Tropical Journal of Pharmaceutical Research August 2011; 10 (4): 483-490

Available online at http://www.tjpr.org
http://dx.doi.org/10.4314/tjpr.v10i4.14

Research Article

Quantitative Structure-Activity Relationship Analysis of the Anticonvulsant Activity of Some Benzylacetamides Based on Genetic Algorithm-Based Multiple Linear Regression

Amir Najafi*[1], Soheil S Ardakani[2] and Mehdi Marjani[3]

[1]Islamic Azad University, Hamedan Branch, Young Researchers Club, [2]Department of Environment, Islamic Azad University, Hamedan Branch, Hamedan, Iran, [3]Department of Clinical Sciences, Faculty of Veterinary Medicine, Karaj Branch, Islamic Azad University, Karaj, Iran

Abstract

Purpose: *To develop the quantitative structure-activity relationship (QSAR) for predicting the anticonvulsant activity of α-substituted acetamido-N-benzylacetamide derivatives.*
Methods: *AM1 semiempirical quantum chemical calculation method was used to find the optimum 3D geometry of the studied molecules. Two types of molecular descriptors, including the 2D autocorrelation and GETAWAY descriptors, were used to derive a quantitative relation between anticonvulsant activity and structural properties. The relevant molecular descriptors were selected by genetic algorithm-based multiple linear regression (GA-MLR) approach.*
Results: *The high value of the correlation coefficient, R^2 (0.900), indicate that the model was satisfactory.*
Conclusion: *The proposed model has good stability, robustness and predictability when verified by internal and external validation.*

Keywords: *Anticonvulsant, Benzylacetamides, 2D Autocorrelation, ,Quantitative structure-activity relationships, Multiple linear regression.*

Received: 10 October 2010 *Revised accepted: 15 May 2011*

Corresponding author:* ***E-mail:* *najafi@iauh.ac.ir, am.najafi@yahoo.com;* ***Tel:*** *+98-811-4494004;* ***Fax:*** *+98-811-4494143*

INTRODUCTION

Epilepsy, a common neurological disorder characterized by recurrent spontaneous seizures arising from excessive electrical activity in some portion of the brain, is a worldwide public health problem which affects approximately 1 % of the population [1]. Over the years, the field of epilepsy has received a great deal of attention from research investigators in the hope of discovering new drugs that are more effective and have minimal adverse effects. Though several new anticonvulsants have been introduced, some types of epilepsies are still not adequately controlled with the current therapy. Adverse reactions and lack of efficacy for certain types of epilepsies are some of the limitations of existing medications [2]. Antiepileptic drugs exert their action by different mechanisms. They include an enhancement of the GABA-ergic neurotransmission, effects on neuronal voltage-gated sodium and/or calcium channels [3]

Quantitative structure-activity relationships (QSAR), as a major factor in drug design, are mathematical equations relating chemical structure to their biological activity [4]. Anticonvulsant agents have been the aim of many SAR and QSAR studies [5-16]. Palludotto et al [5] synthesized a series of 2-aryl-2,5-dihydropyridazino[4,3-b]indol-3-one derivatives and tested them as central benzodiazepine receptor ligands . These workers used 2D and 3D QSAR on these molecules and observed that the molar refractivity (MR) of the substituents was the major factor controlling the binding of the ligands to their receptors. A correlation between the theoretical descriptors of the tricyclic neuro-active drugs and their biological mode of action has been obtained by the theoretical studies of Marone and coworkers [6]. Three dimensional QSAR analyses on the anticonvulsant activity of a series of cinnamamides using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) approaches have been reported by Hou et al. [7]. These investigators found that the interaction of these compounds with receptors is achieved by electrostatic and hydrophobic forces. Marder et al [8] have reported molecular modeling and QSAR analysis of flavone derivatives upon interaction with benzodiazepine binding site. The electronic properties of the ligands were found to be the major factor affecting the ligand-receptor binding [9].

In the present study, we aimed to develop QSAR equations for the anticonvulsant activity of a series of α- substituted acetamido-N-benzylacetamide drugs. We therefore used two types of molecular descriptors including 2D Autocorrelation and GETAWAY parameters to derive a quantitative relation between the anticonvulsant activity and structural descriptors obtained by genetic algorithm-based multiple linear regression (GA-MLR).

COMPUTATIONAL METHODS

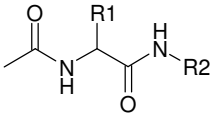
Activity data

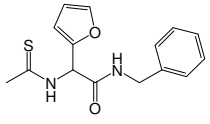
A data set containing α- substituted acetamido-N-benzylacetamide drugs was used in this study. The ED_{50} (mg/kg) data of the anticonvulsant activity, were evaluated by the maximal electroshock seizure (MES test) by Kohn *et al* [17], were taken from the paper of Jin *et al*.[11]. The supposed benzylacetamide anticonvulsant pharmaco-phore consists of a vicinal diamine linkage; an oxygen atom on the ethylene chain bridging two amino groups; and an aromatic ring one carbon removed from an amide group. The chemical structure and anticonvulsant activity of the studied molecules are included in Figure 1 and Table 1.

2D Autocorrelation approach

Three spatial 2D autocorrelation (2DAUTO) vectors were employed for modeling [18]:
Broto-Moreau's autocorrelation coefficients:

Table 1: Molecular structures of the anticonvulsant drugs



Derivative	R_1	R_2	Experimental anticonvulsant	Predicted GA-MLR
1	CH_3	CH_2-Ph	1.88	1.95
2 [a]	CH_3	CH_2-Ph-*m*-F	1.89	1.85
3	2-Furanyl	CH_2-Ph-*o*-F	1.60	1.41
4	2-Furanyl	CH_2-Ph-*m*-F	1.12	1.13
5 [a]	2-Furanyl	CH_2-Ph-*p*-F	1.10	1.17
6	2-Furanyl	CH_2-2,5-$F_2C_6H_3$	1.38	1.58
7	2-Furanyl	CH_2-2,6-$F_2C_6H_3$	1.80	1.80
8	3-Allyl	CH_2-Ph	1.53	1.58
9	2-Tetrahydrofuranyl	CH_2-Ph	1.71	1.56
10	Ph	CH_2-Ph	1.31	1.56
11	2-Furanyl	CH_2-Ph	1.01	1.07
12	2-Furanyl-5-CH_3	CH_2-Ph	1.28	1.20
13 [a]	2-Pyrrolyl	CH_2-Ph	1.21	1.36
14	2-Pyrrolyl-5-CH_3	CH_2-Ph	1.56	1.60
15 [a]	2-Thienyl	CH_2-Ph	1.65	1.54
16	3-Thienyl	CH_2-Ph	1.94	1.82
17 [a]	1-Pyrrole	CH_2-Ph	1.90	1.75
18	1-Pyrazole	CH_2-Ph	1.22	1.33
19	2-Pyridyl	CH_2-Ph	1.03	1.05
20	$C(S)NH_2$	CH_2-Ph	1.94	1.99
21	$NHCH_2CH_3$	CH_2-Ph	1.63	1.57
22	$N(CH_3)_2$	CH_2-Ph	1.66	1.81
23 [a]	$N(CH_3)OH$	CH_2-Ph	1.48	1.34
24	$NPhNH_2$	CH_2-Ph	1.63	1.57
25	OH	CH_2-Ph	1.90	1.98
26	OCH_2CH_3	CH_2-Ph	1.79	1.73
27 [a]	CH_2OCH_3	CH_2-Ph	0.92	0.97
28	$CH_2OCH_2CH_3$	CH_2-Ph	1.23	1.21
29 [a]	2-Pyrazinyl	CH_2-Ph	1.17	1.24
30	2-Pyrimidyl	CH_2-Ph	0.91	0.85
31 [a]	2-Oxazole	CH_2-Ph	1.02	0.91
32 [a]	2-Thiazole	CH_2-Ph	1.08	1.14
33	N(H)Ph(3-NH_2)	CH_2-Ph	1.99	1.82
34			1.26	1.28
35	2-Furan	CH_2-Ph	1.48	1.50

$$ATS_w l = \sum_{i=1}^{A} \sum_{j=1}^{A} \delta_{ij} w_i w_j \quad \text{..... (1)}$$

Moran's indices:

$$MATS_w l = \frac{N}{2L} \frac{\sum_{ij} \delta_{ij} (w_i - \overline{w})(w_j - \overline{w})}{\sum_{i} (w_i - \overline{w})^2} \quad \text{..... (2)}$$

Geary's coefficient:

$$GATS_w l = \frac{(N-1)}{4L} \frac{\sum_{ij} (w_i - \overline{w})(w_j - \overline{w})}{\sum_{i} (w_i - \overline{w})} \quad \text{..... (3)}$$

where $ATS_w l$, $MATS_w l$, and $GATS_w l$ are Broto- Moreau's autocorrelation coefficient, Moran's index, and Geary's coefficient at spatial lag *l*, respectively; where w_i and w_j are the values of any atomic property of atom *i* and *j* respectively; $\overline{w}$ is the average value of property; *L* is the number of nonzero values in the sum, *N* is the number of atoms in the molecule, and $\delta(l, d_{ij})$ is a Dirac-delta function defined as

$$\delta(l, d_{ij}) = \left\{ \begin{array}{lll} 1 & \text{if} & d_{ij} = l \\ 0 & \text{if} & d_{ij} \neq l \end{array} \right\} \quad \text{.....(4)}$$

where d_{ij} is the topological distance or spatial lag between atoms *i* and *j*.

The 2D Autocorrelation descriptors in general explain how the considered property is distributed along the topological structure. Autocorrelation vectors were calculated at spatial lags *l* ranging from 1 up to 8. The physicochemical property considered in the four different weighting schemes: atomic masses (m), atomic van der Waals volumes (v), atomic Sanderson electronegativities (e), and atomic polarizabilities (p). The autocorrelation descriptors are denoted by the scheme: type of descriptor-spatial lag-weighting property; for instance, GATS5p is the Geary autocorrelation of lag 5 weighted by atomic polarizabilities.

GETAWAY approach

The GETAWAY descriptors [19] are recently proposed molecular descriptors derived from a new representation of molecular structure, the *molecular influence matrix* (MIM), denoted by *H* and defined as the following:

$$H = M \cdot (M^T \cdot M)^{-1} \cdot M^T \quad \text{..... (5)}$$

where *M* is the molecular matrix constituted by the centered Cartesian coordinates *x*, *y*, *z* of the molecule atoms (hydrogens included) in a chosen conformation, and the superscript T refers to the transposed matrix.

On the other hand, matrix R, a symmetrical matrix whose elements resemble the single terms in the sums of the gravitational indices, is defined as

$$[R]_{ij} = \left[\frac{\sqrt{H_{ii} H_{jj}}}{r_{ij}} \right], \; i \neq j, \quad \text{..... (6)}$$

Where h_{ii} and h_{jj} are the leverages of the two considered atoms and r_{ij} their geometric distance. Obviously, the diagonal elements of matrix R are zero, and the largest values of its off-diagonal elements derive from the most external atoms (i.e., with high leverages) and simultaneously next to each other in the molecular space (i.e., small interatomic distance).

Finally notice that, in many of these H and R descriptors, the molecule atoms are weighted in such a way as to account for atomic mass, polarizability, van der Waals volume, and electronegativity, with the aim of incorporating relevant chemical information. Two sets of theoretically closely related molecular descriptors have been devised: H-GETAWAY descriptors have been calculated from the MIM H, while R-GETAWAY descriptors are from the *influence/distance matrix* R where the elements of the MIM are combined with those of the geometry matrix.

Model development

HyperChem software was used to draw the chemical structure of the molecules. AM1 semi-empirical quantum-chemical calculation was used to optimize the 3D geometry of the molecules. The geometry optimization was preceded by the Polak-Rebiere algorithm until the root mean square gradient reaches

0.01. Dragon [20] computer software was employed to calculate the 2DAUTO and GETAWAY molecular descriptors.

The calculation of weighted Broto-Moreau, Moran, and Geary 2DAUTO vectors was carried out at spatial lags ranging from 1 to 8. Four different weighting properties including atomic masses (m), atomic van der Waals volumes (v), atomic Sanderson electronegativities (e), and atomic polarizabilities (p) were used. Since we calculated three types of autocorrelation descriptors (Moran's index, Geary's coefficient, and Broto-Moreau's autocorrelation coefficient) weighted by 4 atomic properties at 8 lags, a total of 96 (3 × 4 × 8) descriptors for each compound were computed.

Two sets of GETAWAY descriptors including H-GETAWAY and the R-GETAWAY descriptors have been calculated. The first set consists of 107 H-GETAWAY descriptors calculated from the leverage matrix obtained by the centered atomic coordinates, that is, the molecular influence matrix (H), while the remaining 90 R-GETAWAY descriptors are calculated from the influence/ distance (R) matrix.

The calculated descriptors were gathered in a data matrix. First, the descriptors were checked for constant or near constant values and those detected were discarded from the original data matrix.

Then, the descriptors were correlated with each other and with the activity data. Among the collinear descriptors, one with the lowest correlation with anticonvulsant drug was removed from the data matrix. Multiple linear regressions (MLR) were used to derive the QSAR equation and feature selection was performed by the use of genetic algorithm (GA).

A genetic algorithm is a novel and simple optimization method based on the evolution process of beings that implicitly and effectively has been applied to the various types of optimization problems in many scientific fields. It is based on the simulation of natural genetics and evolutions. The genetic algorithm used was the same as that previously used [21,22]. Each individual of the population was defined by a chromosome of binary values representing a subset of descriptors. 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 at each chromosome was equal to the number of descriptors. The initial population was created randomly. The population size was varied between 50 and 250 for different GA runs. The resulting models were validated by leave one- out (LOO) cross-validation procedures to check their predictability and robustness. Table 2 presents the notation and a short description of the molecular descriptors used to generate the QSAR model.

RESULTS

By using the genetic algorithm-based multiple linear regression (GA-MLR) method, the

Table 2: Description of the molecular descriptors used in this study

Symbol	Definition	Class
MATS6e	Moran autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities	2D
MATS5p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities	2D
ATS6e	Broto-Moreau autocorrelation - lag 6 / weighted by atomic Sanderson electronegativities	2D
H7v	H autocorrelation of lag 7 / weighted by atomic van der Waals volumes	GETAWAY
HATS2u	leverage-weighted autocorrelation of lag 2 / unweighted	GETAWAY
H5m	H autocorrelation of lag 5 / weighted by atomic masses	GETAWAY

models were developed for 35 anticonvulsant analogues.

The correlations performed for the whole set provided the optimal equations for different numbers of descriptors in the range of 1-6. Figure 1 shows the plots of R^2, Q^2 and s^2 (squared standard deviation) as a function of the number of variables in the regression model.

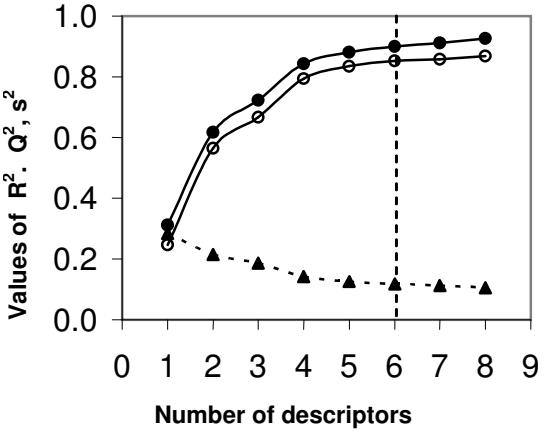


Figure 1: Influence of the number of descriptors on the correlation coefficient (R^2), the LOO cross-validation correlation coefficient (Q^2) and the square of standard error (s^2) of the regression model.. (●) R^2; (○) Q^2; (▲) s^2

It suggests that the best one-descriptor model with highest impact is MATS6e which is defined as Moran autocorrelation vector weighted by Sanderson electronegativities representing topological substructure of sizes 6 in the molecule. Subsequent addition of variables produces monotonously increasing values of R^2 and Q^2 and decreasing values s^2 and the break point is not clearly defined. We decided to select the best model to be the one having the smallest number of parameters and satisfactory statistical parameters. The model with 6 descriptors:

Log ED_{50} = 3.0647 (±1.1867) + 2.4601 (±0.2293) MATS6e + 12.5149 (±1.4119) H7v + 7.6953 (±1.2669) HATS2u + 3.1255 (±0.5477) MATS5p - 3.9054 (±1.0490) ATS6e + 0.9969 (±0.4395) H5m.....(7)

N = 35, R^2 = 0.900, SE = 0.118, RMS = 0.107, F = 41.963, Q^2 = 0.853.....(8)

Some statistical parameters such as squared correlation coefficients (R^2), standard error of estimation (SE), root-mean-square error (RMS), Fisher statistic ratio (F) and LOO cross-validation (Q^2) are given in the Eq. 8. As can be seen, the MLR model has good statistical quality with low prediction error. The predicted activities by using GA-MLR regression method are listed in Table 1.

The robustness of the model and their prediction ability for the anticonvulsant activity, were evaluated by both LOO cross-validation and external validation procedures. In order to estimate the predictive power of the GA-MLR model, an external validation test was performed by splitting the data into two sub-samples with one being used to fit (training set) and the other to test (test set). Models are generated based on training set compounds and predictive capacity of the models is judged based on the predictive R^2 values. Selection of the training set compounds is significantly important in QSAR analysis. One most widely used method for dividing a data set into training and test sets is mere random selection.

In the external validation procedure, 10 analogues are randomly selected and eliminated from the data set as unknown test samples. Then the training set generated using remaining 25 samples and log (ED_{50}) of eliminated samples are predicted using the MLR model. The test set is marked with superscript a in Table 1. A model based on training set is the following:

Log ED_{50} = 3.0590 (±1.2893) + 2.4863 (±0.2678) MATS6e + 12.4037 (±1.5221) H7v + 6.9874 (±1.4062) HATS2u + 2.5769 (±0. 7032) MATS5p - 3.6462 (±1.1633) ATS6e + 0.8814 (±0.5393) H5m …….. (9)

N = 25, R^2 = 0.893, SE = 0.118, RMS = 0.103, Q^2 = 0.823., F = 26.760 ..... (10)

Eq 10 was used to predict the anticonvulsant activity of the test set. The results are presented in Figure 2, which R^2 and RMS for

the test set are 0.898 and 0.119, respectively. The results illustrated once more that the GA-MLR technique is adequate to generate an efficient QSAR model for predicting the anticonvulsant activity of different compounds.

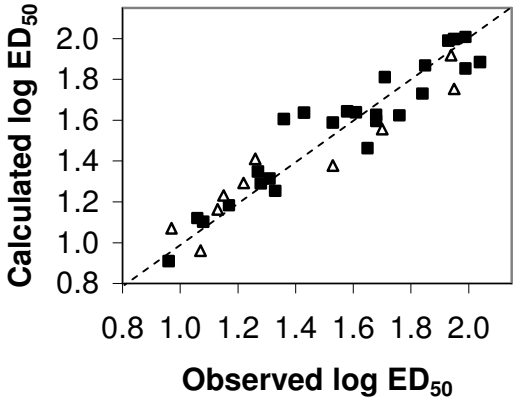


Figure 2: Plot of the anticonvulsant activity predicted by GA-MLR for calibration set (■) and validation set (Δ) against the experimental values. (*Note:* The solid and dashed lines are the calibration and validation regression lines, respectively, while the dotted line is an ideal fit with fit to the straight line)

DISCUSSION

The proposed QSAR model, due to the high predictive ability, can therefore act as a useful aid to the costly and time consuming experiments for determining the maximal electroshock seizure (MES test). We first tried to identify descriptors trends which lead to anticonvulsant activity based on the proposed QSAR equation. The QSAR model of α-substituted acetamido-N-benzylacetamide derivatives (Table 1) has shared six 2DAUTO and GETAWAY class descriptors (Table 2). As mentioned before, Moran autocorrelation lag 6 weighted by Sanderson electronegativities (MATS6e) is the most important variable for predicting the anticonvulsant activity. The remaining five descriptors involve the summations of different functions corresponding to the different fragment lengths and with polarizability (p), electronegativity (e), volume (v) and mass (m) as the weighting parameter (Table 2).

CONCLUSION

In summary, multivariate linear QSAR models were obtained by MLR method combined with genetic algorithm for variable selection (GA-MLR). We have shown that the 2D AUTO and GETAWAY descriptors are able to describe the anticonvulsant activity of different α- substituted acetamido-N-benzylacetamide derivatives. The proposed models have good stability, robustness and predictability when verified by internal validation (LOO-CV) and also external validation.

ACKNOWLEDGEMENT

The authors are grateful to the Young Researchers Club, Hamedan branch, Islamic Azad University, for providing facilities to conduct this study.

REFERENCES

1. OJ McNamara, JG Hardman, Limbird LE, Gilman AG, editors. In The Pharmacological Basis of Therapeutics. New York: McGraw-Hill; 2002; p 554.
2. Löscher W, Schmidt D. New horizons in the development of antiepileptic drug. Epilepsy Res 2002; 50: 3-16.
3. Anger T, Madge DJ, Mulla M, Riddall D. Medicinal Chemistry of Neuronal Voltage-Gated Sodium Channel Blockers. J Med Chem 2001; 44: 115-137.
4. Krogsgaard-Larsen P, Liljefors T, Madsen Ulf, editors. Textbook of Drug Design and Discovery. London: Taylor and Francis; 2002. p 554.
5. Palluotto F, Carotti A, Casini G, Campagna F, Genchi G, Rizzo M, De Sarro GB. Structure-activity relationships of 2-aryl-2,5-dihydropyridazino [4,3-b]indol-3(3H)-ones at the benzodiazepine receptor. Bioorg Med Chem 1996; 4(Pt.12): 2091-2104.
6. Marone S, Rozas I, Weaver DF. Theoretical structural analyses of tricyclic neuroactive drugs: quantum pharmacologic descriptors for clustering anticonvulsant, antidepressant, and antipsychotic activities. J Mol Struct (Theochem) 1999; 467: 25-30.
7. Hou T, Xu X. Three-Dimensional quantitative structure activity relationship analysis of some

cinnamamides. Chemom Intell Lab Syst 2001; 56: 123-132.
8. Marder M, Estiu G, Blanch LB, Viola H, Wasowski C, Medina JH. Molecular modelling and QSAR analysis of the interaction of flavone derivatives with the benzodiazepine site of GABAA receptor complex. Bioorg Med Chem 2001; 9: 323-335.
9. Verli H, Albuquerque MG, Bicca de Alencastro R, Barreiro EJ. Local intersection volume: a new 3D descriptor applied to develop a 3D-QSAR pharmacophore model for benzodiazepine receptor ligands. Eur J Med Chem 2002; 37(Pt 3): 219-29.
10. Hemmateenejad B, Miri R, Tabarzad M, Jafarpour M, Zand F. Molecular modeling and QSAR analysis of the anticonvulsant activity of some N-phenyl-N^0-(4-pyridinyl)-urea derivatives. J Mol Struct (Theochem) 2004; 684: 43-49.
11. Jin A Y, Kohn H, Béguin C, Andurkar SV, Stables JP, Weaver DF. A quantitative structure-activity relationship study for α-substituted acetamido-N-benzylacetamide derivatives -A novel anticonvulsant drug class. Can J Chem 2005; 83: 37-45.
12. Gavernet L, Barrios IA, Sella Cravero M, Bruno Blanch LE. Design, synthesis, and anticonvulsant activity of some sulfamides, Bioorg Med Chem 2007; 15: 5604-5614
13. Librowski T, Kubacka M, Meusel M, Scolari S, Müller CE, Gütschow M, Evaluation of anticonvulsant and analgesic effects of benzyl and benzhydryl ureides. Eur J Pharmacol 2007; 559: 138-149.
14. Kamiński K, Obniska J, Dybała M. Synthesis, physicochemical and anticonvulsant properties of new N-phenylamino derivatives of 2-azaspiro[4.4]nonane- and [4.5]decane-1,3-diones: part V. Eur J Med Chem 2008; 43(Pt 1): 53-61.
15. Jackson PL, Scott KR, Southerland WM, Fang YY, Enaminones 8: CoMFA and CoMSIA studies on some anticonvulsant enaminones. Bioorg Med Chem 2009; 17: 133-140.
16. Amnerkar ND, Bhusari KP. Synthesis, anticonvulsant activity and 3D-QSAR study of some prop-2-eneamido and 1-acetyl-pyrazolin derivatives of aminobenzothiazole. Eur J Med Chem 2010; 45: 149-159.
17. Kohn H, Conley JD, Leander JD. Marked stereospecificity in a new class of anticonvulsants. Brain Res 1998; 457(Pt 2): 371-375.
18. Broto P, Moreau G, Vandycke C. Molecular structures: Perception, autocorrelation descriptor and SAR studies. Eur J Med Chem 1984; 19: 66-70
19. Consonni V, Todeschini R, Pavan M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. J Chem Inf Comp Sci 2002; 42: 682-692.
20. Dragon, Talete. Italian chemometrics Inc.; c1993-2010 [cited 2003 Mar 8]. Available from: http://www.talete.mi.it
21. Ghavami R, Najafi A, Hemmateenejad B. Chemometrics-assisted spectrophotometric methods for simultaneous determination and complexation study of Fe(III), Al(III) and V(V) with morin in micellar media. Spectrochim Acta Part A 2008; 70: 824-834.
22. Ghavami R, Najafi A, Sajadi M, Djanati F. Genetic Algorithm as a Variable Selection Procedure for the Simulation of 13C Nuclear Magnetic Resonance Spectra of Flavonoid Derivatives Using Multiple Linear Regression. J Mol Graphics Modell 2008; 27: 105-115.