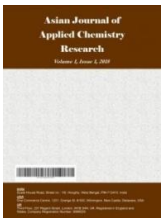


Asian Journal of Applied Chemistry Research

1(2): 1-11, 2018; Article no.AJACR.41186

Application of Substituent Electronic Descriptors QSAR Model of 2-amino-6-arylsulfonylbenzonitriles as HIV-1 Reverse Transcriptase Inhibitors based on the MOLMAP Approach

Razieh Sabet[1*] and Maryam Sabet[2]

[1]Department of Medicinal Chemistry, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran.
[2]Bioinformatic and Computational Research Center, Shiraz University of Medical Sciences, Shiraz, Iran.

Authors' contributions

This work was carried out in collaboration between both authors. Author RS designed the study, performed the statistical analysis, wrote the protocol, wrote the first draft of the manuscript and managed the literature searches. Author MS managed the analyses of the study. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJACR/2018/v1i29617
Editor(s):
(1) Yousry Moustafa Issa, Professor of Analytical Chemistry, Department of Chemistry, Faculty of Science, Cairo University, Egypt.
Reviewers:
(1) Fatma Kandemirli, Kastamonu University, Turkey.
(2) Nobuaki Tanaka, Shinshu University, Japan.
(3) A. Veerareddy, India.
Complete Peer review History: http://www.sciencedomain.org/review-history/24934

Original Research Article

Received 20th February 2018
Accepted 29th April 2018
Published 1st June 2018

ABSTRACT

The HIV-1 reverse transcriptase (RT) is a major target for drug development. Inhibition of this enzyme has been one of the primary therapeutic strategies in suppressing the replication of HIV-1. A series of 2-amino-6-arylsulfonylbenzonitrile derivatives were subjected to quantitative structure-activity relationship (QSAR) analysis. Very recently, we proposed the use of substituent electronic descriptors (SED) instead of the electronic descriptors of whole molecules as new and expedite source of electronic descriptors. In this study, we used SED parameters in QSAR modeling of anti HIV-1 activity of 6-arylsulfonylbenzonitrile derivatives. In SED methodology produces a vector of electronic descriptors for each substituent and thus a matrix of SED is generated for each

**Corresponding author: E-mail: sabet_r@sums.ac.ir;*

molecule. Consequently, a three-dimensional array is obtained by staking the data matrices of different molecules beside each other. As a novel multiway data analysis method, molecular maps of atom-level properties (MOLMAP) approach was also used to transfer a three-dimensional array of SED descriptors into new two-dimensional parameters using Kohonen network, following by genetic algorithm-based partial least square(GA-PLS) to connect a quantitative relationship between the Kohonen scores and biological activity.In unfolding data, HOMO1, $HOMO_B1$, SOF_B1 and EPH_A4 represent the most important indices on QSAR equation derived by PLS analysis. Accurate QSAR models were obtained by both approaches. The resulted GA-PLS model of MOLMAP approach possessed high statistical quality r^2= 0.83 and q^2=0.70. It could explain and predict about 70% of variances in the anti-HIV1 inhibitory activity of the studied molecules. However, the superiority of three-way analysis of SED parameters based on MOLMAP approach with respect to simple unfolding was obtained.

Keywords: Substituent electronic descriptors; MOLMAP; QSAR; Kohonen neural network.

1. INTRODUCTION

An appropriate representation of the structural and physicochemical features of chemical agents is an essential key to the successful application of QSAR models [1-4]. Structural descriptors have been classified into different categories according to different approaches including physiochemical, constitutional, geometrical, topological, and quantum chemical descriptors. Currently, more than 1000 molecular descriptors can be easily calculated using available software such as Dragon [5-6]. In recent years, electronic descriptors obtained from quantum chemical calculations have been found major popularity, and progress in computational hardware and development of efficient algorithms have assisted the routine development of molecular quantum chemical calculations [7-9].

Lately, there is a challenge between calculation complexity and accuracy in the selection of quantum chemical calculation methods (e.g., semi-empirical and *ab initio* methods). In fact, in spite of highly accurate results obtained by the *ab initio* methods, the complexity and long computation times of these methods hinder their wide applicability to the calculation of electronic descriptors. Meanwhile, there is a great demand for the development of electronic descriptors that are calculated with high level of accuracy and low computation times, and can also be tabulated for future works. In this line, we have recently proposed the use of electronic features of substituents instead of those of the whole molecular skeleton [10-12]. These parameters, called as substituent electron descriptors (SED), can be used as an alternative to both substituent constants and molecular descriptors. All of the electronic descriptors that are calculated for the whole molecule can be calculated for substituents, considering the fact that compared with the parent molecules, they have very smaller sizes and consequently their calculation times will be significantly decreased [10].

In SED methodology, instead of a vector of descriptors employed in conventional QSAR studies, a matrix of descriptors is generated. In fact, if the molecules of data set have a similar structural backbone and have *n* common substituent positions, for each of these positions a set of *m* SED properties are calculated. Therefore, a SED data matrix of ($n \times m$) dimension is generated for each molecule, and a three-dimensional array of ($k \times n \times m$) dimension is obtained for *k* molecules.

Recently, a multiway analytical method based on Kohonen network, originated from a method for calculation of molecular descriptors called MOLMAP, has been introduced [13-15]. Kohonen self-organizing maps (SOM) can be used for the reduction of multidimensional objects to 2D [16]. In QSAR modeling based on MOLMAP approach, the resulted Kohonen scores are used as descriptors for classification and regression purposes. In this paper; we illustrate the application of SED and MOLMAP descriptors for QSAR analysis of anti HIV-1 activity of 6-arylsulfonylbenzonitrile derivatives. Models were developed for prediction of anti HIV-1 activity with GA-PLS method.

2. METHODOLOGY

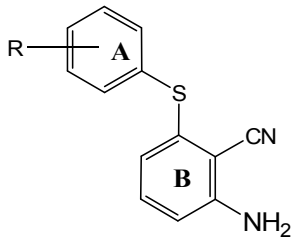
The experiments here described required two major steps, the generation of descriptors and development of predictive models. Generation of descriptors, MOLMAP, was obtained by a Kohonen self-organizing map. A SOM must be trained beforehand with a different of substituents from different structures (each substituents described by the 25 SED properties

calculated by Gaussian 98). Then all the substituents of one molecule are submitted to the trained SOM, and the pattern of activated neurons is a map of the reactivity features of that molecule (MOLMAP)—a fingerprint of the substituents available in that structure. Such MOLMAP (molecular maps of atom-level properties) descriptors can be directly used in QSAR [15].

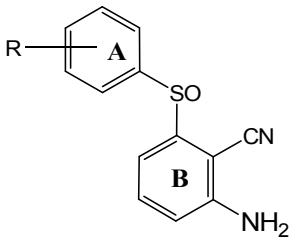
2.1 Data Set

The biological data used in this study were HIV-reverse transcriptase inhibitory activity, (in terms of -log IC_{50}), of a set of sixty one 6-arylsulfonylbenzonitrile derivatives [17]. The structural features and biological activity of these compounds are listed in Table 1.

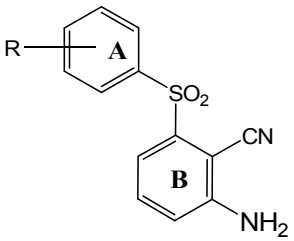
Table 1. Chemical structures of 2-amino-6-arylsulfonylbenzonitriles derivatives used in this study



1-19



20-32



33-61

Compound	R	Experimental pIC_{50} [a]
1	H	1.836
2	2-OCH_3	2.367
3	3-OCH_3	2.222
4	2-CH_3	1.796
5	3-CH_3	2.215
6	4-CH_3	0.939
7	2-Cl	2.387
8	3-Cl	2.131
9b	2-Br	1.523
10 b	3-Br	2.292
11	3-F	2.009
12	3-CN	2.762
13	4-CN	1.359
14	3-CF_3	1.893
15 b	3-NH_2	1.502
16	3,5-$(CH_3)_2$	3.367
17 b	3-Cl,5-CH_3	2.754
18	3-OCH_3, 5-CH_3	2.699
19	3-OCH_3, 5-CF_3	2.292
20 b	2-OCH_3	2.319
21 b	3-OCH_3	1.796
22 b	3-CH_3	1.534
23 b	4-CH_3	1.310
24	2-Br	1.407
25	4-Br	1.694
26	2-CN	2.409
27	3-CN	1.848
28	3-CF_3	1.398
29	3,5-$(CH_3)_2$	3.469
30	2,5-Cl_2	2.007
31 b	3-Cl, 5-CH_3	3.495

32b	3-OCH_3, 5-CF_3	2.684
33	H	2.699
34 b	2-OCH_3	3.222
35	3-OCH_3	3.046
36	4-OCH_3	1.602
37	2-CH_3	2.638
38 b	3-CH_3	3.398
39	4-CH_3	2.022
40	2-Cl	2.387
41	3-Cl	3.229
42	4-Cl	2.523
43	2-Br	2.301
44	3-Br	3.268
45	4-Br	1.699
46	2-F	2.523
47	3-F	2.523
48	2-CN	2.268
49	3-CN	2.620
50	4-CN	1.097
51	3-CF_3	2.456
52	2,5-Cl_2	3.523
53	3,5-Cl_2	4.155
54	3,5-$(CH_3)_2$	5.000
55	3-Br, 5-CH_3	4.699
56	3-Cl, 5-CH_3	4.523
57 b	3-OCH_3, 5–CH_3	4.301
58	3-OCH_3, 5-CF_3	4.046
59	3-OH, 5-CH_3	3.367
60	3-OCH_2CH_3, 5-CH_3	4.222
61	3-$O(CH_2)_2CH_3$, 5-CH_3	4.222

a pIC_{50} = -log (IC_{50}),
b Compounds used as prediction set

2.2 Computational Procedure

A Core i7 personal computer with windows X operating system was used. The SED parameters were calculated according to our previously published article in this subject [10,11]. Here, quantum chemical calculations were performed on radical substituents instead of whole molecular structures. The calculated electronic descriptors for each substituent are summarized in Table 2. The calculated descriptors can be classified into three different electronic categories including local charges, dipoles, and orbital energies. The quantum chemical indices of hardness (HD), softness (SOF), electronegativity (EN), and electrophylicity (EPH) were calculated according to the method proposed by Thanikaivelan et al. [18]. Since most of the substituents are open shell quantum species (due to being in doublet quantum state as a radical molecule), a difference in energy between two electronic energy populations, alpha (spin up) and beta (spin down) can be seen using Gaussian 98. It provides some additional descriptors $HOMO_A$, $HOMO_B$, $LUMO_A$, $LUMO_B$, HD_A, HD_B, SOF_A, SOF_B, EN_A, EN_B, EPH_A, and EPH_B stem from two different alpha and beta electronic population energies, where the subscripts A and B stand for alpha and beta population of electronic energy, respectively. Therefore, a total of 25 electronic descriptors were calculated for each substituent (Table 2).

2.3 Variable Importance in the Projection (VIP)

In order to investigate the relative importance of the variable appeared in the final model obtained by GA-PLS method, variable important in projection (VIP) was employed [19]. VIP values reflect the importance of terms in PLS model. VIP is sum over all model dimensions of the contributions VIN (variable influence) [20].

Table 2. List of the calculated SED parameters

No	Notation	Definition
1	QRMS	Root mean square error of charges
2	SPQ	Sum of positive charges
3	SNQ	Sum of negative charges
4	DRMS	Root mean square of dipole moments at any Cartesian coordinate direction
5	DTOT	Total dipole moment
6	FRMS	Root mean square force that any atom
7	FMAX	Maximum force on molecule
8	HOMO (A&B)	Highest occupied molecular orbital (A and B for alpha & beta population of electronic energy)
9	LUMO (A&B)	Lowest Unoccupied molecular orbital (A and B for alpha & beta population of electronic energy)
10	SOF (A&B)	Softness orbital (A and B for alpha & beta population of electronic energy)
11	HD (A&B)	Hardness orbital (A and B for alpha & beta population of electronic energy)
12	EPH (A&B)	Electrophilicity orbital (A and B for alpha & beta population of electronic energy)
13	EN (A&B)	Electronegativity orbital (A and B for alpha & beta population of electronic energy)

2.4 Training of a Kohonen Self-organizing Map with SED

The Kohonen Self-Organizing Map defines an ordered mapping, a kind of projection from a set of given data items onto a regular, usually two-dimensional grid. Kohonen Neural Network (KNN) is an unsupervised learning method, that revealing similarities between object. A Kohonen SOM consist of a grid of so-called neurons, each containing as many elements (weight) as there are input variables. Fig. 1 shows the architecture of a Kohonen network: each column in this two-dimensional arrangement represents a neuron; each box in such a column represents a weight of a neuron. Each neuron has as many (m) weights, w_{ji}, as there are input data, x_i, for the object that is being mapped into the network.

Here, the input variables are the 25 SED descriptors. During the training, each individual substituent is mapped into the neuron, which has weights most similar to the descriptors of the input object. This is the central neuron, or winning neuron. Not only the winning neuron has its weights adjusted, but also the neurons in its neighborhood. The extent of adjustment depends, however, on the topological distance to the winning neuron—the closer a neuron is to the central neuron the larger is the adjustment of its weights. The objects of the training set are iteratively fed to the map, the weights corrected, and the training is stopped when a pre-defined number of cycles are attained. Then, the pattern of activated neurons can be considered as a fingerprint of the objects and constitute their MOLMAP scores. For numerical processing, each neuron got a value equal to the number of times it was activated by substituents of the molecule [21]. Finally; the map is transformed into a vector by concentration of columns resulting in a fixed length MOLMAP score where each scores of each object have a dimension of (v×v). The best value of v was obtained by trial and error, and the best results were obtained for v =11. It should be noted that output layer dimensions (4×4) to (13×13) was also checked but the best results was achieved using 11×11=121. The obtained MOLMAP descriptors were then subjected to PLS modeling with a GA for variable selection.

2.5 Modeling Procedures

Two different Kohonen maps were implemented to obtain QSAR models using SED parameters. The first one was the traditional unfolding methods, which we used in our previous paper on SED parameters. For a set of molecules with n common substitution position on the molecular basic skeleton, a row vector consist of (25×n) columns (each 25 column is related to SEDs of one substituent position) was provided for each molecules. The descriptor data matrix was then constructed by staking the SED row vectors of different molecules under each others.

In the second approach, a three-dimensional array of SED parameters (molecules on the first mode, substituent positions on the second mode

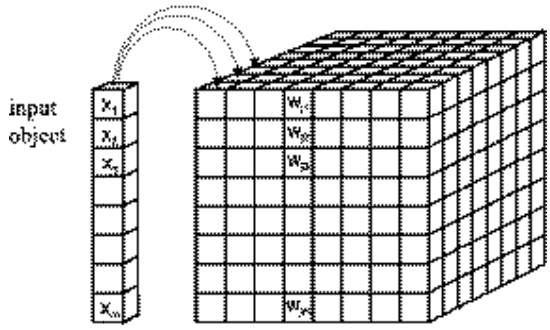


Fig. 1. Architure of a Kohonen neural network; The input object $X=(x_1, x_2,.., x_m)$ is mapped into an n x n arrangement of neurons, j, each having a weight vector $W_j = (w_{j1}, w_{j2}, ..., w_{jm})$

and SED properties of the substituents on the last mode), was provided, which was subjected to MOLMAP analysis to get a two dimensional array of kohonen scores. The MOLMAP scores were produced on the basis of Kohonen map training.

In both approaches (i.e. unfolding and MOLMAP analysis), the Partial least squares (PLS) regression was employed to evaluated the structure-activity relationships and genetic algorithm (GA) was used variable selection. In the case of simple unfolding, the input of GA-PLS was a set of (25×*n*) SED properties, whereas in MOLMAP approach the set of 121 Kohonen scores (i.e., 11×11 Nodes) were used as input. In order to investigate the prediction ability of the models, the data set (n = 61) was divided into two group: calibration set (n = 48) and prediction set (n = 13). Given 48 calibration samples; leave-one out cross-validation procedure was used to find the optimum number of latent variables for each PLS model. GA produces a population of acceptable models in each run. In this work, many different GA-PLS runs were conducted using different initial set of populations (50-250) and therefore a large number of acceptable models were created.

The linear PLS model finds 'new variables' (latent variables or **X** scores) which are linear combination of the original variables. To avoid over-fitting, a strict test for the significance of each consecutive PLS component is necessary and then stopping when the components are non-significant. Cross validation is a practical and reliable method of testing this significance [22]. Application of PLS thus allows the construction of larger QSAR equations while still avoiding over-fitting and eliminating most variables [23,24].

For Kohonen mapping, the MOLMAP toolbox, developed by Milano Chemometrics and QSAR research Group, was used [13]. The PLS regression method used was the NIPALS-based algorithm existed in the chemometrics toolbox of MATLAB software (version 7.1 Math work Inc.). Leave-one-out cross-validation procedure was used to obtain the optimum number of factors based on the Haaland and Thomas F-ratio criterion [25].

3. RESULTS AND DISCUSSION

3.1 Simple Unfolding of SED Parameters

Then attempts were made to obtain a unified QSAR model for a whole set of molecules.

A SED data matrix of (5×25) dimension is generated for each molecule, and a three-dimensional array of (5×25×61) dimension is obtained for 61 molecules. As it is shown in Table 1, all the chemical compounds used in this study share a similar structural backbone but they are different in the oxidation state of sulfur atom. It has oxidation numbers of +2, +4 and +6 for compounds 1-19, 20-32 and 33-61, respectively. We added oxidation number as 126^{th} descriptor. Thus in simple unfolding of SED parameters a data matrix has (126×61) dimension for 61 molecules. In PLS analysis, the descriptors data matrix is decomposed to orthogonal matrices with an inner relationship between the dependent and independent variables. Since redundant variables degrade the performance of PLS analysis, similar to other regression methods, a variable selection method must be employed to find the more convenient set of descriptors. Here, GA was used as variable selection method.

The most useful GA-PLS model that resulted in the best fitness contained 13 descriptors including, the PLS estimate of the regression coefficients of simple unfolding of SED parameters (13 descriptors) are shown in Fig. 2.

Since these constants were calculated based on the normalized descriptor values, they can be used as a measure of the importance of the corresponding descriptor. As it is observed, the SED (SNQ2, SOF2, QRMS3, $HOMO_A3$, EN3, EN_B3, SOF_B3, $LUMO_B4$, EN_B4, $HOMO_A5$, EN5, EN_A5 and Oxidation number) parameters represent the most significant contribution in the obtained QSAR model and all of descriptors are molecular orbital energies.

The PLS estimate of the regression coefficients are shown in Fig. 2. As it is seen, Oxidation number, EN5, EN_A5 and $HOMO_A5$ represent the most significant contribution in the resulted QSAR model followed by the $LUMO_B4$, EN_B4 and QRMS3. The statistical parameters of the resulted PLS-based QSAR model are given in Table 3. The resulted GA-PLS model possessed high statistical quality r^2= 0.81 and q^2=0.66. It could explain and predict about 66% of variances in the anti-HIV1 inhibitory activity of the studied molecules. The predictive ability of the model was measured by application to 13 external test set molecules. The correlation coefficient of prediction set is 0.75, which means that the resulted QSAR model could predict 75% of variances in the inhibitory activity data. To measure the significance of the 13 selected PLS descriptors in the HIV-1 reverse transcriptase activity; variable importance in projection (VIP) was calculated for each descriptor. According to Erikson *et al*., X-variables (predictor variables) could be classified according to their relevance in explaining y (predicted variable) so that VIP > 1.0 and VIP < 0.8 mean highly or less influential, respectively, and 0.8 < VIP < 1.0 means moderately influential [19-20]. The calculated VIP values of the selected SED parameters (shown in Fig. 2) represent the relative significance of the variables in the biological activity. As seen VIP shows that HOMO1, $HOMO_B1$, SOF_B1 and EPH_A4 represent the most important indices on QSAR equation derived by PLS analysis. In addition, EN_A1 and $HOMO_A5$ parameter that has been found as a highly influential parameter.

3.2 Analysis of SED Parameters by MOLMAP Approach

As noted before, in the case of MOLMAP analysis, the SED parameters should be arranged in a three-way array in the direction of molecules, substitution positions and SED parameters. Then, finally, introduce as input of Kohonen network to obtain the related scores. Here, we investigated different dimensions in the range of (4×4) to (13×13) and, in each case, the resulting scores were used as input of GA-PLS model. Best results were obtained by a (11×11) dimension for all data sets. The statistical parameters of the resulted PLS-based QSAR model are given in Table 3. The PLS estimate of the regression coefficients is shown in Fig. 3. The distribution of substituents in the resulted Kohonen map (11×11) is represented in Fig. 4. In the map the numbers denotes the substituent numbering shown in Table 1. It should be noted that the map contain 121 nodes, which can be numbered sequentially from 1 to 121 so that the nodes of the first row are numbered as N_1-N_{11}, those of second row as N_{12}-N_{21} and so on. The map shows a relatively random distribution of substituents. The scores of this map (121 variables) were used as input of PLS regression. The relative importance of selected neurons for GA-PLS model is shown in Fig. 4. The resulted GA-PLS model possessed high statistical quality r^2= 0.83 and q^2=0.70. It could explain and predict about 70% of variances in the anti-HIV1 inhibitory activity of the studied molecules. The predictive ability of the model was measured by application to 13 external test set molecules. The correlation coefficient of prediction set is 0.80, which means that the resulted QSAR model could predict 80% of variances in the inhibitory activity data and standard error of prediction was 0.57. It is clearly observed that the calibration correlation coefficients, correlation coefficients of cross validation and prediction obtained from Kohonen mapping are already higher than those obtained from simple unfolding of SED parameters.

Table 3. GA-PLS based QSAR models obtained by simple unfolding of SED parameters and by MOLMAP approach

Model	r^2_c[a]	q^2[b]	$RMSE_{CV}$[c]	r^2_P[d]
Simple unfolding	0.81	0.66	0.57	0.75
MOLMAP approach	0.83	0.70	0.57	0.80

a r^2_c = Regression coefficient for calibration set, b q^2 = Cross- validation correlation coefficient
c $RMSE_P$ = Root mean square error for calibration set,
d r^2_p = Regression coefficient for prediction set

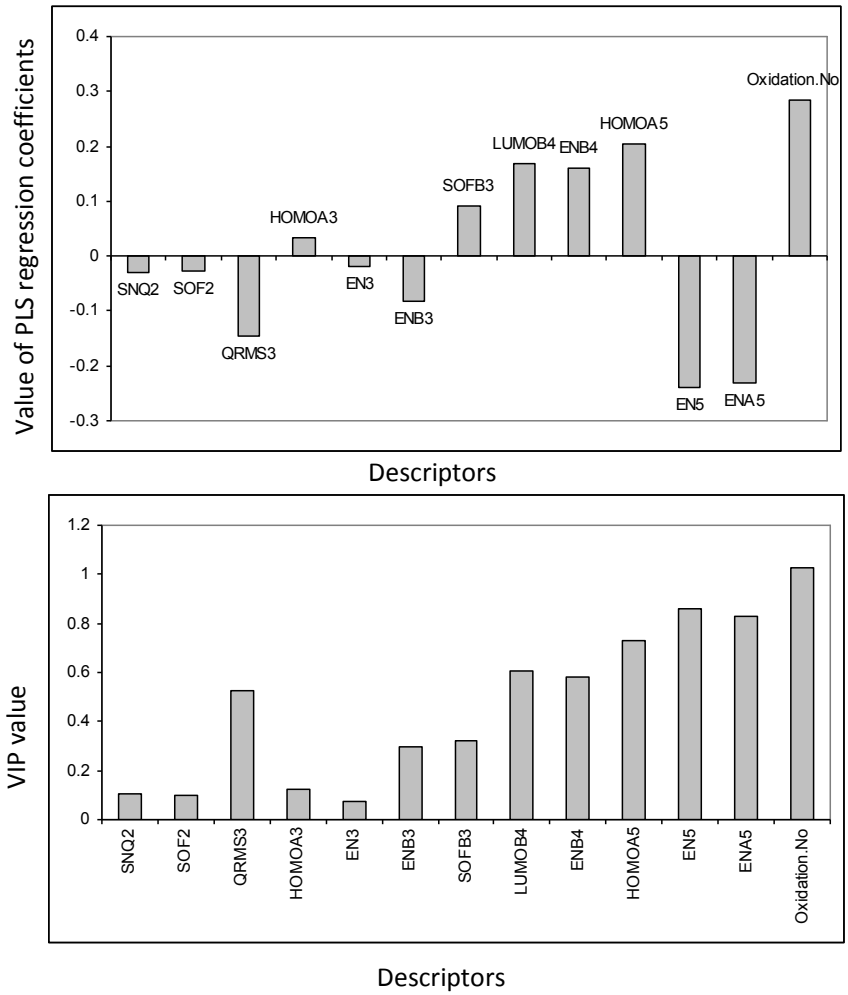
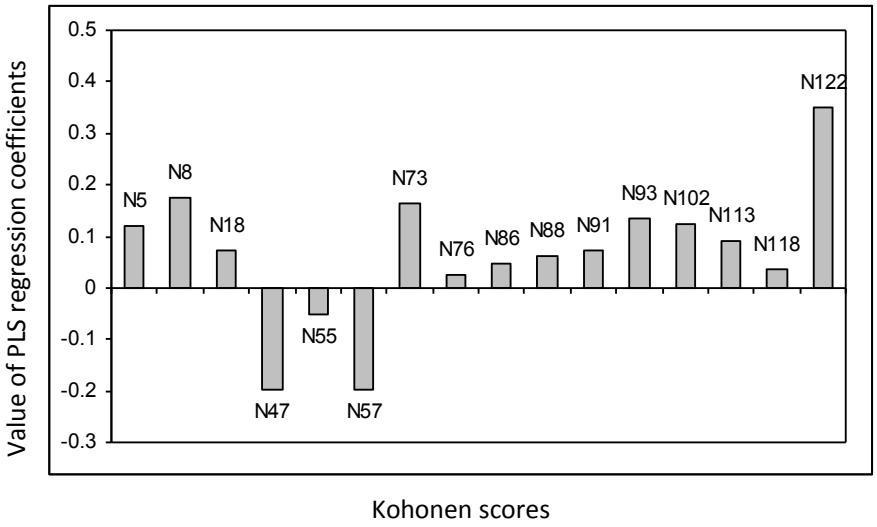


Fig. 2. PLS regression coefficients and the VIP values for the selected simple unfolding of SED parameters used in GA-PLS model



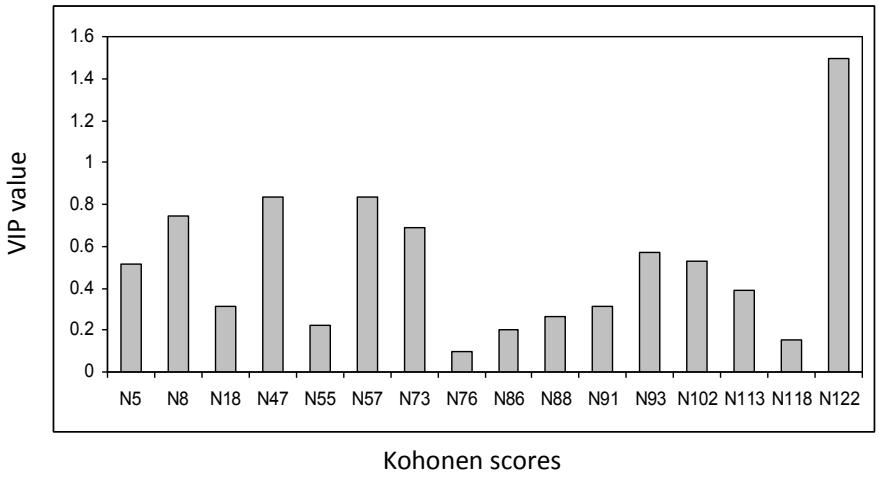


Fig. 3. PLS regression coefficients and the VIP values for the selected Kohonen scores of the above map by GA-PLS used in GA-PLS model

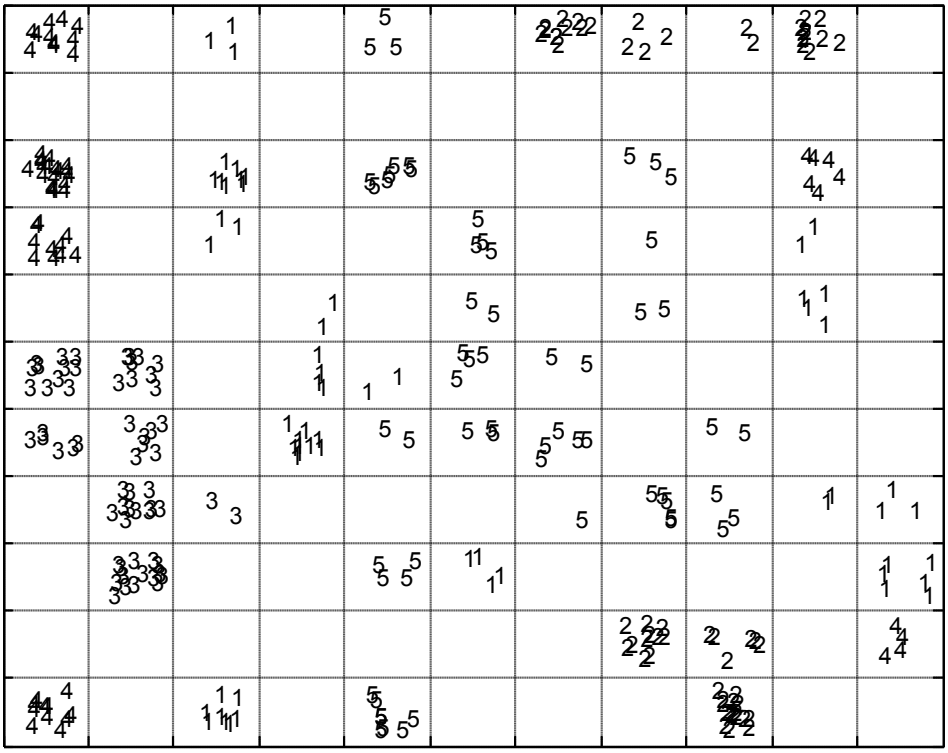


Fig. 4. Kohonen map where input is arranged in the direction of substituents

4. CONCLUSION

The usefulness and applicability of the substituent electronic descriptors as an expedite source of quantum chemical parameters were investigated by QSAR modeling of 2-amino-6-arylsulfonylbenzonitrile derivatives.

In unfolding data, the most convenient GA-PLS model that resulted in the best fitness contained 13 descriptors. The SED (SNQ2, SOF2, QRMS3, $HOMO_A3$, EN3, EN_B3, SOF_B3, $LUMO_B4$, EN_B4, $HOMO_A5$, EN5, EN_A5 and Oxidation number) parameters represent the most significant contribution in the obtained QSAR model. VIP

shows that HOMO1, $HOMO_B1$, SOF_B1 and EPH_A4 represent the most important indices on QSAR equation derived by PLS analysis.

The resulted GA-PLS model of MOLMAP approach possessed high statistical quality r2= 0.83 and q2=0.70. It could explain and predict about 70% of variances in the anti-HIV1 inhibitory activity of the studied molecules. Thus MOLMAP analysis of SED parameters shows more accurate results than those obtain from simple unfolding SED parameters.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Hansch C, Fujita T. ρ-σ-π analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc. 1964;86:1616-1626.
2. Wang J, Zhang L, Yang G, Zhan CG. Quantitative structure-activity relationship for cyclic imide derivatives of protoporphyrinogen oxidase inhibitors: A study of quantum chemical descriptors from density functional theory. J Chem Inf Comput Sci. 2004;44:2099-2105.
3. Hammet LP. The effect of structure upon the reactions of organic compounds. Benzene derivatives. J Am Chem Soc. 1937;59:96-103.
4. Hansch C, Hoekman D, Gao H. Comparative QSAR: Toward a deeper understanding of chemicobiological interactions. Chem Rev. 1996;96:1045-1075.
5. Todeschini R, Consonni V. Handbook of Molecular Descriptors. Wiley-VCH: Weinheim; 2000.
6. Todeschini R. Milano Chemometrics and QSPR Group.
   Available:http://michem.disat.unimib.it/
   (Accessed 14 March, 2009)
7. Sabet R, Fassihi A, Moeinifard B. QSAR study of PETT derivatives as potent HIV-reverse transcriptase inhibitors. J. Mol. Graph & Model. 2009;28:146-155.
8. Hemmateenejad B, Sabet R, Fassihi A. QSAR studies on 2-amino-6-arylsulfonylbenzonitriles as HIV-1 reverse transcriptase inhibitors using descriptors obtained from substituents and whole molecular structures. Chem. Biol & Drug. Des. 2009;74:405-415.
9. Sabet R, Fassihi A. QSAR study of Isatin analogues as *in vitro* anti-cancer agents. Eur. J. Med. Chem. 2010;45:1113-1118.
10. Sabet R, Fassihi A, Hemmateenejad B, Saghaie L, Miri R, Gholami M. Computer-aided drug design of novel antibacterial 3-hydroxypyridine-4-ones: Application of QSAR methods based on the MOLMAP approch. Journal of Computer-Aided Molecular Design. 2012;26:349-361.
11. Hemmateenejad B, Mehdipour AR, Miri R, Shamsipur. Application of MOLMAP Approach for QSAR Modeling of Various Biological Activities Using Substituent Electronic Descriptors M. J Comput Chem. 2009;30:2001-2009.
12. Hemmateenejad B, Mehdipour AR, Popelier, Paul L. A. Chem. Quantum Topological QSAR Models based on the MOLMAP Approach, Biol. Drug. Des. 2008;72:551-563.
13. Ballabio D, Consonni V, Todeschini R. Anal Chim Acta. 2007;605:134-146.
14. Zhang QY, Aires-de-Sousa J. J Chem Inf Model. 2007;47:1-8.
15. Gupta S, Matthew S, Abreu PM. Aires-de-Sousa, J. Bioorg Med Chem. 2006;14: 1199-1206.
16. Zupan J, Gasteiger J. Neural Network in Chemistry and Drug Design, 2nd ed.; Wiley-VCH: Weinheim; 1999.
17. Chan JH, Hong JS, Hunter RN, Orr GF, Cowan JR, Sherman DB, et al. 2-Amino-6-arylsulfonylbenzonitriles as non-nucleoside reverse transcriptase inhibitors of HIV-1. J Med Chem. 2001;44:1866-1882.
18. Thanikaivelan P, Subramanian V, Rao JR, Nair BU. Application of quantum chemical descriptor in quantitative structure activity and structure property relationship. Chem Phys Lett. 2000;323:59-70.
19. Olah M, Bologa C, Oprea TI. An automated PLS search for biologically relevant QSAR descriptors. J Comput. Aided Mol Des. 2004;18:437-449.
20. Erikson L, Johansson E, Kettaneh-Wold N, Wold Multi S, Megavariate Umetrics Academy, Umea; 2001.
21. Zhang QY, Aires-de-Sousa J. J Chem Inf Model. 2005;45:1775-1783.
22. Fan Y, Shi LM, Kohn KW, Pommier Y, Weinstein JN. Quantitative structure-antitumor activity relationships of camptothecin analogues: Cluster analysis

and genetic algorithm-based studies. J Med Chem. 2001;44:3254-3263.
23. Bhattacharya P, Roy K. QSAR of adenosine A3 receptor antagonist 1,2,4-triazolo[4,3-a]quinoxalin-1-one derivatives using chemometric tools. Bioorg Med Chem Lett. 2005;15:3737-3743.
24. Leardi R. Genetic algorithms in chemometrics and chemistry: A review. J Chemometrics. 2001;15:559-569.
25. Haaland DM, Thomas EV. Anal. Chem. 1988;60:1193-1204.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/24934