Structure–electrochemistry relationship in non-aqueous solutions: Predicting the reduction potential of anthraquinones derivatives in some organic solvents ★

Fatemeh Honarasa [a,*], Saeed Yousefinejad [b,c,**], Somayeh Nasr [a], Mohsen Nekoeina [d]

[a] Department of Chemistry, Shiraz Branch, Islamic Azad University, Shiraz, Iran
[b] Department of Chemistry, Shiraz University, Shiraz, Iran
[c] Department of Chemistry, Farhangian University, Tehran, Iran
[d] Department of Chemistry, Payame Noor University (PNU), P.O.BOX 19395-3697 Tehran, Iran

ARTICLE INFO

ABSTRACT

A new strategy is proposed for the quantitative structure–electrochemistry relationship (QSER) in non-aqueous systems, based on the combination of theoretical properties of electrochemical analytes and empirical parameters of solvents. Modeling of the first reduction potential of some anthraquinone derivatives in three organic solvents (acetonitrile, N,N-dimethyl formamide and dimethylsulfoxide) was done using the proposed approach. Combining the properties of electrochemical solvent and solute in a single model, shows that the structural features of anthraquinones and also some properties of the solvents are effective in predicting the electrochemical behavior of this non-aqueous system. The correlation coefficients of cross validation and external test set were 0.93 and 0.96 respectively which indicate stability and prediction ability of the model. To the best of our knowledge, this is the first report on simultaneous including the features of analyte and solvent in a QSER. This work obtains a way to decrease hazardous experiences in non-aqueous solvents.

© 2015 Published by Elsevier B.V.

1. Introduction

It is clear that a chemical reaction could be affected by or controlled with some conditions such as concentration of reactant, pressure, temperature, rate of mass transfer and pH. Doing a reaction in the solution phase, makes it possible to control these conditions. However the water is the most common solvent, but utilizing non-aqueous solvents is inescapable in some cases. For example, low solubility and/or stability and also thermodynamic or kinetic possibility can limit a reaction in water.

One of the earliest applications of non-aqueous solutions was in electrochemistry for exploring new chemical possibilities and also obtaining information about the dynamic and static effects of solvent on different chemical processes [1].

In spite of these advantages, usage of organic solvents in electrochemical reactions (and other applications) could lead to problematic side effects on environment and human health. So nowadays the focus is on making "green" approaches including low-risk organic solvents, ionic liquids, supercritical fluids, immobilized solvents and even solventless processes. One of the other ways toward green approaches is decreasing trial-and-error experiences in non-aqueous solvents by estimating the analyte's electrochemical property using quantitative structure–property relationship (QSER) as a specific form of quantitative structure–property relationships (QSPR).

Different reports have been published previously on the prediction of electrochemical properties by focusing on the structure of the analytes which participate in the electrochemical reaction. One of the first works on this subject was by Driebergen et al. [2]. After that, Tompe and coworkers suggested a QSER to predict the half-wave potential ($E_{1/2}$) of α,β-unsaturated ketones in nongaseous acetonitrile [3]. Li et al. utilized some topological descriptors to model the $E_{1/2}$ of some different organic compounds [4]. As another example, Hemmateenejad and Shamsipur used principle component–artificial neural network and principle component regression to enhance the QSER of some organic substances [5]. Among other studies that utilized QSPR in different branches of electrochemistry, QSER of new sets of chemical compounds was focused by some researchers and others stressed on enhancement of the modeling by stronger regression or variable selection methods [3,6–14].

In QSER studies, two input data are required. The first is the desired electrochemical property for a set of electro-active analytes as the

dependent variable of QSER and the other is the structural descriptors of these analytes, which is expected to reflect the structural features accurately for use as independent variables. In the current work for the first time, a simple and informative approach is suggested for QSER in different non-aqueous solvents by considering both structural properties of analytes and some empirical scales of solvent. The proposed method is a way toward clarification of the involved solvent–electrochemical analyte interactions as well as determination of important structural features of analytes. 30 electrochemical mixtures of anthraquinones (Aq) in three organic solvents were chosen as a model to show the potential of the proposed approach.

2. Experimental

2.1. Data set

The first reduction potential of ten 9,10-anthraquinone derivatives in three different organic solvents (Acetonitrile, AN; dimethylformamide, DMF; dimethylsulfoxide, DMSO) was adapted from literature [15]. Reduction potential of these compounds has been calculated from the volt-ammograms obtained from cyclic voltammetry by the scan rates that varied from 20 mVs^{-1} to 1000 mVs^{-1}[15]. It has been proved previously that the reduction process of anthraquinones is done in two steps [16,17]. In most cases, the first reduction step of Aqs is reversible while the second one is quasi-reversible [6]. So, the scope of this work was focused on the reduction potential of the first step. The structure of 9,10-anthraquinone compounds under study is represented in Table S1 (Supporting information). The electrochemical analysis of ten Aqs in three organic solvents produced 30 electrochemical samples.

2.2. Descriptors and parameters

Theoretical descriptors were utilized to describe Aqs structures and empirical solvent-scales were used to include the information of non-aqueous solvents of electrochemical system in the model. Different categories of descriptors like topological, topological charge indices, constitutional, geometrical, connectivity, RDF, 3D MoRSE, WHIM, GETAWAY, functional group counts and charge descriptors were computed as the structural features of Aqs [18]. These descriptors had enough diversity and ability to show different aspects of Aqs' structures. To generate these descriptors, the chemical structures of the Aqs were drawn using the Hyperchem software (Version 7, Hypercube Inc., http://www.hyper.com, USA) and were optimized by semi-empirical AM1 method. Then, the structural descriptors of Aqs were extracted using DRAGON software (Milano Chemometrics and QSAR research group; http://michem.disat.unimib.it/chm/).

127 "Solvent empirical parameters" or "solvent scales" from different categories, "Equilibrium/Kinetic", "Spectroscopic" and "Multiparameter" [19,20] were used as the descriptors of each electrochemical solvent. The utilized solvent scales could cover most of inter- molecular interactions of which the solvent is capable. It is worthy to mention that the noted solvent scales have shown good ability in various QSPR models [21–24].

2.3. Variable selection and modeling

The calculated descriptors of Aqs and empirical scales of solvent were collected in a data matrix (**X**) whose number of rows and columns were the number of electrochemical samples and total number of descriptors respectively. After deleting constant and near-constant descriptors of our electrochemical system, the correlation of descriptors with each other was investigated, and those with high collinearity were detected. Among the collinear descriptors, one with the highest correlation with the reduction potential values (E^0) of Aqs was retained and the others were removed from **X**.

Before variable selection and modeling, 30 electrochemical samples were divided randomly to training and test sets to further evaluation of model. Stepwise multiple linear regression (SMLR) was utilized as the variable selection method. Among the models suggested by SMLR, the final model was chosen by considering their squared correlation coefficients of calibration (R$^2_{cal}$) and cross validation (Q^2). Cross validation, external test set and y-scrambling (permutation test) were used to evaluate the prediction ability and stability of the proposed QSER model and its applicability domain were defined with the aid of leverage and standardized residual concepts.

3. Results and discussion

3.1. Model development

After removing constant, near constant and collinear descriptors from independent variables' data matrix (**X**), 271 descriptors for each Aq and 121 empirical parameters for each electrochemical solvent were presented in **X** of the size of 30 × 392, where 30 is the number of electrochemical samples and 392 is the total number of descriptors.

6 electrochemical samples out of 30 (about 20%) were chosen randomly as the test set, while the remaining 24 samples (about 80%) were utilized as the training set. It is noteworthy that the E$^0_{CV}$ of electrochemical samples in the training set covered the E$^0_{CV}$ of the whole data set. The reduction potential range of samples in the training set was −1.283 to −0.514 V and the values of test set were in the range of −1.117 to −0.520 V. Some considerations regarding random selection of test set will be discussed in the next parts of the manuscript.

Stepwise regression was run on the original descriptor data matrix (dependent variables, **X**) and E$^0_{CV}$ data vector (dependent variables, **y**) and 5 different models were generated step by step. But, it is clear that some of these models may be over fitted. In order to select the most convenient one, the values of squared correlation coefficient of training set (calibration) and leave one out cross validation (LOO CV) were calculated for each model and are shown in the inset of Fig. 1. As it is seen in Fig. 1 (inset), the performance of model is not increased significantly in the models with more than 4 parameters. So a five parametric linear model would have the risk of over fitting and a model that contained four parameters should be chosen as the final optimum one:

$$E^0_{CV} = 4.989\,(\pm0.637) + 0.602\,(\pm0.143)\,Mor23m - 4.453(\pm0.432) \quad (1)$$
$$E1s - 229.221\,(\pm32.048)\,GATS4m - 0.054\,(\pm0.012)\,E_{NR}$$

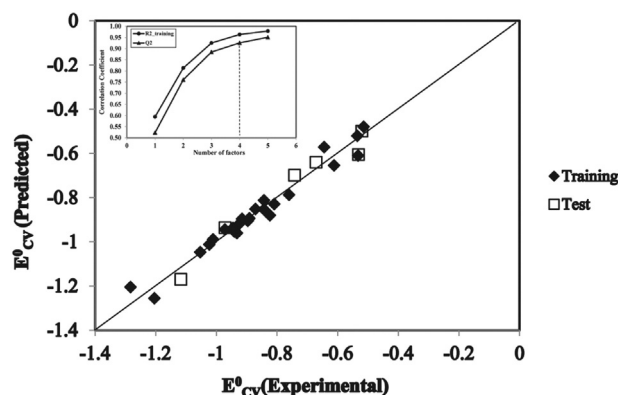$$N = 30, N_{train} = 24, N_{test} = 6, R^2_{train} = 0.96, F = 124.92, F_{crit\,(4,19)} = 2.90.$$

Fig. 1. Plot of predicted E$^0_{CV}$ versus their experimental values for 30 electrochemical mixtures in training and test set using the four parametric model. Inset: Correlation coefficient of the training set (R$^2_{train}$) and cross-validation (Q^2) versus number of parameters to select the optimum number of parameters for model development.