

Original Article

QSAR Studying of Oxidation Behavior of Benzoxazines as an Important Pharmaceutical Property

Elham Baher^{a*}, Naser Darzi^b

^aFaculty of science, Department of chemistry, Golestan University, Gorgan, Iran. ^bFaculty of science, Department of chemistry, Azad University of Mashad, Mashad, Iran.

Abstract

In this work the electrooxidation half-wave potentials of some Benzoxazines were predicted from their structural molecular descriptors by using quantitative structure-property relationship (QSAR) approaches. The dataset consist the half-wave potential of 40 benzoxazine derivatives which were obtained by DC-polarography. Descriptors which were selected by stepwise multiple selection procedure are: HOMO energy, partial positive surface area, maximum valency of carbon atom, relative number of hydrogen atoms and maximum electrophilic reaction index for nitrogen atom. These descriptors were used for development of multiple linear regression (MLR) and artificial neural network (ANN) models. The statistical parameters of MLR model are standard errors of 0.016 and 0.018 for training and test sets, respectively. Also, these values are 0.012 and 0.017 for training and test sets of ANN model, respectively. The predictive power of these models was further examined by leave-eight-out cross validation procedure. The obtained statistical parameters are $Q^2 = 0.920$ and SPRESS = 0.020 for MLR model and $Q^2 = 0.949$ and SPRESS = 0.015 for ANN model, which reveals the superiority of ANN over MLR model. Moreover, the results of sensitivity analysis on ANN model indicate that the order of importance of descriptors is: Relative number of H atom > HOMO energy > Maximum electrophyl reaction index for N atom > Partial positive surface area (order-3) > maximum valency of C atom.

Keywords: Benzoxazines; Half wave potential; Artificial neural network; pharmaceutical property; Quantitative structure-property relationship.

Introduction

Benzoxazine is the reaction product of amine, phenol and formaldehyde which can provide polymers with high glass transition temperature, low water absorption, excellent physical and electrical performances and excellent fire resistance properties (1). One of the interesting features of these compounds is the drug properties. These were used as bacteriocides, fungicides, antitumor agents, herbicides, microbiocides or anti-inflammatory agents, tyrosine mimetics, bacteriostatic, immunomodulating agents, neuroprotective antioxidants and were used as antituberculous agents in various cancer chemotherapy regimens for the treatment of sarcoma and cerebral tumors. (2-10).

Oxidation reactions played an important role in establishing structures and chemical properties of benzoxazines. Since these reactions are the most common pathways at the first phase of drug biotransformation. Therefore the electrochemical half wave potential of benzoxazines can be directly useful for investigation of their biological properties (11). Drugs with low oxidation potential can operate more successfully in the treatment of disease. A common technique in studies of electro-oxidations of benzoxazines is the voltammetric method (12). This method is the electrochemical technique. Since generation and examination of new drugs based on benzoxazine and its investigation by the voltammetric method is limited in time and cost (13) therefore, the development of theoretical model to predict the properties of these compounds are interesting and necessary. Quantitative structure activity relationship (QSAR) methods enable in prediction and interpretation of the properties and activities of a wide range of drugs and organic compounds based on the correlation between their properties and molecular characteristics (molecular descriptors) (14-18). There are some reports about the applications of QSPR/QSAR in electrochemistry (19-22). Wei *et al.* studied the relationship between the reduction characteristics and molecular structure of 87 chlorinated aromatics, such as naphthalenes, biphenyls, benzenes and phenols (19). Shamsipur and Hemmateenejad employed principle component regression (PCR) and principle component artificial neural network (PC-ANN) models in QSPR study of $E_{1/2}$ of some organic compounds. Their best PC-ANN model can explain the 96% of variances in the $E_{1/2}$ data (22). Fatemi *et al.* applied support vector machine (SVM) for the prediction of selectivity coefficients of anion-selective electrode for some univalent anions. The calculated root-mean-square errors of SVM for training and test set of their model are: 0.878 and 0.890 respectively. The correlation coefficients of training and test set are: 0.95 and 0.94, respectively. Also the obtained statistical parameters of cross-validation test on SVM model were: $Q^2 = 0.858$ and $SPRESS = 1.050$, which revealed the reliability of their model (23). Gallegos *et al.* developed some models to predict the logarithm of minimum inhibitory concentration (log MIC) of a subset of 39 substituted benzoxazines using a quantum molecular similarity approach (24). Nesmerak *et al.* was used Hammett substituent constants (25) for calculation of half-wave potentials for 40 benzoxazine (20). Toropov *et al.* calculated optimal descriptors with simplified

molecular input line entry system (SMILES) notation of some Nesmerek group chemicals (26) and used QSPR modeling to calculate the electrochemical half-wave potentials of these compound (26).

In the present work, we tried to predict the half-wave potential of some benzoxazine derives from their molecular descriptors by using artificial neural network (ANN) and multiple linear regression (MLR) techniques.

Experimental

Software

The 3D structures of the studied compounds were optimized using semi-empirical quantum-chemical methods of AM1 in HyperChem (Ver.7) package (27). The structural descriptors are numerical values that encode structural features of the molecular structures. In the present work, the CODESSA software was used to calculate 393 constitutional, topological and geometrical descriptors (28). Then heuristic method (HM) was used to search the best set of descriptors for multilinear correlations (29).

Artificial neural networks are mathematical systems that simulate biological neural networks (30-32). They consist of processing elements (nodes or neurons) which organized in some layers. Back-propagation neural networks are most often used in analytical applications. The back-propagation network receives a set of inputs, which are multiplied by each node and then a nonlinear transfer function is applied for their processing. The goal of training the network is to change the weights between the layers in a direction to minimize the output errors. More details about the theory of the neural networks have been adequately described in many literatures (32-38). The ANN programs were written in FORTRAN 77 in our laboratory. A three-layer network with a sigmoid transfer function was design for each ANN. The generated artificial neural network uses descriptors selected by HM as inputs. The number of nodes in the input layer is dependent on the number of descriptors introduced in the network. The number of nodes in the output layer for both subsets (training and test sets) was set to be one. The inputs and outputs values of ANN were normalized between 0.1 and 0.9. The initial weights were selected randomly between -0.3 and 0.3 . The number of nodes in the hidden layers, learning rate, and momentum would be optimized before training the network. During the training of ANN, the values of weights and biases continuously changed to minimize the differences between ANN outputs and desired activity/property values, using the back propagation of errors. In order to evaluate the performance of the ANN, the standard error of training (SET) and the standard error of prediction (SEP) were used. The training iteration was stopped at overtraining point, where SEP is started to increase. Then the trained network was used to calculation the $E_{1/2}$ values of test set. In order to further investigation of the credibility of obtained ANN model leave-eight-out cross validation method was used. Finally, the sequential zeroing weight (SZW) approach was used for evaluation of the relative importance of selected molecular descriptors.

Dataset and Molecular descriptors

New groups of antimycobacterial agents that were studied in the present work as dataset are derivatives of benzoxazine which their half-wave potentials were obtained from reference (20). The electrochemical measurements on these compounds were performed by an EKO-Tribo Polarograph. The reference electrode was a silver plate which immersed in a solution of acetonitrile that consists 0.01M of AgNO₃ and 1M of NaClO₄. The chemical structures and experimental oxidation half-wave potential of these compounds are shown in Figure 1 and Table 1. The 3D structures of the studied compounds were optimized using semi-empirical quantum-chemical methods of AM1. The data set was separated into two groups: training and test sets. All molecules were placed by Y-ranking method in these sets. The training set, consisted of 35 molecules, was used for the model generation and the test set, consisted of 5 molecules, was used to take care of the overtraining and evaluate the prediction power of the generated model.

Data screening and descriptor selection

The CODESSA software was used to calculate constitutional, topological and geometrical descriptors. Then heuristic method (HM) was used to search the best set of descriptors for multilinear correlations. In the first step of this method, descriptors with constant values for all molecules were eliminated from the pool of descriptors. Also, pairs of variables with a correlation coefficient greater than 0.90 were classified as intercorrelated variables, and only one of them was used in developing the model. In deleting one descriptor from one pair of correlated descriptors we tried to keep descriptor which has these criteria: 1) has the higher correlation with independent variable (half wave oxidation potential), 2) its calculation is simpler and easier, 3) has more information about interested activity/properties, 4) is more interpretable. At the end of this step total of 284 descriptors were reminded to further investigations.

A major decision in developing successive QSPR model is when to stop adding descriptors to the model during the forward selection procedure. A simple technique to control the model expansion is the 'break-point' procedure (39). In this method, improvement of the statistical quality of the models is analyzed by plotting the squared correlation coefficient values (R^2) of the obtained models versus the number of descriptors involved in each model. Consequently, the model corresponding to the break point is considered as the best/optimum model.

Thus, HM procedure was applied to the training set and multilinear regression equations of up to 16 descriptors were developed. Variations of R^2 against the number of descriptors in the models were recorded and are shown in Figure 2. The application of the break-point algorithm led to the conclusion that the best model had five parameters. The specifications of this model w shown in Table 2. Then the artificial network was used to calculation the $E_{1/2}$ values of training and test set, respectively. Also, in order to further investigation of the credibility of obtained ANN model leave-8-out cross validation

method was used. Finally, the sequential zeroing weight (SZW) approach was used for evaluation of the relative importance of selected molecular descriptors.

Result and discussion

Molecular diversity validation

Diversity is a fundamental research subject in chemical database analysis of sampling (40). Molecular diversity analysis explores the way of molecules to cover a determined structural space and underlies many approaches for compound selection and design of combinatorial libraries. The diversity problem involves defining a diverse subset of “representative” compounds so that researchers can scan only a subset of the huge database each time. Therefore, the choice of an optimal metric space that represents the structural diversity of a compound population is determinant in the efficiency of the model (38, 41). In this work, diversity analysis was done for the data set to make sure the structures of the training or test sets could represent those of the whole ones.

For a database of n compounds generated from m highly correlated chemical descriptors, a distance score (d_{ij}) for two compounds X_i and X_j can be measured by the Euclidean distance norm based on the compounds descriptors:

$$d_{ij} = \|X_i - X_j\| = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (1)$$

Each compound X_i is represented as a vector:

$$X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{im})^T \text{ for } i = 1, 2, \dots, n$$

where x_{ij} denotes the value of descriptor j of compound X_i and T indicates vector transposition. The mean distances () of one sample to the remaining ones were computed as follow:

$$\bar{d}_{ij} = \frac{\sum_{j=1}^n d_{ij}}{n-1} \quad (2)$$

Then the mean distances were normalized within the interval (0, 1). In our data sets, the mean distances of samples versus oxidation half-wave potential are plotted in Figure 3. The distribution of points in this figure illustrates the diversity of the molecules in the training and test sets. As can be

seen from this figure, the structures of compounds are diverse in the training and test sets. The training set with a broad representation was adequate to ensure model stability.

Linear modeling

The SPSS software (Ver. 14) was used to developing many MLR models (43). The best model was selected based on the statistics of correlation coefficient (R), standard error (SE) and Fisher-statistics value (F). Consequently, among different models, the five-parameter model was chosen based on the break point procedure. Descriptors which were selected by this method are: high occupied molecular orbital energy(HOMO), partial positive surface area, maximum valency of carbon atom, relative number of hydrogen atoms and maximum electrophilic reaction index for nitrogen atom that have shown in Table 2.

Multicollinearity for the selected parameters (descriptors) was also checked and its result was presented in Table 3. As can be seen in this table there are not any high correlation between these descriptors. Then the MLR model was used to calculate of $E_{1/2}$ for test set as well as training set. The MLR predicted values of $E_{1/2}$ were shown in Table 1. Finally, the leave 8-out cross-validation (L8O) was used to evaluate credibility and robustness of these models. The statistical parameters of this test were shown in Table 4. Other statistical parameters of MLR model are: average error = 0.0002, relative error = 0.0022 and absolute error = 0.0102, respectively.

Non-linear modeling

A three-layer network with a sigmoid transfer function was designed for ANN model. The network was trained using the training set by the back propagation strategy for optimization of the weights and bias values. To obtain the best result the weight and bias learning rate and momentum value as well as ANN's topology were optimized. The procedure for optimization of ANN's parameters is given elsewhere (37, 38). The optimized values of these terms and ANN characteristics are given in Table 5. Then the constructed ANN model was used to calculate the $E_{1/2}$ for test set as well as training set. The predicted values of $E_{1/2}$ by ANN model were shown in Table 1. Moreover, the leave-8-out cross-validation (L8O) was used to evaluate the credibility and robustness of the ANN model. The statistical parameters of this test were shown in Table 4. Other statistical parameters of ANN model are, average error = 0.0046, relative error = 0.0040 and absolute error = 0.0137, respectively. In comparison whit MLR statistical parameters and other statistical values in Table 4, it can be seem that the performance of ANN model was better than MLR ones. Figure 4 indicates the variation of ANN predicted against experimental values of $E_{1/2}$ that the agreement between the predicted and experimental values is clear ($R_{(training\ set)}=0.0983$ and $R_{(test\ set)}=0.971$). Also, the residual values between ANN predicted and experimental values of half-wave electrooxidation potential of benzoxazines were traced in Figure 5.

The random distribution of residuals about zero line confirms that there is no systematic error in developed ANN model. To verify the chemical domain of the consensus model and the distribution of the studied chemicals in this new multidimensional space, the chemicals are plotted in a principal components 3D-graph (Figure 6), which was obtained by applying PCA on all molecular descriptors used by these models. This PCA plot shows that chemicals have fine distribution in the molecular descriptors domain.

Sensitivity analysis and descriptor interpretation

By interpreting selected descriptors in the ANN model, it is possible to gain some insight into the factors that are likely to govern the $E_{1/2}$ of benzoxazines. Here, a brief interpretation of these factors in order to determine the relative importance of each variable is given based on the results of sensitivity analysis. The procedure of this approach is based on the sequential removal of variables by zeroing the specific connection weights for that specific input variable in the first layer of the ANN (44). For each sequentially zeroed input variable, root mean square error of prediction set (RMSEP) as the prediction error of this network was calculated. Generally RMSEP value increases in this way. Then, differences between RMSEP and root mean square error of established ANN (RMSE) was calculated and shown as DRMSE. Each variable which causes greater value of DRMSE is more important. The DRMSE values are shown for each descriptor in Figure 7. As the mentioned earlier, five descriptors were used for ANN model to comprise: relative number of H atom, HOMO energy, maximum electrophyl reaction index for N atom, partial positive surface area (order-3), maximum valency of C atom that belonging for constitutional, quantum chemical and charge descriptors and encode electronic aspects of the molecular structure. The order of importance of descriptors is: Relative number of H atom > HOMO energy > Maximum electrophyl reaction index for N atom > Partial positive surface area (order-3) > maximum valency of C atom.

First important descriptor in the model is relative number of H atom that is a simple constitutional type descriptor. This factor indicates the size of molecules as well as the degree of saturation of molecule. The second one is the highest occupied molecular orbital energy which is belonging to quantum chemical descriptors and determines the needed energy to drawing the electron in oxidation process (45). Molecule with high HOMO energy values can donate its electron more easily than the molecule with lower HOMO value, and hence is more reactive (26). Next descriptor is maximum electrophylic reaction index for N atom that is the quantum chemical descriptor too. This index provides feasible chemical interaction with electrophilic attack as electron affinity (46) and is important in molecular properties and reactivity in particular for radical reactions.

The forth descriptor is partial positive surface area (order-3) which is a charge descriptor and contains the electronic and structural information of molecule (46). This descriptor encodes the solvent accessible surface area of molecule in electrochemical reaction, and can well estimate the absolute hardness and can affect on electrooxidation of benzoxazines. The last descriptor is maximum

valency of C atom. This parameter is a charge type descriptor which can affect on electron affinity of the molecule and therefore can correlate to the $E_{1/2}$ of a molecule. Thus these descriptors can encode different aspects of molecules which can effect on their $E_{1/2}$ values.

Conclusion

The obtained results indicate that QSPR approaches can be used to predict the electrooxidation half-wave potentials of benzoxazine derivatives from their structural molecular descriptors. Also, comparison between statistical parameters of ANN and MLR models indicates that the ANN model produces better results due to non-linear characteristic of ANN. Finally, total descriptors which were appeared in this model can encode features of molecules which were responsible in electrooxidation characteristics of these molecules.

References

1. Burke WJ, Morteson-Glennie EL and Weatherber C. Condensation of halophenols with formaldehyde and primary amines. *J. Org. Chem.* (1964) 29:909- 912.
2. Burke WJ, Weatherbee C, Lau H, Van Lear G and Goken G. New Catalysts for the Oxidation of Sulfides to Sulfones with Hydrogen Peroxide. *J. Am. Chem. Soc.* (1963) 28:1098- 1100.
3. Kuehne ME and Konopka EA. Dihydro-1, 3-oxazine as antitumor agents. *J. Med. Pharm. Chem.* (1962) 5:257-280.
4. Moloney GP, Craik DJ, Iskander MN. Qualitative analysis of the stability of the oxazine ring of various benzoxazine and pyridooxazine derivatives with proton nuclear magnetic resonance spectroscopy. *J. Pharm. Sci.* 81(7) (1992) 692-297.
5. Colin JL and Loubinoux B. Nouvelle voie d'accès aux dihydro-3, 4-2H-benzoxazines-1,3. *Tetrahedron Lett.* (1982) 23:4245- 4246.
6. Dewar MK, Johns RB, Kelly DP, Wei-Let S and Yates JF. Cross-Linking of Amino Acids by Formaldehyde. Preparation and ^{13}C NMR Spectra of Model Compounds. *Adv. Exp. Med. Biol.* (1977) 86:641-647.
7. Wagner G, Singer D and Weuffen W. Untersuchungen uber 2-hydroxythiobenzamide und 2-hydroxythiobenzanilide. *Pharmazie.* (1966) 21:161-166.
8. Fringuelli R, Schiaffella F and Vecchiarelli A. Antifungal and immunomodulating activities of 1, 4-benzothiazine azole derivatives: review. *J. Chromatogr.* (2001) 13:9-14.
9. LARGERON M, LOCKHART B, PFEIFFER B, FLEURY MB. Synthesis and *in-vitro* evaluation of new 8-amino-1,4-benzoxazine derivatives as neuroprotective antioxidants. *J. Med. Chem.* (1999) 42:5043–5052.
10. Paci A, Martens T and Royer J. Anodic oxidation of ifosfamide and cyclophosphamide: a biomimetic metabolism model of the oxazaphosphorinane anticancer drugs. *Bioorg. Med. Chem. Lett.* (2001) 11:1347-1349.
11. Guengerich FP, Willard RJ, Shea JP, Richards LE and Macdonald TL. Mechanism-based inactivation of cytochrome P-450 by heteroatom-substituted cyclopropanes and formation of ring-opened products. *J. Am. Chem. Soc.* (1984) 106:6446-6447.
12. Scholz F, Schroder U and Gulaboski R. *Electrochemical Methods: Guide to Experiments and Applications*, Springer – Verlag, Berlin Heidelberg, Germany (2005).
13. Macchiarulo A, Constantino G, Fringuelli D, Vecchiarelli A, Schiaffella F and Fringuelli R. 1, 4-Benzothiazine and 1, 4-benzoxazine imidazole derivatives with antifungal activity: a docking study. *Bioorg. Med. Chem.* (2002) 10:3415-3423.
14. Fatemi MH. QSPR Prediction of Migration Index in Microemulsion Electrokinetic Chromatography Using Artificial Neural Networks. *J. Chromatogr. A.* (2003) 1002:221-229.
15. Fatemi MH. Prediction of ozone tropospheric degradation rate constant of organic compounds by using artificial neural network, *Anal. Chim. Acta* (2006) 556:355-363.
16. M. Karelson, A. Perkson, QSPR Prediction of Densities of Organic Liquids, *Comput. Chem.*, 23(1999) 49-59.

17. Kaliszán R, Kaliszán A, Noctor TAG, Purcell WP and Wainer IW. Mechanism of retention of benzodiazepines in affinity, reversed phase and adsorption high performance liquid chromatography in view of quantitative structure retention relationships. *J. Chromatogr.* (1992) 609:69-81.
18. Hansch C, Kurup A and Gao H. Chem-bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms. *Chem. Rev.* (2001) 101:619-672.
19. Wei D, Hu H, Hiramoto K and Fujie K. Determination and QSPR study on the electrochemical reduction characteristics of chlorinated aromatic compounds. *Huanjing Kexue.* (2003) 24:19-22.
20. Nesmerak K, Nemecc L, Sticha M, Waisser K and Palat K. Quantitative structure–property relationships of new benzoxazines and their electrooxidation as a model of metabolic degradation. *Electrochim. Acta.* (2005) 50:1431-1437.
21. Tompe P, Clementis G, Petnehazy L and Toke L. Quantitative structure-electrochemistry relationships of α , β -unsaturated ketones. *Anal. Chim. Acta.* (1995) 305:295-303.
22. Hemmateenejad B and Shamsipur M. Quantitative Structure-Electrochemistry Relationship study of Some Organic compounds Using PC-ANN and PCR. *Internet Electronic Journal of Molecular Design.* (2003) 3:316-334.
23. Fatemi MH, Gharaghani S, Mohammadkhani S and Rezaie Z. Prediction of selectivity coefficients of univalent anions for anion-selective electrode using support vector machine. *Electrochim. Acta.* (2008) 53:4276-4282.
24. Gallegos A, Carbo-Dorca R, Pones R and Wiasser K. milarity approach to QSAR: Application to antimycobacterial benzoxazines. *Int. J. Pharm.* (2004) 269:51-60.
25. Hansch C, Leo A and Hoekman D. *Exploring QSAR, Hydrophobic, electronic and steric constants*, American Chemical Society, Washington, USA (1995)
26. Toropov A, Nesmerak K, Raska JrI, Waisser K and Palat K. QSPR modeling of the half-wave potentials of benzoxazines by optimal descriptors calculated with the SMILES. *Comp. Biol. Chem.* (2006) 30:434-437.
27. Hyperchem, Molecular Modeling System, Hyper Cube, Inc. and Auto Desk, Inc., Developed by Hyper Cube, Inc, 1993.
28. Katritzky AR, Lobanov VS and Karelson M. CODESSA Version 2.0 Reference Manual, 1995–1997.
29. Katritzky AR, Perumal S, Petrukhin R and Kleinpete E. CODESSA–based theoretical QSPR model for hydantoin HPLC–RT lipophilicities. *J. Chem. Inf. Model.* (2001) 41:569-547.
30. Patterson DW. *Artificial Neural Networks: Theory and Applications*, Simon and Schuster, New York, UAS (1996)
31. Zupan J and Gasteiger J. Neural networks: A new method for solving chemical problems or just a passing phase. *Anal. Chim. Acta.* (1991) 248:1-30.
32. Bose NK and Liang P. *Neural Network-Fundamentals*, McGraw-Hill, New York, USA (1996)
33. Jalali-Heravi M and Fatemi MH. Simulation of mass spectra of noncyclic alkanes and alkenes using artificial neural network. *Anal. Chim. Acta.* (2000) 415:95-103.
34. Jalali-Heravi M and Fatemi MH. Artificial neural network modeling of Kovats retention indices for noncyclic and monocyclic terpenes. *J. Chromatogr. A.* (2001) 915:177-183.
35. Jalali-Heravi M and Fatemi MH. Prediction of flame ionization detector response factors using an artificial neural network. *J. Chromatogr. A.* (1998) 825:161-169.
36. Fatemi MH. Simultaneous modeling of Kovats retention indices on OV-1 and SE-54 stationary phases using artificial neural networks. *J. Chromatogr. A.* (2002) 955:273-280.
37. Haykin S. *Neural Network*, Prentice Hall, Englewood Cliffs, NJ (1994)
38. Beal MT, Hagan HB and Demuth M. *Neural Network Design*, PWS, Boston, USA (1996)
39. Kartitzky AR, Dobchy DA, Hur E, Fara DC and Karelson M. QSAR treatment of drugs transfer into human breast milk. *Bioorg. Med. Chem.* (2005) 13:1623-1632
40. Maldonado AG, Doucet JP, Petitjean M and Fan BT. Molecular similarity and diversity in chemoinformatics: from theory to applications. *Molec. Divers.* (2006) 10:39-79.
41. Richon AB. *A history of computational chemistry*, Network Science, Kidlington, UK (1996)
42. Rouvray DH. The pioneering contributions of Cayley and Sylvester to the mathematical description of chemical structure. *J. Chem. Inf. Model.* (1992) 32:580-586.
43. SPSS/PC, Statistical Package for IBMPC, Quiad software, Ontario (1986)
44. Nord LI and Jacobsson SP. A novel method for examination of the variable contribution to computational neural network models. *Chemometr. Intell. Lab* . (1998) 44:153-160.
45. Karelson M and Lobanov VS. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* (1996) 96:1027-1043.
46. Todeschini R and Consonni V. *Handbook of Molecular Descriptors*, Wiley/VCH, Weinheim, USA (2000).

* **Corresponding author:**
E-mail: e.baher@gu.ac.ir

Table 1. Structures, experimental, MLR and ANN-predicted values of oxidation half-wave

| Derivate | X | R1 | R2 | $E_{(1/2)}\text{-Exp}$ | $E_{(1/2)}\text{-MLR}$ | $E_{(1/2)}\text{-ANN}$ |
|-----------------|---|--------------------|---------------------|------------------------|------------------------|------------------------|
| 1 | O | 7-OCH ₃ | | 1.420 | 1.413 | 1.419 |
| 2 | O | 7-OCH ₃ | 4-F | 1.430 | 1.434 | 1.424 |
| 3 | O | 7-OCH ₃ | 4-Br | 1.440 | 1.465 | 1.453 |
| 4 | O | 7-OCH ₃ | 3-F | 1.445 | 1.458 | 1.441 |
| 5 | O | 7-OCH ₃ | 3-Cl | 1.450 | 1.461 | 1.447 |
| 6 | O | 7-CH ₃ | 4-CH ₃ | 1.415 | 1.412 | 1.418 |
| 7 | O | 6-CH ₃ | 4-CH ₃ | 1.420 | 1.416 | 1.423 |
| 8 ^T | O | | 4-Br | 1.490 | 1.522 | 1.533 |
| 9 | O | 6-OCH ₃ | 4-CH ₃ | 1.450 | 1.425 | 1.433 |
| 10 | O | 6-OCH ₃ | 4-F | 1.460 | 1.480 | 1.460 |
| 11 | O | 6-OCH ₃ | 4-Br | 1.465 | 1.487 | 1.474 |
| 12 | O | 6-OCH ₃ | 4-Cl | 1.470 | 1.487 | 1.473 |
| 13 | O | 6-OCH ₃ | 3-F | 1.480 | 1.481 | 1.462 |
| 14 | O | 6-OCH ₃ | 4-CN | 1.510 | 1.516 | 1.512 |
| 15 | O | 6-Cl | | 1.530 | 1.525 | 1.535 |
| 16 | O | 6-Cl | 3-Cl | 1.590 | 1.589 | 1.591 |
| 17 | S | 7-OCH ₃ | 4-CH ₃ | 1.280 | 1.312 | 1.309 |
| 18 | S | 7-OCH ₃ | | 1.315 | 1.323 | 1.318 |
| 19 | S | 7-OCH ₃ | 4-F | 1.350 | 1.348 | 1.366 |
| 20 | S | 7-OCH ₃ | 4-Br | 1.360 | 1.378 | 1.359 |
| 21 | S | 7-OCH ₃ | 4-Cl | 1.370 | 1.368 | 1.362 |
| 22 | S | 7-OCH ₃ | 3-F | 1.390 | 1.363 | 1.376 |
| 23 ^T | S | 7-OCH ₃ | 3-Cl | 1.395 | 1.370 | 1.364 |
| 24 | S | 7-OCH ₃ | 4-CF ₃ | 1.405 | 1.433 | 1.430 |
| 25 ^T | S | 7-OCH ₃ | 3,4-Cl ₂ | 1.420 | 1.417 | 1.408 |
| 26 | S | 7-CH ₃ | 4-CH ₃ | 1.305 | 1.323 | 1.308 |
| 27 ^T | S | 6-CH ₃ | 4-CH ₃ | 1.320 | 1.328 | 1.308 |

| | | | | | | |
|-----------------|---|--------------------|---------------------|-------|-------|-------|
| 28 | S | | 4-Br | 1.420 | 1.449 | 1.428 |
| 29 | S | 6-OCH ₃ | 4-CH ₃ | 1.330 | 1.336 | 1.326 |
| 30 | S | 6-OCH ₃ | | 1.360 | 1.353 | 1.350 |
| 31 | S | 6-OCH ₃ | 4-F | 1.380 | 1.395 | 1.403 |
| 32 | S | 6-OCH ₃ | 4-Br | 1.400 | 1.406 | 1.408 |
| 33 | S | 6-OCH ₃ | 4-Cl | 1.400 | 1.402 | 1.402 |
| 34 | S | 6-OCH ₃ | 3-F | 1.410 | 1.399 | 1.407 |
| 35 | S | 6-OCH ₃ | 3-Cl | 1.430 | 1.405 | 1.404 |
| 36 | S | 6-OCH ₃ | 4-CF ₃ | 1.440 | 1.455 | 1.451 |
| 37 ^T | S | 6-OCH ₃ | 3,4-Cl ₂ | 1.445 | 1.451 | 1.444 |
| 38 | S | 6-OCH ₃ | 4-CN | 1.450 | 1.437 | 1.438 |
| 39 | S | 6-Cl | | 1.420 | 1.443 | 1.420 |
| 40 | S | 6-Cl | 3-Cl | 1.520 | 1.498 | 1.503 |

T: denotes the test set.

Table 2. Specification of multiple linear regression model

| Name of descriptors | Symbol | Coefficient | SE | Mean effect |
|---|----------------|-------------|--------|-------------|
| Relative number of H atom | X ₁ | -0.13 | ±0.027 | -0.796 |
| Partial positive surface area(order-3) | X ₂ | -0.1 | ±0.004 | -0.086 |
| Maximum electrophyl reaction index for N atom | X ₃ | 0.023 | ±0.002 | 0.075 |
| HOMO energy | X ₄ | -0.079 | ±0.051 | 1.010 |
| Maximum valency of C atom | X ₅ | 2.298 | ±1.012 | 8.880 |
| Constant | | -7.903 | ±3.65 | |

n =35, R =0.97, SE = 0.016, F =512

Table 3. Internal correlation matrix between molecular descriptors

| | X ₁ | X ₂ | X ₃ | X ₄ | X ₅ |
|----------------|----------------|----------------|----------------|----------------|----------------|
| X ₁ | 1.000 | 0.255 | 0.075 | 0.650 | 0.670 |
| X ₂ | | 1.000 | -0.027 | -0.010 | -0.163 |
| X ₃ | | | 1.000 | -0.352 | 0.009 |
| X ₄ | | | | 1.000 | 0.253 |
| X ₅ | | | | | 1.000 |

Table 4. The statistical results of ANN and MLR models

| Models | Training set | | Test set | | Cross-validation Test | |
|--------|--------------|-------|----------|-------|-----------------------|--------|
| | R | SE | R | SE | Q ² | SPRESS |
| ANN | 0.983 | 0.012 | 0.971 | 0.017 | 0.949 | 0.015 |
| MLR | 0.969 | 0.016 | 0.970 | 0.018 | 0.920 | 0.020 |

R, SE, Q² and SPRESS are regression coefficient, standard error, correlation coefficient of cross validation and square of predictive error sum of squares respectively.

Table 5. Architecture of ANN

| Transfer Function | Sigmoidal |
|--------------------------|-----------|
| No. of Hidden Layer Nods | 2 |
| Weight Learning Rate | 0.2 |
| Bias Learning Rate | 0.6 |
| Momentum | 0.3 |
| No. of Input Layer Nods | 5 |
| No. of Output Layer Nods | 1 |

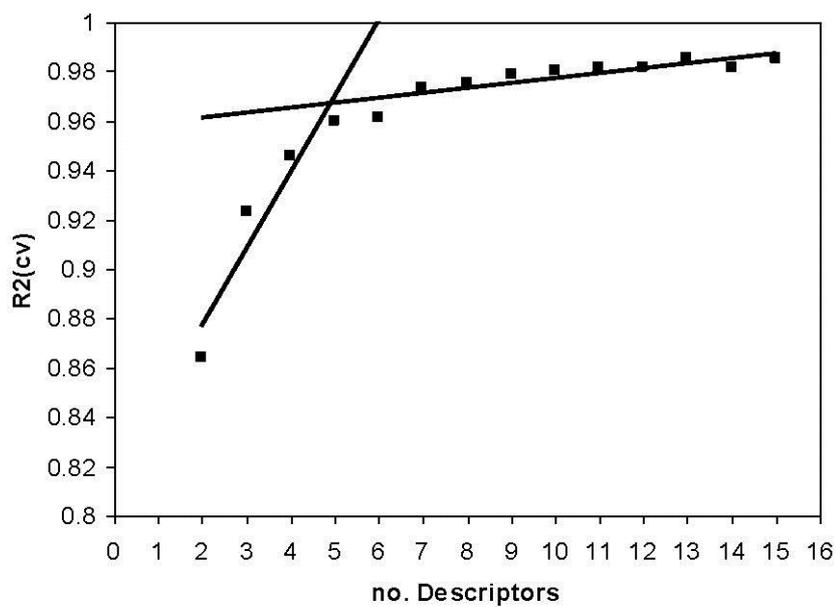


Figure 1. Plot of R^2 for the obtained models versus the number of descriptors involved.

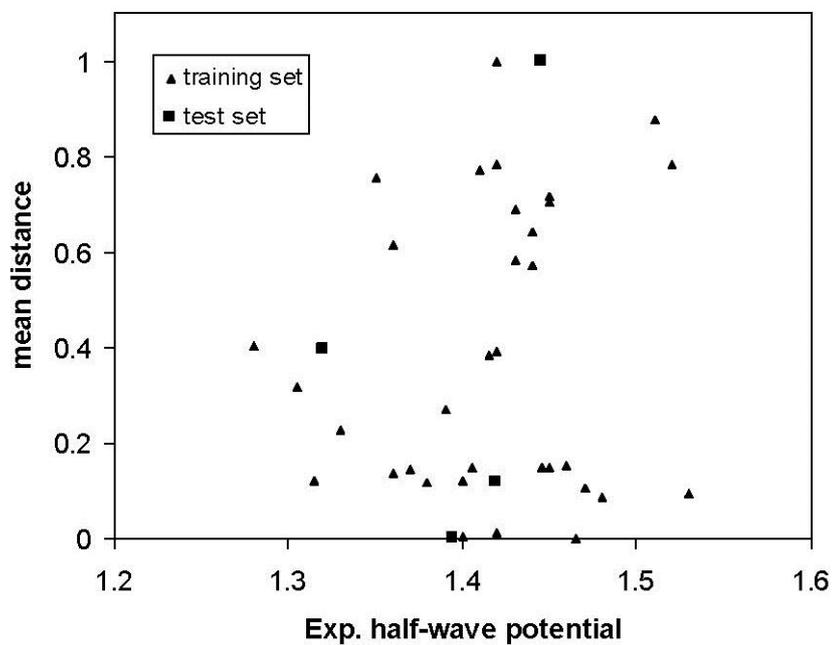


Figure 2. Scatter plot of samples for training and test sets according to the mean distances distribution.

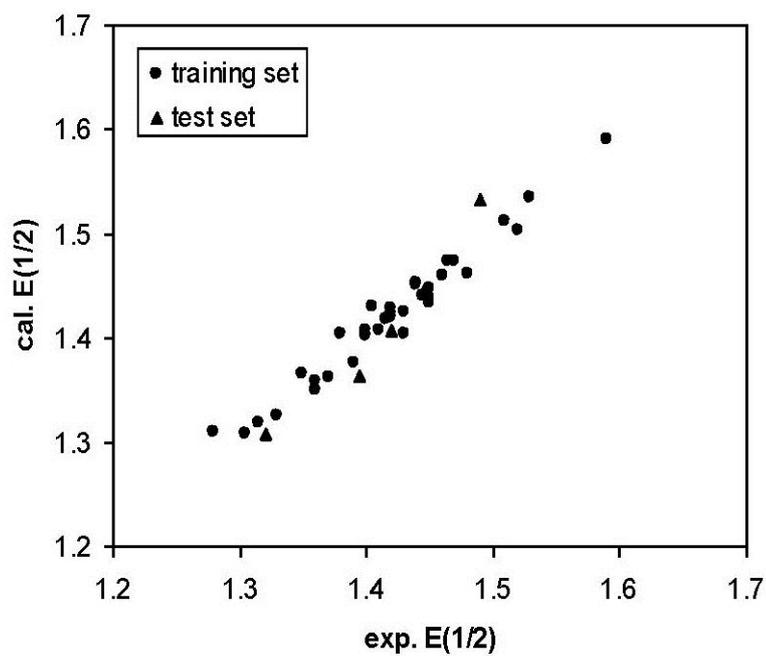


Figure 3. Calculated. $E_{1/2}$ versus Experimental $E_{1/2}$ plot.

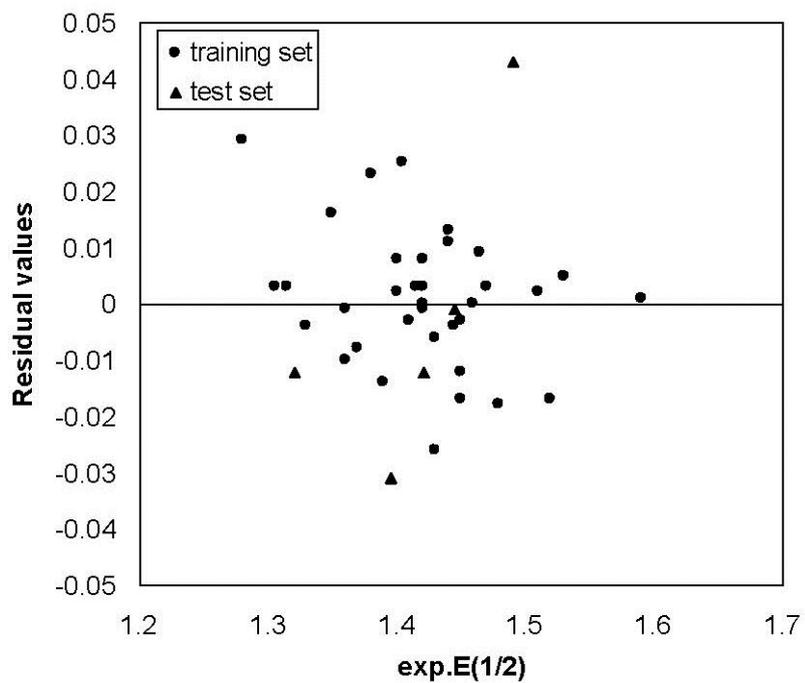


Figure 4. Residual versus Experimental $E_{1/2}$ plot.

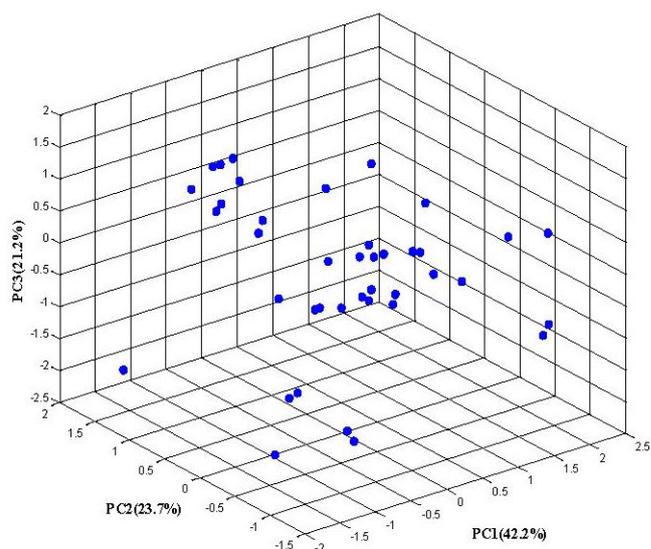


Figure 5. Principal component analysis on the selected molecular descriptors for the consensus model.

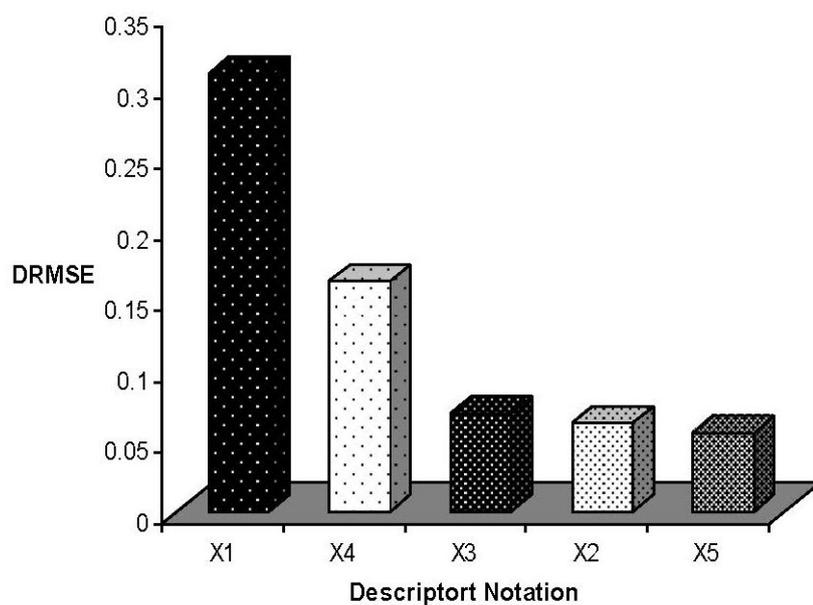


Figure 6. Sensitivity analysis results.