H_5	C_2H_5	$C_6H_5CH_2$	SCH_3	1.00993	1.058379	11
15	CH_3	CH_3	$2-ClC_6H_4CH_2$	SCH_3	1.01749	1.033185	12
16	C_2H_5	C_2H_5	$2-ClC_6H_4CH_2$	SCH_3	1.012685	1.039569	12
17	CH_3	CH_3	$2-ClC_6H_4CH_2$	SC_2H_5	1.00088	0.796317	12
18	C_2H_5	C_2H_5	$2-ClC_6H_4CH_2$	SC_2H_5	0.858992	0.524032	12
19	CH_3	C_2H_5	CH_3	SO_2CH_3	0.600871	0.592495	10
20	CH_3	n-pr	CH_3	SO_2CH_3	0.537459	0.557937	10
21	C_2H_5	n-pr	CH_3	SO_2CH_3	0.550631	0.568916	10
22	CH_3	i-Pr	CH_3	SO_2CH_3	0.567289	0.619903	10
23	CH_3	n-Bu	CH_3	SO_2CH_3	0.590333	0.6612	10
24	C_2H_5	n-Bu	CH_3	SO_2CH_3	0.619669	0.684053	10
25	CH_3	tert-Bu	CH_3	SO_2CH_3	0.55005	0.492167	10
26	C_2H_5	tert-Bu	CH_3	SO_2CH_3	0.570664	0.581974	10
27	CH_3	iso-bu	CH_3	SO_2CH_3	0.643257	0.556631	10
28	C_2H_5	iso-bu	CH_3	SO_2CH_3	0.722763	0.550418	10
29	CH_3	$CH_2C_6H_5$	CH_3	SO_2CH_3	0.905541	0.949922	10
30	C_2H_5	$CH_2C_6H_5$	CH_3	SO_2CH_3	0.72203	0.739644	10
31	CH_3	$CH_2CH_2C_6H_5$	CH_3	SO_2CH_3	1.017519	0.898361	10
32	C_2H_5	$CH_2CH_2C_6H_5$	CH_3	SO_2CH_3	0.799125	0.851237	10
33	CH_3	Cyclohexyl	CH_3	SO_2CH_3	0.611643	0.757533	10
34	C_2H_5	Cyclohexyl	CH_3	SO_2CH_3	0.587993	0.66771	10
35	CH_3	$CyclohexylCH_2$	CH_3	SO_2CH_3	0.661675	0.675148	10
36	C_2H_5	$CyclohexylCH_2$	CH_3	SO_2CH_3	0.676576	0.67648	10

(33). The genetic algorithm used in this paper is an evolution of the algorithm described in the reference #34, from which, parameters are reported in Table 2. Each wavelength subset selected in the spectrum will be represented by a p-dimensional vector and w, with binary coordinates. If the *ith* wavelength is selected then the *ith* coordinate of w is one, otherwise it is considered as zero. Each w is a chromosome. Given a chromosome (w), a MLR calibration is constructed using, from each spectrum, only the wavelengths represented by w. Each chromosome is evaluated using the PRESS (w) value reached in the calibration. The genetic algorithm searches for the minimum PRESS (w) in the space of all the possible chromosomes without establishing, a priori, the latent structure of the calibration.

Table 2. Parameters of the genetic algorithm

| |
|---|
| Population size: 30 chromosomes |
| In average, five variables per chromosome in the original population |
| Regression method: PLS |
| Response: cross-validated % explained variance (five deletion groups; the number of components is determined by cross validation) |
| Maximum number of variables selected in the same chromosome: 30 |
| Probability of mutation: 1% |
| Probability of crossover: 50% |
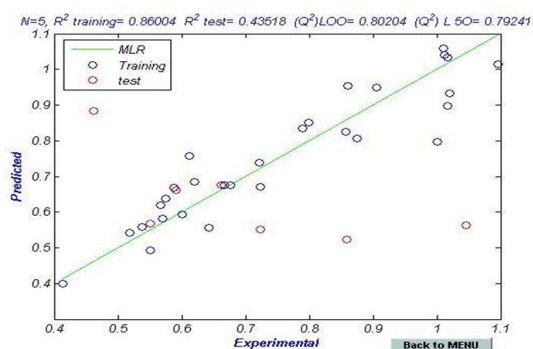
Figure 1. The predicted versus the experimental pIC_{50} measured by GA-MLR
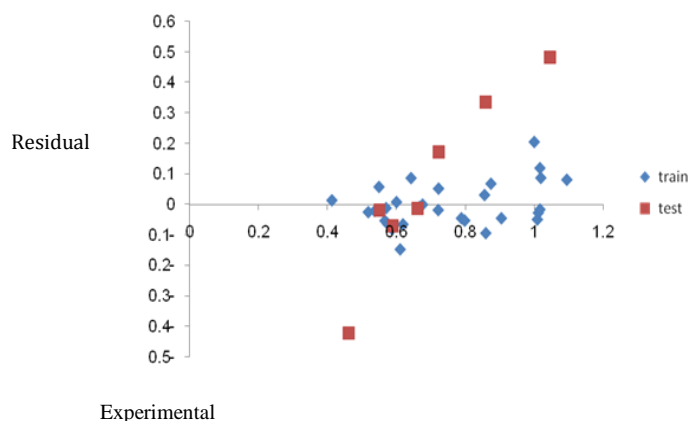

Figure 2. The residual versus the experimental pIC_{50} by measured GA-MLR.

Results

In a QSAR study, generally, the quality of a model is expressed by its fitting ability and prediction ability, from which the latter is more important. With the selected descriptors, we have built a linear model using the set data and the following equation was obtained.

$pIC_{50}=$ -2.1301(0.951352)+1.457617(0.585674) BELm6 -1.08595(0.233582) E1m -2.25419(0.369137)E2v+3.7547(1.007998) HATS8m+19.65472(4.570518) R2e+   (1)

$N_{train}=36$ $N_{test}=7$ $R^2_{train}= 0.86$ $R^2_{test}=0.435$
$R^2_{adj}=0.830$ $F_{train}=28.26$ $F_{test}= 0.04$ $Q^2_{LOO}=0.802$
$Q^2_{LGO}=0.792$ $Q^2_{BOOT}=0.796$ RMSE $_{train}$: 0.0715 RMSE $_{test}$: 0.2826

In this equation, N is the number of compounds, R^2 is the squared correlation coefficient, Q^2_{LOO}, Q^2_{BOOT} and $Q2_{LGO}$ are the squared cross-validation coefficients for leave one out, bootstrapping and external test set, respectively, RMSE is the root mean square error and F is the Fisher F statistic. The figures in parentheses are the standard deviations. The built model was used to predict the test set data and the prediction results are given in Table 1. As it is seen in Table 1, calculated values for the pIC_{50} are in good agreement
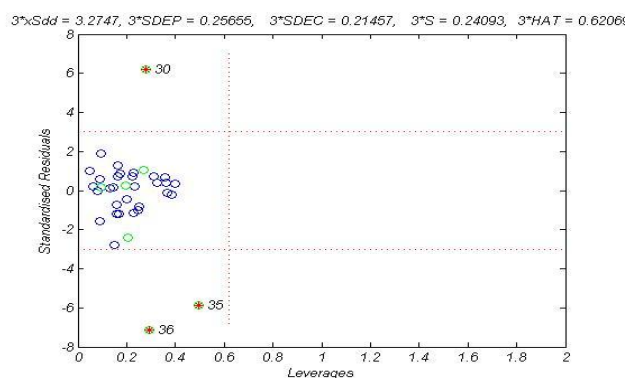

Figure 3. The William plot of the GA-MLR model

with those of the experimental values. The predicted values for pIC_{50} for the compounds in the training and test sets using equation 1 were plotted against the experimental pIC_{50} values in Figure 1. A plot of the residual for the predicted values of pIC_{50} for both training and test sets against the experimental pIC_{50}values are shown in Figure 2. Clearly, the model did not show any proportional and systematic error, because the propagation of the residuals on both sides of zero is random. The real usefulness of QSAR models is not just their ability to reproduce known data verified by their fitting power (R^2), but mainly it is their predictive application potential .

For this reason, the model calculations were performed by maximizing the explained variance in prediction verified by the leave-one-out cross-validated correlation coefficient (Q^2_{LOO}). To avoid the risk of over fitting and the possibility of overestimating the model predictivity by using Q^2_{LOO}, and Q^2_{LGO}, the internal predictive ability of the models was also verified using the bootstrap Q^2_{BOOT} procedure, as is strongly recommended for QSAR modeling. The robustness of the proposed models and their predictive ability was guaranteed by the high Q^2_{BOOT} based on the bootstrapping being repeated 5000 times. The Q^2_{LOO}, Q^2_{LGO} and Q^2_{BOOT} for the MLR model are shown in equation 1. This indicates that the obtained regression model has a good internal and external predictive power. Also, in order to assess the robustness of the model, the Y-randomization test was applied in this study. The dependent variable vector (pIC_{50}) was randomly shuffled and a new QSAR model developed using the original independent variable matrix. The new QSAR models (after several repetitions) would be expected to have low R^2 and Q^2_{LOO} values Table 3. If the opposite happens, acceptable QSAR model cannot be obtained for the specific modeling method and data. The Williams plot (Figure 3), the plot of the standardized residuals versus the leverage, was exploited to visualize the applicability domain (36). The leverage indicates a compound's distance from the centroid of X. The leverage of a compound in the original variable space is defined as (37, 38).

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

Where *xi* is the descriptor vector of the considered compound and X is the descriptor matrix derived from the training set descriptor values. The warning leverage (h*) is defined as:

$$h^* = \frac{3p}{n}$$

Where *n* is the number of calibration compounds, *p* is the number of model variables plus one. The leverage (*h*) greater than the warning leverage (*h**) suggested that the compound was very influential on the model.

The MLR analysis was employed to derive the QSAR models for different 1, 4- dihydropyridines. MLR and correlation analyses were carried out by the statistics software SPSS (Version 16.0) Table 4.

Table 3. The R^2 train and Q^2_{LOO} values after several Y-randomisation tests

No	Q^2	R^2
1	0.019534	0.287701
2	0.000557	0.242321
3	1.79E-05	0.228966
4	0.047437	0.119975
5	0.000167	0.19419
6	0.316241	0.060208
7	0.026796	0.127187
8	0.141785	0.09585
9	0.19683	0.063782
10	0.000406	0.215877

Discussion

After analyzing we spillted the data set into the training set and test set, the next step was to select the main factors which were the most important for the L-type calcium channel blocking inhibition activity of of 1,4- dihydropyridines. As we do not know yet which descriptors or which particular combinations are related to the studied response and can be used in the predictive models, we applied genetic algorithms as the variable selection procedure to select only the best combinations (most relevant) for obtaining the models with the highest

predictive power by using the training set. Five most significant descriptors according to the GA-MLR algorithm are lowest eigenvalue n. 6 of Burden matrix / weighted by atomic masses (BELm6), 1st component accessibility directional WHIM index / weighted by mass (E1m), 2nd component accessibility directional WHIM index / weighted by van der Waals volume (E2v), leverage-weighted autocorrelation of lag 8/weighted by mass (HATS8m) and R maximal autocorrelation of lag 2 / weighted by Sanderson electronegativity (R2e+).

The multi-collinearity between the above five descriptors were detected by calculating their variation inflation factors (VIF), which can be calculated as follows.

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

Where r is the correlation coefficient of the multiple regression between the variables in the model. If VIF equals 1, no inter-correlation exists for each variable; if VIF falls into the range of 1–5, the related model is acceptable; and if VIF is larger than 10, the related model is unstable and a recheck is necessary (39). The corresponding VIF values of the seven descriptors are shown in Table 5. Based on this table, most of the variables had VIF values of less than 5, indicating that the obtained model has statistical significance. 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 following equation.

$$MF_j = \frac{\beta \sum_{i=1}^{i=n} d_{ij}}{\sum_j^m \beta_j \sum_i^n \beta_{ij}}$$
(3)

Where *MF_j* epresents the mean effect for the considered descriptor *j*, *βj* is the coefficient of the descriptor *j*, *dij* tands for the value of the target descriptors for each molecule and eventually, *m* is the descriptors number for the model. The MF value indicates the relative importance of a descriptor, compared with the other descriptors in the model. Its sign (+, -) indicates the variation direction in the values of the activities as a result of the increase (or decrease) in the descriptor values. The mean effect values are shown in Table 5.

Table 4. The correlation coefficient existing between the variables used in different GA-MLR

	BELm6	E1m	E2v	HATS8m	R2e+
BELm6	1	0	0	0	0
E1m	-0.32678	1	0	0	0
E2v	0.023974	-0.20311	1	0	0
HATS8m	-0.59	0.688288	-0.16102	1	0
R2e+	-0.19761	0.223311	-0.35277	0.508854	1

Table 5. The linear model based on seven parameters selected by the GA-MLR method

Descriptor	Chemical meaning	MF[a]	VIF[b]
Constant	Intercept	0	0
BELm6	lowest eigenvalue n. 6 of Burden matrix / weighted by atomic masses	0.746010645	1.47149272
E1m	1st component accessibility directional WHIM index / weighted by mass	-0.138362847	1.889935834
E2v	2nd component accessibility directional WHIM index / weighted by van der Waals volume	-0.259229493	1.062263659
HATS8m	leverage-weighted autocorrelation of lag 8 / weighted by mass	0.22801657	3.166979174
R2e+	R maximal autocorrelation of lag 2 / weighted by Sanderson electronegativity	0.423565125	1.500297586

[a] Mean effect
[b] Variation inflation factors

Conclusion

In this article, a QSAR study of 36 molecules showing L-type calcium channel blocking activity was performed based on the theoretical molecular descriptors calculated by the DRAGON software. The built model was assessed comprehensively (internal and external validation) and all the validations indicated that the QSAR model built was robust and satisfactory and that the selected descriptors could account for the structural features responsible for the 1, 4 DHPs. The QSAR model developed in this study can provide a useful tool to predict the activity of new compounds and also to design new compounds with high calcium channel blocking activity.

Acknowledgment

References

1. Isom LL, De Jongh KS, Catterall WA. Auxiliary subunits of voltage-gated ion channels. Neuron 1994; 12:1183–1194.

2. Kim MS, Morii T, Sun LX, Imoto K, Mori Y. Structural determinants of ion selectivity in brain calcium channel. FEBS Lett 1993; 318:145–148.

3. Catterall WA, Striessnig J. Receptor sites for Ca2+ channel antagonists. Trends Pharmacol Sci 1992; 13:256–262.

4. Fleckenstein A. Specific pharmacology of calcium in myocardium, cardiac pacemakers, and vascular smooth muscle. Annu Rev Pharmacol Toxicol 1977; 17:149–166.

5. Weiner DA. Calcium channel blockers. Med Clin North Am 1988; 72:83–115.

6. Coburn RA, Wierzba M, Suto MJ, Solo AJ, Triggle AM, Triggle DJ. 1,4-Dihydropyridine antagonist activities at the calcium channel: a quantitative structure-activity relationship approach. J Med Chem 1988; 31:2103–2107.

7. Dagnino L, Li-Kwong-Ken MC, Wolowyk MW, Wynn H, Triggle CR, Knaus EE. Synthesis and calcium channel antagonist activity of dialkyl 1,4-dihydro-2,6-dimethyl-4-(pyridinyl)- 3,5-pyridinedicarboxylates. J Med Chem 1986; 29:2524–2529.

8. Iqbal N, Akula MR, Vo D, Matowe WC, McEwen CA, Wolowyk MW, Knaus EE. Synthesis, rotamer orientation, and calcium channel modulation activities of alkyl and 2-phenethyl 1,4-dihydro-2,6-dimethyl-3-nitro-4-(3- or 6-substituted- 2-pyridyl)-5-pyridinecarboxylates. J Med Chem 1998; 41:1827–1837.

9. Vo D, Matowe WC, Ramesh M, Iqbal N, Wolowyk MW, Howlett SE, et al. Syntheses, calcium channel agonistantagonist modulation activities, and voltage-clamp studies of isopropyl 1,4-dihydro-2,6-dimethyl-3-nitro-4-pyridinylpyridine-5- carboxylate racemates and enantiomers. J Med Chem 1995; 38:2851–2859.

10. Hemmateenejad B, Miri R, Safarpour MA, Khoshneviszadeh M, Edraki N. Conformational analysis of some new derivatives of 4-nitroimidazolyl-1,4-dihydropyridine-based calcium channel blockers. J Mol Struct (Theochem) 2005; 717:139–152.

11. Shafiee A, Dehpour AR, Hadizadeh F, Azimi M. Synthesis and calcium channel antagonist activity of nifedipine analogues with methylsulfonyl- imidazolyl substituent. Pharm Acta Helv 1998; 73:75-79.

12. Hadizadeh F, Fatehi Hassanabad M, Baghban-Golabadi B, Mohammadi M. Synthesis and calcium channel antagonist activity of 2-dimethylamino / 4-benzylimidazolyl substituted dihydropyridines. Bol Chim Farm 2005; 14:1-8.

13. Hadizadeh F, Imenshahidi M, Mohammadpour F, Mihanparast P, Serif M. Synthesis and calcium channel antagonist activity of 4-[(halobenzyl) imidazolyl] dihydropyridines. Saudi Pharm J 2009; 17:170-176.

14. Iwanami M, Shibanuma T, Fujimoto M, Kawai R, Tamazawa K, Takenaka T, et al. Synthesis of new water-soluble dihydropyridine vasodilators. Chem Pharm Bull (Tokyo) 1979; 27:1426– 1440.

15. Hansch C, Kurup A, Garg R, Gao H. Chembioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms. Chem Rev 2001; 101:619–672.

16. Hemmateenejad B, Miri R, Jafarpour M, Tabarzad M, Foroumadi A. Multiple linear regression and principal component analysis-based prediction of the anti-tuberculosis activity of some 2-aryi-1,3,4-thiadiazole derivatives. QSAR Comb Sci 2006; 25:56–66.

17. Riahi S, Pourbasheer E, Dinarvand R, Ganjali MR, Norouzi P. QSAR Study of 2-(1-Propylpiperidin-4-yl)-1H-Benzimidazole-4-Carboxamide as PARP Inhibitors for Treatment of Cancer. Chem Biol Drug Des 2008; 72:575–584.

18. Riahi S, Pourbasheer E, Ganjali MR, Norouzi P. Investigation of different linear and nonlinear chemometric methods for modeling of retention index of essential oil components: Concerns to support vector machine. J Hazard Mater 2009; 166:853-859.

19. Riahi S, Pourbasheer E, Ganjali MR, Norouzi P. Support Vector Machine Based Quantitative Structure–Activity Relationship Study of Cholesteryl Ester Transfer Protein Inhibitors. Chem Biol Drug Des 2009; 73:558-571.

20. Depczynski U, Frost VJ, Molt K. Genetic algorithms applied to the selection of factors in principal component regression. Anal Chim Acta 2000; 420:217-227.

21. Alsberg BK, Marchand-Geneste N, King RD. A new 3D molecular structure representation based on quantum topology with application to structure-property relationships. Chem Intell Lab Sys 2001; 54:75-91.

22. Jouan-Rimbaud D, Massart DL, Leardi R, De Noord OE. Anal Chem 1995; 67:4295-4301.

23. Riahi S, Ganjali MR, Pourbasheer E, Divsar F, Norouzi P, Chaloosi M. Development and validation of a rapid chemometrics-assisted spectrophotometry and liquid chromatography methods for the simultaneous determination of the phenylalanine, tryptophan and tyrosine in the pharmaceutical products. Curr Pharm Anal 2008; 4:231–237.

24. Riahi S, Pourbasheer E, Ganjali MR, Norouzi P, Zeraatkar Moghaddam A. QSPR study of the distribution coefficient property for hydantoin and 5-arylidene derivatives. A genetic algorithm application for the variable selection in the MLR and PLS methods. J Chin Chem Soc 2008; 55:1086–1093.

25. Riahi S, Ganjali MR, Moghaddam AB, Pourbasheer E, Norouzi P. Development of a new combined chemometrics method, applied in the simultaneous voltammetric determination of cinnamic acid and 3, 4-dihydroxy benzoic acid. Curr Anal Chem 2009; 5:42–47.

26. Riahi S, Pourbasheer E, Dinarvand R, Ganjali MR, Norouzi P. Exploring QSARs for antiviral activity of 4-alkylamino-6-(2-hydroxyethyl)- 2-methylthiopyrimi-dines by support vector machine. Chem Biol Drug Des 2008; 72:205–216.

27. Stewart JPP. MOPAC 6.0: Quantum Chemistry Program Exchange QCPE. No. 455. Bloomington, IN: Indiana University; 1989.p. 250–260.

28. Katritzky A, Karelson, M, Petrukhin, R. http://www.codessa-pro.com.

29. Aires-De-Sousa J, Hemmer MC, Gasteiger J. Prediction of H-1 nmr chemical shifts using neural networks. Anal Chem 2002; 74: 80-90.

30. Holland H. Adaption in Natural and Artificial Systems. Ann Arbor, MI: The University of Michigan; 1975.p. 342–375.

31. Cartwright HM. Applications of Artificial Intelligence in Chemistry. Oxford: Oxford University; 1993.p. 760–765.

32. Hunger J, Huttner G. Optimization and analysis of force field parameters by combination of genetic algorithms and neural networks. J Comput Chem 1999; 20:455–471.

33. Waller CL, Bradley MP. Development and Validation of a Novel Variable Selection Technique with Application to Multidimensional Quantitative Structure-Activity Relationship Studies. J Chem Inf Comput Sci 1999; 39:345-355.

34. Leardi R, Lupianez Gonzalez A. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chem Intell Lab Syst 1998; 41:195-207.

35. OECD. . Guidance Document on the Validation of (Quantitative) Structure– Activity Relationships [(Q) SAR] Models. . Paris: Organisation for Economic Co-Operation and Development; 2007.p.256–278.

36. Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, *et al.* ECVAM WORKSHOP REPORT Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. ATLA 2005; 33: 155-177.

37. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect 2003; 111:1361–1375.

38. Agrawal VK, Khadikar PV. QSAR prediction of toxicity of nitrobenzenes. Bioorg Med Chem 2001; 9:3035–3040.