Int. J. Electrochem. Sci., 7 **(2012)** 4811 - 4821

International Journal of
ELECTROCHEMICAL
SCIENCE
www.electrochemsci.org

A Simple and Robust Model for Predicting the Reduction Potential of Quinones Family; Electrophilicity Index Effect

A. Beheshti, P. Norouzi[*]*, M. R. Ganjali*

Center of Excellence in Electrochemistry, Faculty of Chemistry, University of Tehran, Tehran, Iran
[*]E-mail: norouzi@khayam.ut.ac.ir

Received: 1 April 2012 / *Accepted:* 25 April 2012 / *Published:* 1 May 2012

This paper describes a procedure revealing the relationship between quantum chemical indices and electronic descriptors as independent variables and half-wave potential ($E_{1/2}$) of quinones family as a dependent variable. In this manner, three different quinone families including; benzoquinone, naphthoquinone and anthraquinone derivatives were subjected to structure–property relationship, quantitatively. Density functional theory (DFT) method at the level of B3LYP employing 6-31+g(d) basis set was performed to complete geometrical optimization. After calculating the electronic descriptors and quantum chemical indices for each compound, the quantitative relationship between the molecular property and the calculated descriptors was obtained by the multiple linear regression (MLR) and the features (or variables) were selected by genetic algorithm. The results showed that the electrophilicity index has significant correlation with reduction potential of quinones family. The accuracy of the proposed GA-MLR model was illustrated using the following evaluation techniques: cross-validation, validation through an external test set, and Y-randomization. This model, demonstrating high statistical qualities (R^2_{Train} = 0.983, Q^2_{LOO}= 0.979, Q^2_{LGO} = 0.976 and F_{Train} = 1051.100). An excellent agreement between the predictions and the experimentally obtained half-wave potential values for test set data was found (Q^2_{Ext}=0. 970), with a root mean square error of (RMSE) of 0.057 V.

Keywords: quantum chemical calculations, electrophylicity index, QSPR, half-wave potentials, quinones, chemometrics

1. INTRODUCTION

A wide range of quinones, for example anthraquinones, naphthoquinones, and benzoquinones, are now known and widely distributed in nature. Compounds containing a quinone group are well known to demonstrate various physiological activities as antibacterial [1] antifungal [2], antiviral [3], antimicrobial [4] and anticancer [5]. This is relevant, considering that a large part of the biological

activity of quinonoid systems is related to its capacity to generate free radicals via redox reactions. Biochemically, The electrogenerated radical anion species (semiquinone) have long half-life periods and able to transfer the electron to another species in vivo [6]. Because of this fact, considering the redox properties of quinones is important for understanding their mechanisms of action. Thus, the ability of calculating redox potentials accurately using the theoretical methods would be advantageous in a number of different areas, particularly where the experimental measurements are difficult due to the complex chemical equilibria, and reactions of the chemical species involved. Our literature survey showed that there have been several computational studies calculating redox potentials [6-12], in which the researchers have focused on computational methods to improve the calculation accuracy and decrease the prediction error, while the effects of chemical structure and substituent groups on the redox potential have not been investigated well.

Quantitative structure–activity/property relationship (QSAR/QSPR) studies are one of the most important areas in modern chemistry and biochemistry. The QSAR/QSPR is mathematical equations relating chemical structure to a wide variety of physical, chemical and biological properties. The main task of QSPR is to obtain a reliable statistical model for the prediction of properties/behaviors of new chemical substances and analytical systems. These relationships also take an approach to the identification and isolation of the most important structural descriptors that affect physicochemical properties. Thus, it is possible to select the most suitable compounds to be synthesized and tested in the laboratory. Hence, the QSAR/QSPR approach conserves resources and accelerates the process of the development of new molecules for use as drugs, materials, additives, or for any other purpose [13]. This provides information that is useful for molecular design and medicinal chemistry [14,15]. And also has used to prediction the corrosion inhibitory [16]. In continues of our previous success in developing QSPR studies for predicting different chemical and physical properties [17-19], the main goal of this study is selecting the best structural descriptors which differentiate the redox value of benzoquinone, naphthoquinone and anthraquinones derivatives, and propose a reliable model to predict half-wave potential of the quinones family.

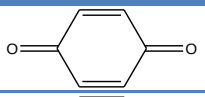
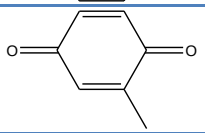
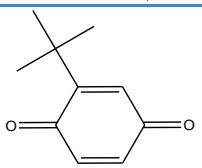
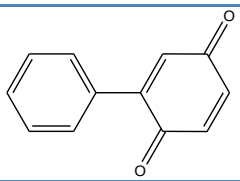
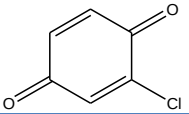
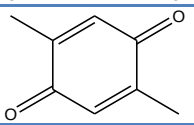
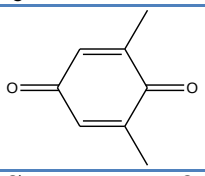
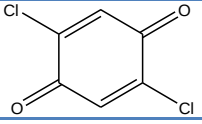
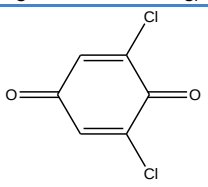
2. MATERIALS AND METHODS

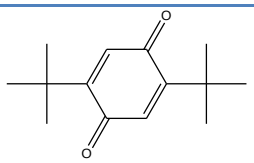
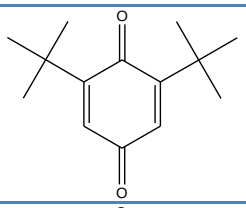
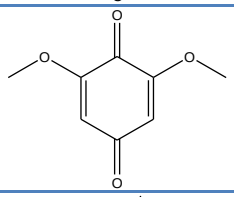
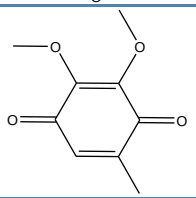
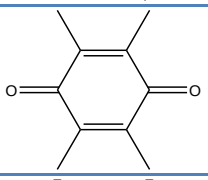
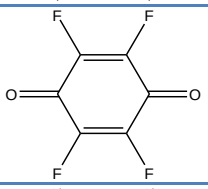
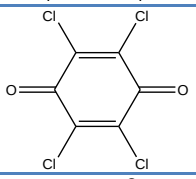
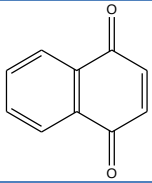
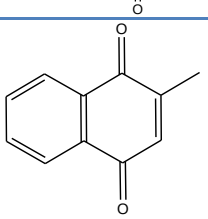
2.1. Data set and Data pretreatment

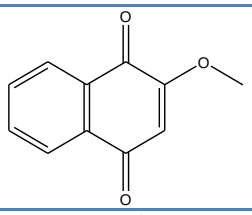
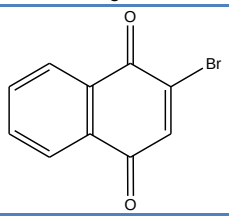
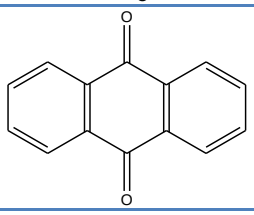
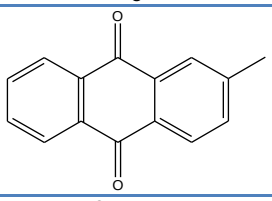
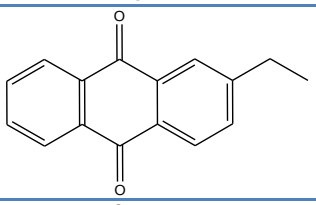
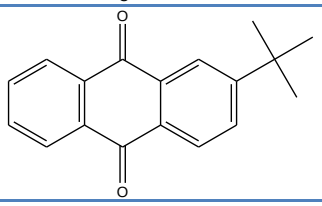
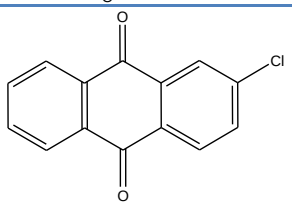
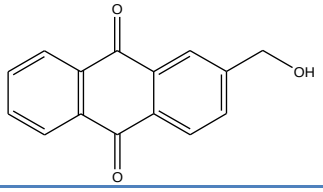
The half-wave potentials ($E_{1/2}$) of 26 1,4-benzoquinone, 1,4-naphthoquinone, and 9,10-anthraquinone derivatives were taken from the literature [6]. Structures and half-wave potentials taken from the literature are presented in Table 1. The reduction potential values have been measured in acetonitrile solvent. The first step to obtain a QSPR model is to encode the structural features of the molecules, which are named molecular descriptors. The successful application of the QSPR method needs to have a proper description of the variance between the individual molecular descriptors within a set of compounds. The choice is very important for groups of structurally similar congeners (‘congeners’ are defined as compounds having the same carbon skeleton but differing substitution patterns). Because the compounds in these groups are highly similar, the relative differences between the descriptor values for the data set are so small. Since the descriptors must be determined as

precisely as possible, the error in the calculated descriptor value must be significantly lower than the real variance of that descriptor between the congeners. Density function theory calculations are thus an attractive source of new and precise molecular descriptors, which can, in principle, express all the electronic and geometric properties of molecules and their interactions. So, in this work, because of the considerably large size of the studied molecules, density functional theory (DFT) method at the level of B3LYP employing 6-31+g(d) basis set was performed to complete geometrical optimization. Some electronic descriptors and quantum chemical indices for each molecule were obtained from the Gaussian output.

Table 1. The Structure, experimental and calculated values of half-wave potentials of quinone derivatives

No	Name	Structure	$E_{1/2}$(Exp.)	$E_{1/2}$(Calc.)
1	1,4-benzoquinone		-0.851	-0.808
2	2-methyl-1,4-benzoquinone		-0.928	-0.903
3[T]	2-tertbutyl-1,4-benzoquinone		-0.958	-0.926
4	2-phenyl-1,4-benzoquinone		-0.842	-0.806
5[T]	2-chloro-1,4-benzoquinone		-0.602	-0.692
6	2,5-dimethyl-1,4-benzoquinone		-1.002	-0.998
7	2,6-dimeth-yl-1,4-benzoquinone		-1.01	-0.989
8	2,5-dichloro-1,4-benzoquinone		-0.535	-0.565
9	2,6-dichloro-1,4-benzoquinone		-0.516	-0.565

10	2,5-ditertbutyl-1,4-benzoquinone		-1.059	-1.033
11[T]	2,6-ditertbutyl-1,4-benzoquinone		-1.074	-1.026
12	2,6-dimethoxy-1,4-benzoquinone		-1.05	-1.133
13	5-methyl-2,3-dimethoxy-1,4-benzoquinone		-0.99	-1.043
14	tetramethyl-1,4-benzoquinone		-1.175	-1.119
15[T]	tetrafluoro-1,4-benzoquinone		-0.358	-0.29
16	tetrachloro-1,4-benzoquinone		-0.340	-0.318
17	1,4-naphthoquinone		-1.029	-1.048
18	2-methyl-1,4-naphthoquinone		-1.113	-1.126

19[T]	2-methoxy-1,4-naphthoquinone		-1.163	-1.208
20	2-bromo-1,4-naphthoquinone		-0.92	-0.935
21	9,10-anthraquinone		-1.259	-1.261
22	2-methyl-9,10-anthraquinone		-1.316	-1.295
23	2-ethyl-9,10-anthraquinone		-1.298	-1.302
24	2-tertbutyl-9,10-anthraquinone		-1.3	-1.306
25	2-chloro-9,10-anthraquinone		-1.184	-1.165
26[T]	2-hydroxymethyl-9,10-anthraquinone		-1.283	-1.312

a: Half-wave potentials (V vs Fc^+ /Fc)
T: Test set

The calculated descriptors for each molecule are listed in Table 2. Eventually, the calculated descriptors were collected in a (26×19) data matrix where: n and m were the number of the compounds and the descriptors, respectively.

Table 2. List of descriptors that were used as input for the GA-MLR method

No.	Notation	Definition
1	EP	Exact polarizability
2	AP	Approximate polarizability
3	TDM	Total Dipole Moment
4	MNQ	Maximum of Negative Charges
5	SNQ	Sum of Negative Charges
6	ANQ	Average of Negative Charge
7	MPQ	Maximum of Positive Charges
8	SPQ	Sum of Positive Charges
9	APQ	Average of Positive Charge
10	RMSENQ	Root Mean Square Error of Negative Charges
11	RMSEPQ	Root Mean Square Error of Positive Charges
12	RMSETQ	Root Mean Square Error of Total Charges
13	HOMO	Highest Occupied Molecular Orbital
14	LUMO	Lowest Unoccupied Molecular Orbital
15	η	Hardness
16	μ	Chemical Potential
17	S	Softness
18	ω	Electrophilicity
19	χ	Electronegativity

2.2. Descriptor selection and model development

The calculated structural descriptors and the experimental half-wave potential values were analyzed with the aid of Genetic algorithms – multiple linear regression. Genetic algorithms (GAs) were introduced by Holland, and mimic nature's evolutionary method of adapting to a changing environment [20]. They are stochastic optimization methods and provide a powerful means to perform directed random searches in a large problem space as encountered in chemometrics and drug design. Each individual in a population is represented by a chromosome. After initialization of the first generation (step 1), the fitness of each individual is evaluated by an objective function (step 2). In the reproduction step (step 3), the genetic operators of parent selection, crossover and mutation are applied, thereby providing the first offspring generation. Iteration of steps 2 and 3 is performed until the objective function converges [21]. Multiple linear regression is one of the most used modeling methods in QSPR. According to Todeschini et al [22], the best fitness function, leave-one-out cross-validated correlation coefficient (Q^2_{LOO}), was used as criteria for evaluating the credibility of each model.

3. RESULTS AND DISCUSSION

The chemical structures of the three quinone families including;1,4-benzoquinone, 1,4-naphthoquinone, and 9,10-anthraquinone derivatives were optimized with B3lyp method employing 6-

31+g(d) basis set in gas phase, and the quantum-chemical descriptors were obtained from the Gaussian output. The calculated descriptors can be classified into four different electronic categories including: local charges, dipoles, orbital energies and the quantum chemical indices. the quantum chemical indices of hardness (η), softness (S), electronegativity (χ), and electrophylicity (ω), were calculated according to the method proposed by Thanikaivelan et al [23], are the important electronic futures used to describe stability, reactivity, chemical potential and other related properties of molecules [24]. Hardness has been used to understand chemical reactivity and stability of molecules [25,26]. Electronegativity was introduced by Pauling as a power of an atom in a molecule to attract electron to itself. Softness is a property of molecule that measures the extent of chemical reactivity. Electrophilicity was proposed by Parr et al. [27] as a measure of energy lowering due to maximal electron flow between donor and acceptor. After calculating the electronic descriptors for each compound, the data set of 26 compounds was randomly separated into training and test sets. The training set was used to build a model, and the test set was applied to validate the predicting power. The quantitative relationships between the molecular property and the calculated descriptors were obtained by the multiple linear regression and the features (or variables) were selected by genetic algorithm. In the model development step, leave-one-out cross-validation procedure (LOO-CV) was employed to evaluate the performances of the resulting models. In the LOO-CV procedure, n-1 samples from a total data set were used to construct a calibration set (assessment set), and to build a QSPR model between the selected descriptors and the half-wave potential. Then, the property of the left out sample was estimated by the designed model. This procedure was repeated until every sample in the total data set was used for a prediction. The Q^2_{LOO} was calculate for each subset of descriptors then the best model was chose based on the high value of this parameter. To avoid over parameterization of the model, an increase of Q^2_{LOO} values of less than 0.02 was chosen as the breakpoint criterion. Eventually, the following equations were obtained between the half-wave potential of quinone derivatives and their structural descriptors.

$$E_{1/2} = -2.532 + 2.615\omega \quad (1)$$

R^2_{Train}=0.983, R^2_{adj}=0.982, Q^2_{LOO}=0.979, Q^2_{LGO}=0.976, F_{Train}=1051.100, $RMSE_{Train}$= 0.034, R^2_{Test}=0.974, Q^2_{Ext}=0. 970, $RMSE_{Test}$= 0.057

where ω is the electrophilicity index, and some statistical quantities such as the; square of the correlation coefficient (R^2), square of the correlation coefficient of the leave one out cross validation (Q^2_{LOO}), square of the correlation coefficient of the leave group out cross validation (Q^2_{LGO}), adjusted R^2 (R^2_{adj}), rout mean square error (RMSE) of train and test sets and F-values are presented to validating the purpose model. The primary value of QSPR is its predictivity, that is, how well it is able to predict endpoint values of compounds not used to develop the correlation, i.e. not in the training set. Goodness of fit is determined by the square of the correlation coefficient, but this dose not show much about the capability of prediction. So, two main methods are applied to determine predictivity, internal

cross-validation and external validation method with a test set of compounds. From the internal validation technique, the value of Q^2_{LOO}=0.979 and Q^2_{LGO}=0.976 were determined an excellent internal prediction ability, but an inconsistency between internal and external predictivity was reported in a few QSAR/QSPR studies [28-30]. It was reported that, in general, there is no relationship between internal and external predictivity [31]: high internal predictivity may result in low external predictivity and vice versa. In order to test the external validation of proposed model 25% of the data (6 compounds out of 26) were selected as the external test set compounds. The half-wave potential values of these compounds were calculated using proposed model. The statistical external validation (R^2_{Test} = 0.974, Q^2_{Ext}=0.970, $RMSE_{Test}$=0.057) also confirmed the high satisfactory prediction ability for chemicals not used in the model development.

The statistical parameters in equation 1 describe the stability, robustness, satisfactory fitting and prediction ability of the resulted model to predict the half-wave potential of quinone family. The excellent agreement between the experimental and calculated values of half-wave potential for quinone compounds are demonstrate in Table 1 and Figure 1.

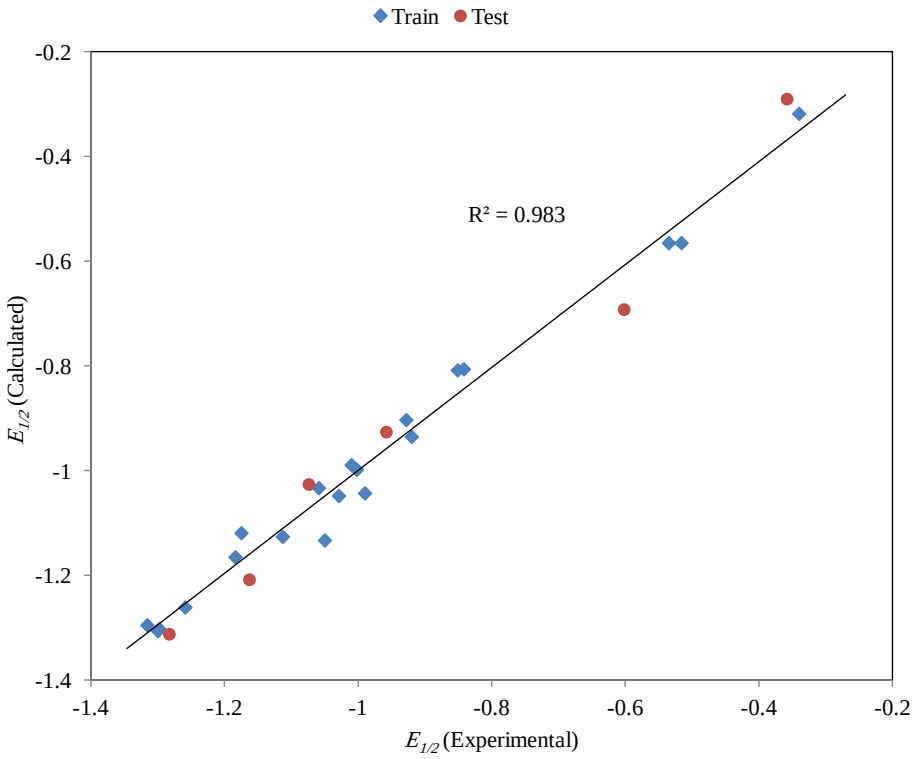


Figure 1. Plot of calculated values of $E_{1/2}$ against the experimental values of $E_{1/2}$

In the QSPR studies a model may contain descriptors which are statistically well correlated to **y** but in reality there is no cause-effect relationship encoded in the respective correlations with **y** because they are not related to the mechanism of action. The Y-randomization test was applied in this contribution to test the later. The dependent variable vector (half-wave potentials) was randomly shuffled and the original descriptors matrix is kept fixed then a new QSPR model was developed. The models obtained under such conditions should be of poor quality and without real meaning. The new

QSPR models (after several repetitions) were expected to have low R^2 and Q^2_{LOO} values. If the R^2 and Q^2_{LOO} values of these models were much lower than those of the original model, the model could be considered as reasonable, and had not been obtained by the chance. The average values of R^2 and Q^2_{LOO} after 10 repetitions were 0.016 and 0.060 respectively.

As even a robust, significant and validated QSPR model, it cannot be expected to reliably predict the modeled property for the entire universe of compounds, it is a domain of application that must be defined, and the predictions for only those compounds that fall in this domain can be considered as reliable. The Williams plot, the plot of the standardized residuals versus the leverage (h_i), was exploited to visualize the applicability domain [32]. In this plot, the horizontal and vertical straight lines indicate the limits of normal values: the first for the outliers and the second for influential compounds. A compound with leverage value more than warning leverage (h*) seriously influences the regression performance, but it doesn't appear to be an outlier because its standardized residual may be small, even though it has been excluded from the applicability domain. Moreover, a value of 3 for standardized residual is commonly used as a cut-off value for accepting predictions. The Williams plot for train and test sets was shown in Figure 2.

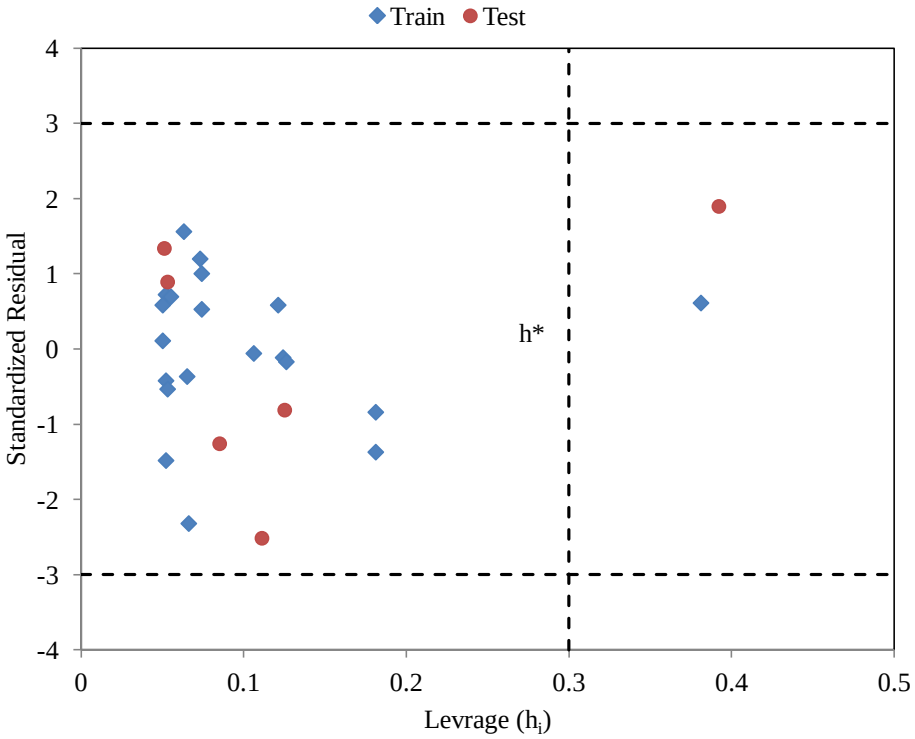


Figure 2. The Williams plot; the plot of the standardized residual versus the leverage (h_i)

The equation 1 shows that only electrophylicity index can explain 98% of variance between the half-wave potential of qoinune derivatives. Maynard et al. [33] suggested that electrophilicity index of a species is obtained by chemical potential (μ) squared divided by hardness (η) as following equation:

$$\omega = \mu^2 / 2\eta \quad (2)$$

Since, the chemical potential is equal to the negative electronegativity ($\mu = -\chi$) [23], the equation 2 rearranges as:

$$\omega = \chi^2 / 2\eta \quad (3)$$

The equation 1and 3 show that the half-wave potential of qoinune derivatives is related to the first-order η and second-order χ. In other word, the positive sign of the coefficient of ω in the equation 1 describes the increase in the electronegativity and the decrease in the hardness of molecule cause the molecule to be reduced at more positive potentials.

4. CONCLUSIONS

1. The main goal of the present work was discovering the main electronic factors of structure and the quantum chemical indices that significantly affect the half-wave potentials of quinone family. Density functional theory was performed to complete geometrical optimization and calculate structural features. Applying the GA-MLR methods proposed a simple and robust model between reduction potential and electrophilicity index; $E_{1/2}$ = -2.532 + 2.615ω.
2. We proposed that a specific quantum chemical index, electrophilicity, had excellent correlation with half-wave potential of quinones derivatives. Notably, the molecular index that correlated strongest with reduction potential was the ratio of electronegativity squared (χ^2) to hardness (η). So with increasing the electronegativity and the decreasing the hardness, the molecule is reduced at more positive potentials. The excellent predictive powers for both the internal and external sets showed the model could provide a useful tool to predict the half-wave potential of new quinone compounds.

References

1. J. R. E. Hoover and A. R. Dag, *J. Am. Chem. Soc.*, 76, (1954) 4148
2. N. G. Clark, *Pestic. Sci.*, 16, (1985) 23
3. K. W. Stagliano and A. Emadi, US20040087663A1 (2004)
4. S. A. Vichkanova, S. B. Izosimova, V. V. Adgina and L. D. Shipulina, Search for antimicrobial drugs among quinones of plant origin. *Rastitel'nye Resursy*, 15 (1979) 167
5. A. B. Pardee, Y. Z. Li and C. J. Li, *Cancer Drug Targets*, 2 (2002) 227
6. C. Frontana, A. Vazquez-Mayagoitia, J. Garza, R. Vargas and I. Gonzalez, *J. Phys. Chem. A,* 110 (2006) 9411
7. Y. Paukku and G. Hill. *J. Phys. Chem. A,* 115 (2011) 4804
8. R. Benassi, P. Ferrarini, C. Fontanesi, L. Benedetti and F. Paolucci, *J. Electroanal. Chem.,* 564 (2004) 231
9. K. L. Phillips, S. I. Sandler and P. C. Chiu, *J. Comp. Chem.*, 32 (2011) 226

10. M. Namaziana, P. Norouzi and R. Ranjbar, *J. Mol. Struct.*, 625 (2003) 235
11. D. Nematollahi, A. Taherpour, S. Jameh-Bozorghi, A. Mansouri, B. Dadpou, *Int. J. Electrochem. Sci.,* 5 (2010) 867
12. H. Mohammad-Shiri, M. Ghaemi, S. Riahi and A. Akbari-Sehat, *Int. J. Electrochem. Sci.,* 6 (2011) 317
13. M. Karelson, V. S. Lobanov and A. R. Katrizky, *Chem. Rev.*, 96 (1996) 1027
14. C. Hansch, A. Kurup, R. Garg and H. Gao, *Chem. Rev.*, 101 (2001) 619
15. B. Hemmateenejad, R. Miri, M. Akhond and M. Shamsipur, *Arch. Pharm. Pharm. Med. Chem.*, 10 (2002) 472
16. N. O. Eddy, S. R. Stoyanov and E. E. Ebenso, *Int. J. Electrochem. Sci.,* 5 (2010) 1127
17. S. Riahi, A. Beheshti, M. R. Ganjali and P. Norouzi, *Electrophoresis*, 29 (2008) 4027
18. S. Riahi, A. Beheshti, M. R. Ganjali and P. Norouzi, *Spectrochim. Acta Part A*, 74 (2009) 1077
19. A. Beheshti, S. Riahi, M. R. Ganjali and P. Norouzi, *J. Comp. Chem.,* (2012) DOI: 10.1002/jcc.22892
20. J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press (1975)
21. L. Terfloth and J. Gasteiger, *DDT*, Vol. 6, No. 15 (Suppl.) (2001)
22. R. Todeschini, V. Consonni, A. Mauri and M. Pavan, *Anal. Chim. Acta*, 515 (2004) 199
23. P. Thanikaivelan, V. Subramanian, R. J. Rao and B. U. Nair, *Chem. Phys. Lett.*, 323 (2000) 59
24. B. Hemmateenejad, M. Sanchooli and A. Mehdipour, *J. Phys. Org. Chem.*, 22 (2009) 613
25. R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.*, 105 (1983) 7512
26. R. G. Parr and P. K. Chattaraj, *J. Am. Chem. Soc.*, 113 (1991) 1854
27. R. G. Parr, L. V. Szentpaly, S. Liu. *J. Am. Chem. Soc.*, 121 (1999) 1922
28. E. Novellino, C. Fattorusso and G. Greco, *Pharm. Acta Helv.*, 70 (1995) 149
29. U. Norinder, *J. Chemom.*, 10 (1996) 95
30. H. Kubinyi, *A general view on similarity and QSAR studies. In Computer-Assisted Lead Finding and Optimization*, H. van de Waterbeemd, B. Testa and G. Folkers, (Eds.) VHChA and VCH: Basel, Weinheim, pp. 9-28, (1997)
31. H. Kubinyi, F. A. Hamprecht and T. Mietzner, *J. Med. Chem.*, 41 (1998) 2553
32. OECD, *Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships [(Q)SAR] Models*, Organisation for Economic Co-Operation and Development, Paris, France (2007)
33. A. T. Maynard, M. Huang W. G. Rice and D. G. Covel, *Proc. Natl. Acad. Sci.*, 95 (1998) 11578