M. Lazar et al. / Journal of Pharmacy Research 2014,8(2),118-122

Research Article
ISSN: 0974-6943

Available online through
http://jprsolutions.info

Artemisinin analogues selection in linear and non-linear QSAR

M. Lazar[1]*, A. Mouzdahir,[2] M. Badia[3] and M. Zahouily[1].

[1]Department of Chemistry, Laboratory of Materials, catalysis and development of natural resources (URAC24) University of Hassan II–Mohammedia, Faculty of sciences and Technologies. B.P.146 (20650) Mohammedia, Morocco.

[2]Department of Chemistry, Laboratory of Bioorganic Chemistry University of Chouaïb Doukkali, Faculty of Sciences El Jadida.: Road Ben Maâchou B.P.: 20, (24000), El Jadida, Morocco.

[3]Ecole Royale de l'air, Marrakech, Morocco. Mechanic of Department,[3]Ecole Royale de l'air, Marrakech, Morocco. Mechanic of Department, ERA, BEFRA Marrakech 4000, Morocco.

Received on:07-11-2013; Revised on: 26-12-2013; Accepted on:11-12-2013

ABSTRACT

Artificial neural networks (ANNs) and multiple linear regressions (MLR) controlled non-linear and linear quantitative structure activity relationship (QSAR) models were established for a set of (32) Artemisinin derivative (ART). QSAR models derived from MLR and ANN yielded appreciable internal and external predictability. Using the pertinent descriptors calculate from this study that the activity dependent especially on substituent R and hydrogen liaison du substituent R_2 by different descriptors coding for same structural properties and energies and associated properties of the molecules are identified and discussed as overlapping structure features in linear and non-linear QSAR models. The independent variables selected in relation with concentration of 50% inhibition are BHA(R), W(R), and WM (R_2), processed by leave one-out cross validation and Y-scrambling technique. The test set (N=6) presented an external prediction power of 94%: In conclusion these overlapping features may reveal fundamental structural properties which convert a linear relationship to non-linear and better identifications of bio-chemical aspects of QSAR models to medicinal chemists.

Keywords: Artemisinin derivatives, Multiple Linear Regression, Structure-Activity Relationships, Artificial Neural Network.

INTRODUCTION

Artemisinin (ART), a natural product isolated from the plant *Artemesia annual*, is widely used as an antimalarial drug [1, 2, 3]. Recently, more and more evidences have emerged to elucidate that Artemisinin and its derivatives show potent anticancer activities in a variety of human cancer cells [4,5,6,7]. Artemisinin contains an endoperoxide bridge, which is cleaved in Fenton reaction mediated by iron and produces free-radical reactive oxygen species (ROS) [8,9]. However, oxidative damage alone is not sufficient to explain all of the anticancer activities of ART[10]. The advantage of this family of antimalarial and anticancer, that we have chosen to develop that it's able to reduce the number of parasites by a factor of about 10,000 to each asexual cycle, which is far more than other antimalarial drugs that reduce only by a factor of 100 to 1000 the number of parasites per cycle. It is that very few molecules developer to derivatives of Artemisinin on terrain.[11, 12,] The quantitative structure-activity relationships (QSARs) are certainly a major factor in contemporary drug design. Thus, it is quite clear why a large number of users of QSAR[13, 14] are located in industrial research units. Quantitative structure-activity relationship (QSAR) has been proven as the most successful tool in the comparative evaluation of the structure of a drug with its biological activity.[15]

A comparative study for estimation the pertinent descriptors between the postulate that statistical fitness and predictability of QSAR models are not related terms and should be treated and analyzed separately and a combination between MLR and ANN analysis for a set of 32 Artemisinin derivatives (Fig 1).

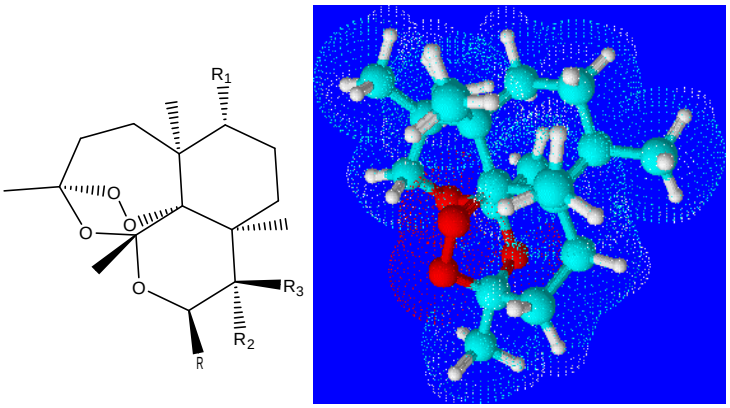

Figure 1: General structure of Artemisinin derivatives.

***Corresponding author.**
M. Lazar
Department of chemistry,
Laboratory of Materials,
catalysis and development of natural resources (URAC24)
University of Hassan II–Mohammedia, Faculty of sciences and Technologies. B.P.146 (20650) Mohammedia, Morocco.
E-mail: mo_lazar@yahoo.fr

MATERIALS AND METHODS

Biological Data

In this study, we are considering 32 substituted molecules. The chemical structure along with observed activity data of the compounds used in this work are shown in Table 1. The activity data were taken from various studies.[16]

Table 1: Chemical structure of anti-malaria activity of the date a set of Artemisinin analogues and observed activity malaria.

S.No	R	R_1	R_2	R_3	*log $(RA)_{obs}$
1	OH	CH_3	H	CH_3	0.854
2	O-$CH_2COOCH_2CH_3$	CH_3	H	CH_3	0.689
3	O-$(CH_2)_2COOCH_3$	CH_3	H	CH_3	0.202
4	O-$CH_2C_6H_4COOCH_3$	CH_3	H	CH_3	0.580
5	O-CH_2COO-	CH_3	H	CH_3	-1.264
6	O-$(CH)_2COO$-	CH_3	H	CH_3	-1.463
7	O-$(CH_2)_3COO$-	CH_3	H	CH_3	-1.411
8	O-$CH_2C_6H_4COO$	CH_3	H	CH_3	0.226
9	O-$(CH_2)_3COOH$	CH_3	H	CH_3	-0.786
10	O-CH_3	CH_3	H	CH_3	0.432
11	O-C_2H_5	CHO	H	CH_3	0.079
12	NH-phényl	CH_3	H	CH_3	0.146
13	2-N-phényl	H	Br	CH_3	-0.786
14	2,6 -N-phényl	H	Br	CH_3	-1.139
15	O-$CH(CH_2CH_2CH_3)(CH_2C_6H_5)$	H	Br	CH_3	-1.666
16	O-$CH(CH_2C_6H_5)(CH_2CH_2CH_3)$	H	H	CH_3	-0.122
17	O-$CH(C_6H_5)(COOCH_2CH_3)$	H	H	CH_3	0.375
18	O-$CH(COOCH_2CH_3)(C_6H_5)$	H	H	CH_3	0.904
19	O-$CH(CH_3)(p\text{-}CF_3\text{-}C_6H_4)$	H	H	CH_3	0.904
20	O-$CH(CH_2COOCH_2CH_3)(C_6H_5)$	H	H	CH_3	0.199
21	O-$CH(C_6H_5)(CH_2COOCH_2CH_3)$	H	H	CH_3	0.667
22	O-$CH(CH_2COOCH_2CH_3)(p\text{-}NO_2C_6H_4)$	H	H	CH_3	0.700
23	O-$CH(p\text{-}NO_2\text{-}C_6H_4)(CH_2COOCH_2CH_3)$	H	H	CH_3	0.612
24	O-$CH(p\text{-}CO_2CH_3C_6H_4)(CH_3)$	H	H	CH_3	0.971
25	O-$CH(p\text{-}CO_2CH_3C_6H_4)(CH_3)$	H	H	CH_3	0.522
26	O-$CH(CH_3)(p\text{-}CO_2CH_3C_6H_4)$	H	H	CH_3	-0.399
27	O-$CH(p\text{-}CO_2HC_6H_4)(CH_3)$	H	H	CH_3	-0.105
28	O-$CH(CH_2COOH)(p\text{-}NO_2C_6H_4)$	H	H	CH_3	-0.094
29	O-CH_2CF_3	H	OH	CH_3	0.255
30	O-CH_2CF_3	H	CH_3	OH	-0.824
31	O-CH_2CH_3	H	OH	CH_3	-0.347
32	**O-CH_2CH_3**	**H**	**CH_3**	**OH**	-1.097

$$*\log RA = \log\left[\left(IC_{50}(i)/\sum_{i=1}^{n=32} IC_{50}(\text{analogues})\right)\left(\sum_{i=1}^{n=32} M(\text{analogues})\right) / M(i)\right]$$

with: $IC_{50}(i)$: Concentration which inhibits 50% of component i.

$\sum_{i=1}^{n=32} IC_{50}(\text{analogues})$: The sum of the concentrations necessary for 50% inhibition

$\sum_{i=1}^{n=32} M(\text{analogues})$: The sum of Molecular mass of component i.

M (i) : Molecular mass of component i.

Molecular Descriptors Used

The objective of QSAR resides in parameterising the variation in chemical structure. It is evident that the performance of QSAR models depends mostly on the parameters used to describe the molecular structures. In this study, a set of descriptors related to physicochemical and geometric properties of the molecules was used in order to study their influence on the inhibition activity of these compounds. All these descriptors were calculated for the separate substituents R, R_1, R_2, R_3, and molecule structure (Fig. 1).

Molecular properties used for substituents R, R_1, R_2, R_3 and molecular structure, were: Size and shape described by means of surface (S) [17], and Van Der Waals volume (V) Molecular dimensions (length, width and height) [18]. Length (L): is the distance along the screen x-axis between the left and rightmost atoms plus their Van Der Waals radii. Width (W): is the distance along the screen y-axis between the top and bottommost atoms plus their Van Der Waals radii. Height (H): is the distance along the screen z-axis between the nearest and farthest atoms plus their Van Der Waals radii[19], Ratio V/L, V/W, W/H were also calculated[20], Electronegativity[21, 22], Molecular weight (MW), log P, the partition coefficient between n-octanol and water[23], Molar refractivity (MR)[24], hydrogen-bonding acceptors (HBA)[25], and Hydrogen-bonding donors (HBD) Some topological descriptors such as the winner indices w(x) , connectivity indices from Kier and Hall [26,27] up to fourth order. All these descriptors were calculated with Molecular descriptors for all dataset are calculated using an online server E-Dragon[15] (Pclient), an advanced version of well known tool Dragon and molecular modelling program (MMP) demo version[28] for molecule structure and the separate substituents.

Methodologie

The multiple linear regression method was used to generate linear models between the antimalarial activity and the molecular descriptors. Because of the large number of descriptors considered, a stepwise procedure combining the forward and backward algorithms was used to select the pertinent descriptors.

In order to avoid all difficulties in the interpretation of the resulting models, pairs variables with a correlation coefficient larger than 0.8 were classified as inter-correlated, and only one of these was classified in the screened model. The quality of the model was considered as statistically satisfactory on the basis of squared coefficient of determination (R^2), (r) coefficient of correlation, Standard error of estimate (SE), and F is the Ficher's statistics,

Artificial neural networks (ANN) are able to create internal models for complex input-output relationships based on learning from examples and therefore are useful in prediction. In medicine chemistry, ANN was successively used to predict the synthesis of a new organic compound structure.

The purpose of the current work is to provide an application of ANN to the structure antimalarial activity relationship of dihydroartemisinin compounds[16]. The result obtained by the ANN will be compared to those given by multiple linear regressions (MLR).

RESULTS AND DISCUSSION

Regression Linear Multiple (MLR) models and validation

Forward and Backward selections regression without constant can be achieved either by including all potential independent variables in the model and eliminating those that are not statistically significant, are presented in Eq. (1), (3) and (5).

Log (RA) = (- 0.913 ± 0.246) HBA(R) - (0.0007 ± 0.0001) W(R) – (0.207 ± 0. 08) J(R) + (0.596 ± 0.241) L/B_5(R) + (0.2482 ± 0.217) log P (R_2) – (0.019 ± 0.004) WM (R_2)---------- **(1)**

n = 32(R^2 = 0.67) (r = 0.82) SE = 0.49 = 8.71---------- **(2)**

Log (RA) = (- 0.827 ± 0.236) HBA(R) - (0.0006 ± 0.0001) W(R) – (0.190 ± 0.077) J(R) + (0.630 ± 0.242) L/B_5(R) – (0.0197 ± 0.0042) WM (R_2) ---------- **(3)**

n = 32 (R^2 = 0.65) (r = 0.81) SE = 0.50 F = 10.11---------- **(4)**

Log (RA) = (- 0.73 ± 0.2) HBA(R) - (0.0005 ± 0.0001) W(R) – (0.016 ± 0.004) WM (R_2)---------- **(5)**

n = 32 (R^2 = 0.57) (r = 0.75) SE = 0.53 F = 12.65---------- **(6)**

Where n is the number of compounds in the training dataset, R^2 is the coefficient of determination, r Correlation coefficient, SE Standard error of estimate, and F is the Ficher's statistics.

Table 2: Statistical parameters for MLR and ANN models

Variables independents	ANN			MLR		
	R^2	r	SE	R^2	r	SE
Eq. 1 with hexvariable: HBA(R); W(R); J(R); L/B_5(R); log$P(R_2)$ and WM (R_2).	0.95	0.97	0.19	0.67	0.82	0.49
Eq. 3: with Pentavariable HBA(R); W(R); J(R); L/B_5(R) and WM (R_2).	0.99	0.99	0.08	0.65	0.81	0.50
Eq. 5 with Trivariable: HBA(R); W(R) and WM (R_2).	0.94	0.97	0.18	0.57	0.75	0.53

All These equations shown a regression coefficient for hydrogen-bond acceptor and Winner indices of substituent R, followed by the molecular weight of substituent R_2. These equations are statistically significant with the highest F statistic or lowest p- value and it's notable that there no significant intercorrelation between descriptors appearing in the selected model, as seen in Table 3 from Trivariable**,** **B**ut, this observation isn't the case for equations 1 and 3 which present a correlation between Balaban Indices J(R) and report of Sterimol Parameters L/B_5(R). The input variable s may be correlated with another variable the response. If this is the case, the presence of the one variable input in the model may mask the effect of another variable input.

Table 3: Correlation matrix of descriptors trivariables.

	HBA (R)	W (R)	WM (R_2)
HBA (R)	1.0000		
W (R)	0.4995	1.0000	
WM (R_2)	0.0712	-0.1068	1.0000

Artificial neural network (ANN) models and validation

Artificial neural network was employed to deduce the possibility of non-linear effects on the data and to found a more accurate model, [29].

The ANN was produced by using the Gaussian transfer function and the pertinent descriptors appearing in the MLR model as input. A 3-9-1 neural network architecture was developed with the learning rate and optimum momentum of 0.01 and 0.9, respectively with 10.000 iterations. The nine hidden neurons were chosen to keep ? = 2.0 ± 0.2 [30]. To verify this condition, we also tried three to nine neurons in the hidden layer and it was found that nine hidden neurons gives the best result for the training and test sets

.

To evaluate the neural network, the results of its coefficient R^2 was compared with the R^2 of the regression model developed in this work. The R^2 values were 0.57 and 0.94 and were attributed respectively to models MLR and ANN. The corresponding standard error estimate (SE) for both models was respectively 0.53 and 0.18. This preliminary study enables us to conclude that the ANN with the (3-9-1) architecture was able to establish a satisfactory relationship between the pertinent descriptors and activity of analogues Artemisinin.

This method of using descriptors from the MLR has some advantages to focus the descriptors that govern the activities observed and predicted compounds in the modelling of this family of molecules, with the exception of a few compounds; ANN predictions are found more accurate and precise.

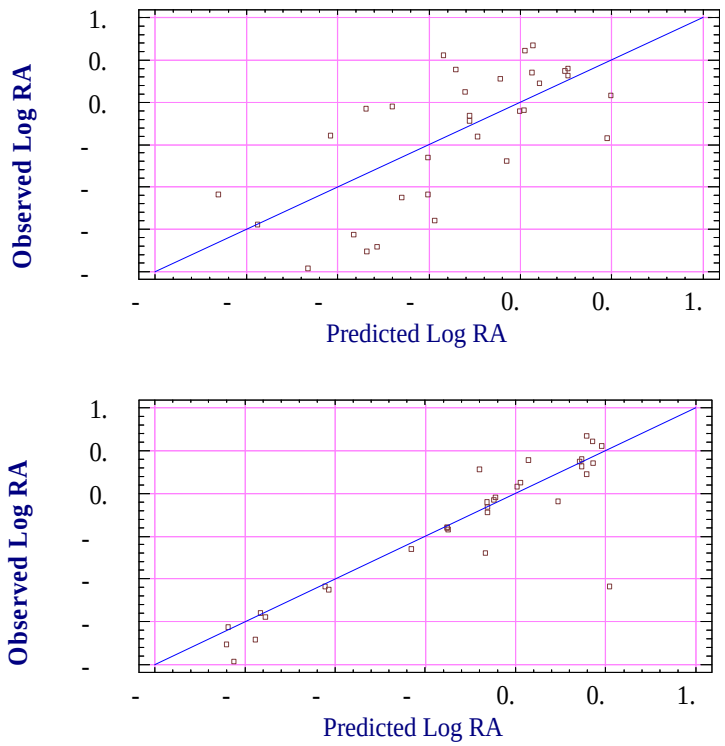

Figure 2: Graphical correlation of observed log RA and predicted log RA usingTridescriptors non-linear (ANN) QSAR model.

Table 4: Cross-validation parameters for MLR and ANN trivariables models

Model	R^2_{cv}	PRESS	SSY	PRESS/SSY
MLR	0.46	10.24	18.90	0.54
ANN	0.80	3.81	18.90	0.20

To insure that the results obtained were not due to chance and lend credence to our results, we have run a scrambling experiment and calculated PRESS and SSY parameters.

Firstly, the dependant variable log (RA) is randomly scrambled and then the same algorithms used in MLR and ANN run once again. The statistics of results is listed in Table 4 with R^2_{cv}, PRESS, SSY and PRESS/SSY. The graphical correlations of observed and predicted log RA for training and test sets are recorded in Fig.2. R^2_{cv} approved model stability and Y-scrambling dismissed any chance of by chance modelling. It's worthy to mention that ANN models (non-linear) found statistical superior than MLR models (linear). Observations conceived on predicted correlation of observed and estimated log RA values revealed a unique feature of non-linear models.

This test confirms and clearly shows that the descriptors selected in this study describe very well the activity studied. Note that in the aforementioned expression, SSY is the variance of the biological activity of the molecules around the mean value. PRESS is a good estimate of real prediction of error of the model, provided that the observations were independent. If PRESS is smaller than the sum of the squares of the response value (SSY), the model predicts better than chance and can be considered statistically significant. Table 4 shows that in all the two models PRESS are significantly smaller than SSY indicating them to be statistically significant.

The statistics of results by ANN from Eq. 1 and Eq. 3 with pentavariable and hexavariable are listed in Table 2 with R^2, r, and S.E. The descriptors chosen in Forward and Backward selections of MLR and ANN (Gaussian function) observed the big interference between all descriptors. It's worthy to mention that ANN models (non-linear) with interference between descriptors found statistical superior than MLR models (linear) and the postulate that statistical fitness and predictability of QSAR models are not related terms and should be treated and analyzed separately, is probably true just if we have not consider the interference between descriptors appearing in the selected model.

CONCLUSION

The search for new molecule structure, directs us to the use of descriptors easy to interpret and taking into account the complexity of modelled phenomena, we were able to show, with the fewer descriptors (three descriptors) and in 3D QSAR study, that the activity of the Artemisinin substituted derivatives, was strongly depending on the hydrogen-bonding acceptors (HBA) and winner indices of the substituent R and Molecular Weight of the substituent R_2 attached to the ART- substituted derivatives.

The model obtained using the ANN approach is more effective than the regression analysis alone because it reveals the non-linearity in the data, but the choice of descriptor is considered in this work for the most serious application and launch the synthesis a new structure molecule. In addition, the approach of ANN and MLR combination gives a good generalization of the model for predicting data.

ACKNOWLEDGEMENT

The authors are thankful to the University of Hassan II – Mohammedia, Faculty of Sciences and Technologies, Mohammedia, Morocco for providing necessary facilities to carry out this research work.

REFERENCES

1. Hsu E: The history of qing hao in the Chinese materia medica. Trans R, Soc Trop Med Hyg, 2006, 100: 505–508. doi: 10.1016/j.trstmh.2005.09.020.
2. Miller LH, Su XZ: Artemisinin Discovery from the Chinese Herbal Garden, Cell, 2011, 146: 855–858. doi: 10.1016/j.cell.2011.08.024.
3. Tu Y: The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine, Nat Med, 2011, 17: 1217–1220. doi: 10.1038/nm.2471.
4. Hou J, Wang D, Zhang R, Wang H: Experimental therapy of hepatoma with artemisinin and its derivatives: in vitro and in vivo activity, chemosensitization, and mechanisms of action, Clin Cancer Res, 2008, 14: 5519-5530. doi: 10.1158/1078-0432.ccr-08-0197.
5. Chen T, Li M, Zhang R, Wang H: Dihydroartemisinin induces apoptosis and sensitizes human ovarian cancer cells to carboplatin therapy, J Cell Mol Med, 2009, 13: 1358-1370. doi: 10.1111/j.1582-4934.2008.00360.x.
6. Morrissey C, Gallis B, Solazzi JW, Kim BJ, Gulati R, and al: Effect of Artemisinin derivatives on apoptosis and cell cycle in prostate cancer cells. Anti-Cancer Drugs, 2010, 21: 423-432. doi: 10.1097/cad.0b013e328336f57b.
7. O'Neill PM, Barton VE, Ward SA: The molecular mechanism of action of Artemisinin the debate continues. Molecules, 2010, 15: 1705–1721.
8. Firestone GL, Sundar SN: Anticancer activities of Artemisinin and its bioactive derivatives, Expert Rev Mol Med, 2009, 11: e32. doi: 10.1017/s1462399409001239.
9. Chaturvedi D, Goswami A, Saikia PP, Barua NC, Rao PG: Artemisinin and its derivatives: a novel class of anti-malarial and anti-cancer agents, Chem Soc Rev, 2010, 39: 435-454. doi: 10.1039/b816679j.
10. Lu JJ, Meng LH, Cai YJ, Chen Q, Tong LJ, and al: Dihydroartemisinin induces apoptosis in HL-60 leukemia cells dependent of iron and p38 mitogen-activated protein kinase activation but independent of reactive oxygen species, Cancer Biol Ther, 2008, 7: 1017-1023. doi: 10.4161/cbt.7.7.6035.
11. Jambou R., Legrand E., Niang M., Khim N, Lim P, Volney B,

Ekala M.T, Bouchier C, Esterre P, Fandeur T. Mercereau-Puijalon, O: Resistance of plasmodium falciparum Field isolates to in vitro artemether and point mutations of the SERCA–type PfATPase6. Lancet, 2005, 366, 1960-1963.

12. Menard D, Mastika-Claquin M.D, DolmazonV, Sarda J, Talarmin A: Association of failures of seven-day courses of artesunate in a population in Banquin Central African Republic with decreased sensitivity of plasmodium falciparum, Am. J. Trop. Med. Hyg, 2005, 73, 616-621.
13. Zahouily M, Lazar M, Bnoumarzouk M: A Quantitatice Structure-Activity Relationship Model, Journal of Chemical Product and process Modeling, 2008, Volume 5, Issue 1
14. Zahouily M, Lazar M, Elmaksoudi A, Rakik J, Elaychi S, Rayadh A: QSAR for antimalarial activity of aziridinyl and 2, 3 –bis (aziridinyl)-1, 4 naphtaoquinonyl sulfonate and acylate derivatives, J. mol. Model, 2005, Doi 10.1007/s008 94-005-0059x.
15. Zhang H, Zhang L, Peng L, et al: Quantitative structure-activity relationships of antimicrobial fatty acids and derivatives against Staphylococcus aureus, J. Zhejiang Univ Sci. B, 2012, 13(2):83-93.
16. Tonmunphean S, Kokpol S, Parasuk V, Wolschann P, Winger R. H, Liedl K. R, Rode B. M: *J.* Comput. Aided. Mol, 1988, 12, 397.
17. Balaban A. T: Applications of graph theory in chemistry, J. Chem. Inf. Comput. Sci, 1985, 25 334.
18. Hansen P. J, and Jurs P. C: Chemical application of graph theory, J. Chem. Educ, 1988, 65, 574.
19. Rouvary D. H: The modeling of chemical phenomena using topological Indices, J. Comput. Chem, 1987, 8: 4 470.
20. Labanowski J. K, Motoc I, and Dammkoehler R: A. Computers Chem, 1991, 1: 15 47.
21. Zahouily M., Rayadh A., Aadil M. and Zakarya D: Quantitative structure-diastereoselectivity relationships arylsulfoxides derivatives in radical chemistry, J. Mol. Model, 2003, 9, 242.
22. Nohair M, Zakarya D: Prediction of solubility of aliphatic alcohols using the restricted components of autocorrelation method, J. Mol. Mod, 2003, 9: 6 365.
23. Nohair M, Zakarya D, and Berrada A: Autocorrelation method adapted to generate new atomic environnements) J. Chem. Comp. Sci, 2002, 42, 3 586.
24. Coburn R. A, Weirzba M, Suto M. J, Solo A.J, Triggle A.M, Triggle D.J: 1, 4-Dihydropyridine antagonist activities at the calcium channel: a quantitative structure-activity relationship approach J. Med. Chem, 1988, 31 2103.
25. Devillers J, and Balaban A. T: Topological indices and related Descriptors in QSAR and QSPR, Cordon and Breach Science Publishers, 1999.
26. Takahata Y., Costa M. C. A. and Gaudio A.C: Pattern recognition and neural networks applied to structure-activity, J. Chem. Inf. Comput. Sci, 2003, 43, 540.
27. Kier L.B, and. Hall L.H: The Molecular Connectivity Chi Indexes and kappa Shape Indexes in Structure-Property Modelling. In Reviews in Computational Chemistry (Lipkowitz, K. B., and Boyd, D. B., Eds. Vol. 2, VCH Publishers, New York, 1995, p. 367.
28. MMP, molecular modelling pro-Demo (TM) Revision 301 demo published by ChemSW Software (TM).
29. Meenakshi Ojha, Mukesh Yadav, Anuraj Nayarisseri, Jyoti Prajapati, Neetesh Pandey, Payal Chatterjee: Overlapping structure features selection in linear and non-linear QSAR, Journal of pharmacy research,6, 2013, 183-187.
30. Cherqaoui D, Esseffar M, Villemin D, Cence JM, Chastrette M, Zakarya D: Structure-musk odor relationships studies of tetralin and indan compound using neural networks, New J. Chem New J Chem, 1998, 839–843 31. Tetko, I.V, Villa, A.E.P, Livingstone, D.J: J. Chem. Inf. Comput, 1996, 393, 36, 794.

Source of support: Nil, Conflict of interest: None Declared