Quantitative Structure - Micellization Relationship Study of Cationic Surfactants Using Ordinary Least Squares Regression

J.B. Ghasemi*
Chemistry Department, Faculty of Sciences, Khaje NasiredinToosi University of Technology, Tehran, Iran
S. Saaidpour
Chemistry Department, Faculty of Sciences, Razi University, Kermanshah, Iran

Abstract

Introduction: The quantitative structure-property relationship (QSPR) is a successful strategy for prediction of surfactant properties based on modeling between calculated descriptors from molecular structures of the surfactants and chemical or physical properties of the solution. There are a great number of molecular descriptors that have been used in such QSPR studies, which can be divided into six types, namely constitutional descriptors, topological descriptors, electrostatic descriptors, geometrical descriptors, quantum chemical descriptors and thermodynamic descriptors. There are some reports about the applications of QSPR approaches to predict the CMC of anionic, nonionic and Gemini surfactants.

Aim: In the present work, the *logCMC* of some tetra-alkylammonium and alkylpyridinium salts was mathematically related to the molecular structure properties.

Material and Methods: All critical micelle concentrations data of this investigation were obtained from a set of cationic surfactants. They are measured in water at 25 °C. The data set consists of 44 surfactants were divided into two groups with 29 tetra-alkyl ammonium and 15 alkylpyridinium salts. The 3D molecular structures generated by ChemDraw 2005 and optimized by AM1 rotuine of MOPAC. The molecular descriptors generated ChemSAR and Dragon ver 3.0

Results: OLS regression analysis provided useful equations that can be used to predict the *logCMC* of cationic surfactants in this study. Model (I) which was used to estimate the logarithm of CMC tetra-alkyl ammonium surfactants using four structural descriptors could be represented as:

$$\log CMC = -1.0097 - 0.1258L_C - 0.0123V_H + 0.0960A_{HG} + 0.0053R_{HCI}$$

$$n = 20, R^2 = 0.9860, s^2 = 0.0210, F = 135, \mathrm{mod}el(I).$$

The *logCMC* of alkylpyridinium surfactants with three descriptors can be effectively predicted using following Eq. for model (II).

$$\log CMC = 6.0291 - 0.2461L_C - 0.0011V_H + 0.0249R_{HCI}$$

$$\mathrm{mod}el(II), n = 10, R^2 = 0.9940, s^2 = 0.0098, F = 159, \mathrm{mod}el(II).$$

The simultaneous model, which was used to estimate *logCMC* all cationic surfactants using four molecular structure descriptors, could be represented as

*Corresponding author

$$\log CMC = -1.4055 - 0.1529L_C - 0.0101V_H + 0.1214A_{HG} + 0.0063R_{HCI}$$

$$n = 30, R^2 = 0.9820, s^2 = 0.0228, F = 173, \ final \bmod el.$$

where n is the number of compounds used for regression, R^2 is the squared correlation coefficient, s^2 is the standard error of the regression, and F is the Fisher ratio for the regression.

Conclusion: The results indicate that the CMC decreases as the hydrophobic character (L and V) increases and CMC increases as the hydrophilic character (A) of the surfactant increases.

***Keywords*:** Cationic surfactants, Critical micelle concentration, Molecular descriptors, QSMR**,** OLS, Prediction.

Abbreviations

CMC, critical micelle concentration; QSPR, Quantitative structure property relationship;AM1, Austin model 1, QSMR, Quantitative structure-micellization relationship; OLS, ordinary least squares regression; L_C, hydrophobic chain length; V_H, hydrophobic volume; A_{HP}, area of hydrophilic portion; R_{HCI}, radius of hydrated counter ions.

Introduction

Surfactants are organic molecules that have a chemical structure combining both a polar (amphiphobic) and a nonpolar (amphiphilic) group into a single molecule.The value of the critical micelle concentration (CMC) is an important parameter in a wide variety of industrial applications involving adsorption of surfactant molecules at interfaces, such as foams, froths, emulsions, suspensions, and surface coatings. Micellization is observed in surfactant solutions when concentration exceeds the CMC, whereas the physicochemical properties of the aqueous solution change abruptly. The CMC is affected by factors such as relative size of hydrophobic and hydrophilic parts in the molecule, the counter-ions, and presence of electrolytes, pH and temperature.[1] The CMC of surfactants has been determined using many methods e.g. dye solubilization,[2] polarographic method,[3]flow injection systems,[4]turbidimetry,[5]light scattering,[6] ^{13}C NMR techniques,[7]luminescence spectroscopy, uv/visible spectroscopy, electrical conductivity, [8] surface tension,[9] capillary electrophoresis [10] and etc.

Cationic surfactants are comprised of a long chain hydrocarbon with a positively charged polar head group. Most cationic surfactants are long chain amines and their salts, quaternary ammonium salts, alkylpyridinium and picolinium compounds, or polyoxyethylenated long chain amines. In most cases, the head group is either a quaternary ammonium or phosphonium ion-paired with a counter ion such as chloride, bromide, iodide, hydroxide, carboxylates and nitrate.[11] Most cationic surfactants have hydrophilic groups, which are based on a nitrogen atom carrying the positive charge, few exceptions are based on phosphor and sulfur atoms. Most solid surfaces are negatively charged and the surface properties can therefore be modified by using cationic surfactants, as they carry a positive charge. First, their positive charge allows them to adsorb on negatively charged substrates, as most solid surfaces are at neutral pH. This capacity confers to them an antistatic behavior and a softening action for fabric and hair rinsing and conditioners. The positive charge enables them to operate as floatation collectors, hydrophobating agents, corrosion inhibitors, as well

as solid particle dispersant. They are used as emulsifiers in emulsions and coatings in general, in inks, wood pulp dispersions, herbicides, germicides, textiles auxiliaries, toiletries, etc. On the other hand, many cationic surfactants are bactericides. They are used to clean and asepsis surgery hardware, to formulate heavy-duty disinfectants for domestic and hospital use, and to sterilize food bottle or containers, particularly in the dairy and beverage industries.[12, 13]

The quantitative structure-property relationship (QSPR) is a successful strategy for prediction of surfactant properties based on modeling between calculated descriptors from molecular structures of the surfactants and chemical or physical properties of the solution.[14-16] There are a great number of molecular descriptors that have been used in such QSPR studies, which can be divided into six types, namely constitutional descriptors, topological descriptors, electrostatic descriptors, geometrical descriptors, quantum chemical descriptors and thermodynamic descriptors.[17-19] There are some reports about the applications of QSPR approaches to predict the CMC of anionic,[20-22] nonionic [23-25] and Gemini surfactants.[26, 27] In our previous papers, we reported on the application of QSPR techniques in the development of a new, simplified approach to prediction of compounds properties.[28-32]

In the present work, the *logCMC* of some tetra-alkylammonium and alkylpyridinium salts was mathematically related to the molecular structure properties.Ordinary Least squares (OLS) method was applied in quantitative structure-micellization relationship (QSMR) for modeling the relationship between *logCMC* of 44 cationic surfactants and their structural descriptors in aqueous solution.

Materials and Experimental Methods

The strategy used in this study consisted of six fundamental stages:
(1) molecular structure input and generation of the files containing the chemical structures stored in a computer–readable format, (2) quantum mechanics geometry optimization with a semi–empirical (AM1) method, (3) structural descriptors computation, (4) structural descriptors selection, (5) structure–*logCMC* model generation with the ordinary least squares regression method, and (6) statistical analysis.

Data set

All critical micelle concentrations data of this investigation were obtained from a set of cationic surfactants from references. [33, 34] They are measured in water at 25 °C. The data set consists of 44 surfactants were divided into two groups with 29 tetra-alkyl ammonium for model (I) and 15 alkylpyridinium salts for the model (II) and a simultaneous model was developed for all 44 cationic surfactants. The data set was split randomly into a training set (a-I and a-II) and a prediction set (b-I and b-II). The data set are shown in Table 1**.**

Table 1-Experimental values of logCMC cationic surfactants for train set (a-I, a-II) and prediction set (b-I, b-II).

No.	Compound	LogCMC(Exp.)	No.	Compound	LogCMC(Exp.)
1	$(C_{12}H_{25})_2N^+(CH_3)_2Br^-$ (a-I)	-3.7545	23	$C_{14}H_{29}N^+(CH_2Ar)(Me)_2Cl^-$(b-I)	-2.6990
2	$C_{18}H_{37}N^+(Me)_3Br^-$ (a-I)	-3.5376	24	$C_{14}H_{29}N^+(CH_3)_3NO_3^-$ (b-I)	-2.5686
3	$C_{16}H_{33}N^+(CH_3)_3NO_3^-$ (a-I)	-3.0915	25	$C_{14}H_{29}N^+(CH_3)_3Br^-$ (b-I)	-2.4437
4	$C_{12}H_{25}N^+OCT(Me)_2Br^-$ (a-I)	-2.9586	26	$C_{12}H_{25}N^+Bu(Me)_2Br^-$ (b-I)	-2.1249
5	$C_{16}H_{33}N^+(CH_3)_3Cl^-$ (a-I)	-2.8539	27	$C_{12}H_{25}N^+(CH_3)_3Br^-$ (b-I)	-1.7959

Tabel 1 countinue

6	$(C_{10}H_{21})_2N^+(CH_3)_2Br^-$	(a-I)	-2.7328	28
	$C_{12}H_{25}NH_2{}^+CH_2CH_2OHCl^-$	(b-I)	-1.4559	
7	$C_{14}H_{29}N^+(nPr)_3Br^-$	(a-I)	-2.6778	29
	$C_{10}H_{21}N^+(CH_3)_3Cl^-$	(b-I)	-1.1675	
8	$C_{16}H_{33}N^+(CH_3)_3OH^-$	(a-I)	-2.6383	30
	$C_{18}H_{37}PY^+Cl$-	(a-II)	-3.6198	
9	$C_{12}H_{25}N^+Hex(Me)_2Br^-$	(a-I)	-2.5086	31
	$C_{16}H_{33}PY^+Cl^-$	(a-II)	-3.0458	
10	$C_{14}H_{29}N^+(Et)_3Br^-$	(a-I)	-2.5086	32
	$C_{14}H_{29}PY^+Br^-$	(a-II)	-2.5686	
11	$C_{14}H_{29}N^+(CH_3)_3OH^-$	(a-I)	-2.3468	33
	$C_{14}H_{29}PY^+Cl^-$	(a-II)	-2.4559	
12	$C_{14}H_{29}N^+(CH_3)_3Cl^-$	(a-I)	-2.3468	34
	$C_{13}H_{27}PY^+Br^-$	(a-II)	-2.2757	
13	$C_{12}H_{25}N^+(CH_2Ar)(Me)_2Cl^-$	(a-I)	-2.0555	35
	$C_{12}H_{25}PY^+MeI^-$	(a-II)	-2.2636	
14	$C_{12}H_{25}N^+Et(Me)_2Br^-$	(a-I)	-1.8539	36
	$C_{12}H_{25}PY^+Br^-$	(a-II)	-1.9431	
15	$C_{12}H_{25}N^+(CH_3)_3Cl^-$	(a-I)	-1.6990	37
	$C_{12}H_{25}PY^+Cl^-$	(a-II)	-1.7242	
16	$C_{12}H_{25}N^+(CH_2CH_2OH)_3Cl^-$	(a-I)	-1.6021	38
	$C_{11}H_{23}PY^+Br^-$	(a-II)	-1.7100	
17	$C_{10}H_{21}N^+(CH_2Ar)(Me)_2Cl^-$	(a-I)	-1.5376	39
	$C_{10}H_{21}PY^+Cl^-$	(a-II)	-1.0614	
18	$C_{12}H_{25}NH^+(CH_2CH_2OH)_2Cl^-$	(a-I)	-1.4437	40
	$C_{16}H_{33}PY^+Br^-$	(b-II)	-3.1938	
19	$C_{10}H_{21}N^+(CH_3)_3Br^-$	(a-I)	-1.1675	41
	$C_{15}H_{31}PY^+Br^-$	(b-II)	-2.8861	
20	$C_8H_{17}N^+(CH_3)_3Br^-$	(a-I)	-0.7447	42
	$C_{10}H_{21}CHMePY^+MeI^-$	(b-II)	-2.3799	
21	$C_{16}H_{33}N^+(CH_3)_3Br^-$	(b-I)	-3.0088	43
	$C_6H_{13}C$=-C-CPY^+MeI^-	(b-II)	-1.8861	
22	$C_{14}H_{29}N^+(nBu)_3Br^-$	(b-I)	-2.9208	44
	$C_{10}H_{21}PY^+Br^-$	(b-II)	-1.3565	

Computer Hardware and Software

All calculations were run on a Pentium IV personal computer with windows XP operating system. The ChemDraw Ultra version 9.0 (ChemOffice 2005, CambridgeSoft Corporation) software was used for drawing the molecular structures.[35] The optimizations of molecular structures were done by the MOPAC 7.0 (AM1 method) and structural descriptors were calculated by Molecular Modeling Pro (MMP) Version 6.0 (ChemSW, Inc.) softwares.[36, 37] OLS regression performed by XLSTAT-PLS 1.8 software [38] and other calculations have done in the MATLAB (version 7.0, Mathworks, Inc.) environment.

Molecular Modeling and Calculation of Descriptors

The derivation of theoretical molecular descriptors proceeds from the chemical structure of the compounds. In order to calculate the theoretical descriptors, molecular structures were sketched with ChemDraw Ultra version 9.0 and molecular structures were optimized using AM1 algorithm.[39] The computational chemistry software Chem3D Ultra version 9.0 using in built MOPAC with 0.1 gradient cutoff was used to build the molecules and perform the necessary geometry optimizations. Eleven theoretical descriptors were

calculated for each compound in the data sets (44 compounds) by Molecular Modeling Pro (MMP) Version 6.0 (ChemSW, Inc.) software. The calculated descriptors were hydrophilic-lipophilic balance (*HLB*), hydrophobic volume (V_H), area of slice through hydrophilic portion (A_{HP}), hydrophobic portion's longest chain length(L_C), *V/(A*L)*, radius of hydrated counter ions (R_{HCl}), number of ethylene oxide units, number of propylene oxide unites, %molecular weight that is ethylene oxide, %molecular weight that is propylene oxide and surface tension of 1% aqueous.

Selection of Descriptors

All descriptors with zero, constant and/or near constant values for all the molecules in the data set were eliminated. The correlation matrix was calculated for the descriptors, one of the two descriptors which has the pair wise correlation coefficient above 0.7 ($r > 0.7$) and it has a large correlation coefficient with the other descriptors was eliminated.

The stepwise regression method was used as the variable selection method to select the suitable descriptors among eleven theoretical descriptors generated by Molecular Modeling Pro Plus software. This method combines the forward and backward procedures. Stepwise model-building techniques for regression designs with a single dependent variable involve identifying an initial model, repeatedly altering the model from the previous step by adding (forward stepwise) or removing (back stepwise) a predictor variable and terminating the search when stepping does not further improve the model. The forward stepwise method employs a combination of the forward entry of independent variables and backward removal of insignificant variables. The best single predictor, which is the most significant variable, was used for the initial linear regression step. Next, descriptors were added one at a time, always adding the one that most improved the fit, until the fit was not significantly improved. Once all the significant variables were determined, the regression equation was constructed. The number of variables retained in the model is based on the levels of significance assumed for inclusion and exclusion of variables from the model. [40, 41]

By using these criteria, seven out of the initial eleven original descriptors were eliminated and the remaining descriptors were used to generate the models using the XLSTAT-PLS software package. The result shows that four calculated descriptors are the most feasible ones. The selected descriptors are hydrophobic portions longest chain length (L_C), hydrophobic volume (V_H), area of slice through hydrophilic portion (A_{HP}) and the radius of the hydrated counter ions (R_{HCl}).

Ordinary Least Squares Regression Modeling

Regression can be used to describe and interpret the relationship between the X-variables (predictors) and the y-response, and to predict the y-values of new samples from the values of the X-variables. The linear relationship between the descriptors and *logCMC* of the cationic surfactants was modeled using an ordinary least squares regression (OLS) technique. [42] In the case of a model with *p* explanatory variables, the OLS regression model writes

$$y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \varepsilon \qquad (1)$$

where y is the dependent variable, β_0 is the intercept of the model, β_j's, are regression coefficients, X_j corresponds to the jth explanatory variable of the model (j= 1 to p), and ε is the random error with expectation 0. An ideal model is one that has high squared correlation coefficient (R^2) and Fisher ratio (F) values, a low standard error (s^2), and the fewest independent variables.

Results and Discussion

A number of good models were obtained using the OLS technique. Among these models, three models showed higher R^2 and F values and lower s^2 compared with the others.

However, the training set used to develop models (I) and (II) consisted of 20 tetra-alkylammonium and 10 alkylpyridinium salts, respectively. The specifications for models are given in Eqs. 2, 3 and 4. Inspection of the models revealed the superiority of models (I) and (II) and simultaneous model, owing to better predictive power. It should be noted that simultaneous model was obtained using all tetra-alkylammonium and alkylpyridinium salts (30 molecules). On the other hand, the parameters appearing in these models were only for the geometric-type descriptors.The geometric descriptors were calculated using the optimized Cartesian coordinates and the van der Waals radius of each atom in the molecule. Four descriptors appeared in model (I) as geometric descriptors, i.e., hydrophobic portions longest chain length (L_C), hydrophobic volume (V_H), area of slice through hydrophilic portion (A_{HP}) and the size of the counter ions (R_{HCI}) of the molecules. Model (II) consisted of three descriptors in common with model (I), with the A_{HP} (area of slice through hydrophilic portion) absent from this model. The correlation matrix for logCMC and the calculated descriptors are shown in Table 2. Of the four descriptors, L_C and V_H have the higher negative values of correlation coefficients. The lowest correlation coefficients were for the A_{HP} and R_{HCI} values.

Table 2- Correlation matrix for the dependence of logCMC with the descriptors.

	LogCMC	L_C	V_H	R_{HCI}	A_{HP}
LogCMC	1.0000				
L_C	-0.8756	1.0000			
V_H	-0.8082	0.6087	1.0000		
R_{HCI}	0.1103	-0.0015	0.0000	1.0000	
A_{HP}	0.3036	-0.0255	-0.0685	0.0055	1.0000

OLS analysis

OLS regression analysis provided useful equations that can be used to predict the *logCMC* of cationic surfactants in this study. The Eq. (2) for model (I) which was used to estimate the logarithm of CMC tetra-alkyl ammonium surfactants using four structural descriptors could be represented as:

$$\log CMC = -1.0097 - 0.1258L_C - 0.0123V_H + 0.0960A_{HG} + 0.0053R_{HCI}$$
$$n = 20, R^2 = 0.9860, s^2 = 0.0210, F = 135, \mathrm{mod}\,el(I). \quad (2)$$

The *logCMC* of alkylpyridinium surfactants with three descriptors can be effectively predicted using Eq. (3) for model (II).

$$\log CMC = 6.0291 - 0.2461L_C - 0.0011V_H + 0.0249R_{HCI}$$
$$\mathrm{mod}\,el(II), n = 10, R^2 = 0.9940, s^2 = 0.0098, F = 159, \mathrm{mod}\,el(II). \quad (3)$$

The simultaneous model (Eq. (4)), which was used to estimate *logCMC* all cationic surfactants using four molecular structure descriptors, could be represented as

$$\log CMC = -1.4055 - 0.1529L_C - 0.0101V_H + 0.1214A_{HG} + 0.0063R_{HCI}$$
$$n = 30, R^2 = 0.9820, s^2 = 0.0228, F = 173, \textit{final}\ \mathrm{mod}\,el. \quad (4)$$

where n is the number of compounds used for regression, R^2 is the squared correlation coefficient, s^2 is the standard error of the regression, and F is the Fisher ratio for the regression.

The squared correlation coefficient, R^2, is a measure of the fit of the regression model. Correspondingly, it represents the part of the variation in the experimental data that is explained by the model; the higher the value of correlation coefficient, the better the model.

The correlation coefficient values closer to 1 represents the better fit of the model. The s^2 is the standard error measured by the error mean square, which expresses the variation of the residuals or the variation about the regression line. Thus, the standard error measures the model error. If the model is correct, it is an estimate of the error of the data variance (σ^2). The F-test reflects the ratio of the variance explained by the model and the variance due to the error in the model, and high values of the F-test indicate the model is statistically significant.

Positive values in the regression coefficients indicate that the descriptor contributes positively to the value of CMC, whereas negative values indicate that the greater the value of the descriptor the lower the value of CMC.

The addition of a fourth descriptor (A_{HP}) to the model (II) was not statistically justified, resulting in a very minor improvement of the correlation coefficient and standard error and a decrease of the *F*-test statistic, finally this descriptor was excluded from the equation (3). Since the variation in the head group of alkylpyridinium surfactants in this work is minimum. Final model consists of four descriptors in common with model (I) and three descriptors in common with model (II).

The *logCMC* is assumed highly dependent upon the hydrophobic chain length (r^2=0.8756) and hydrophobic volume (r^2= 0.8086). Figure 1 (a, b) show the excellent correlations between the experimental *logCMC* of the all cationic surfactants with the (a) hydrophobic portions longest chain length (L_C) and (b) hydrophobic volume (V_H). The nature of the selected descriptors is in agreement with what is known qualitatively about structural effects on CMC. Finding a good correlation with only four descriptors can be considered a success

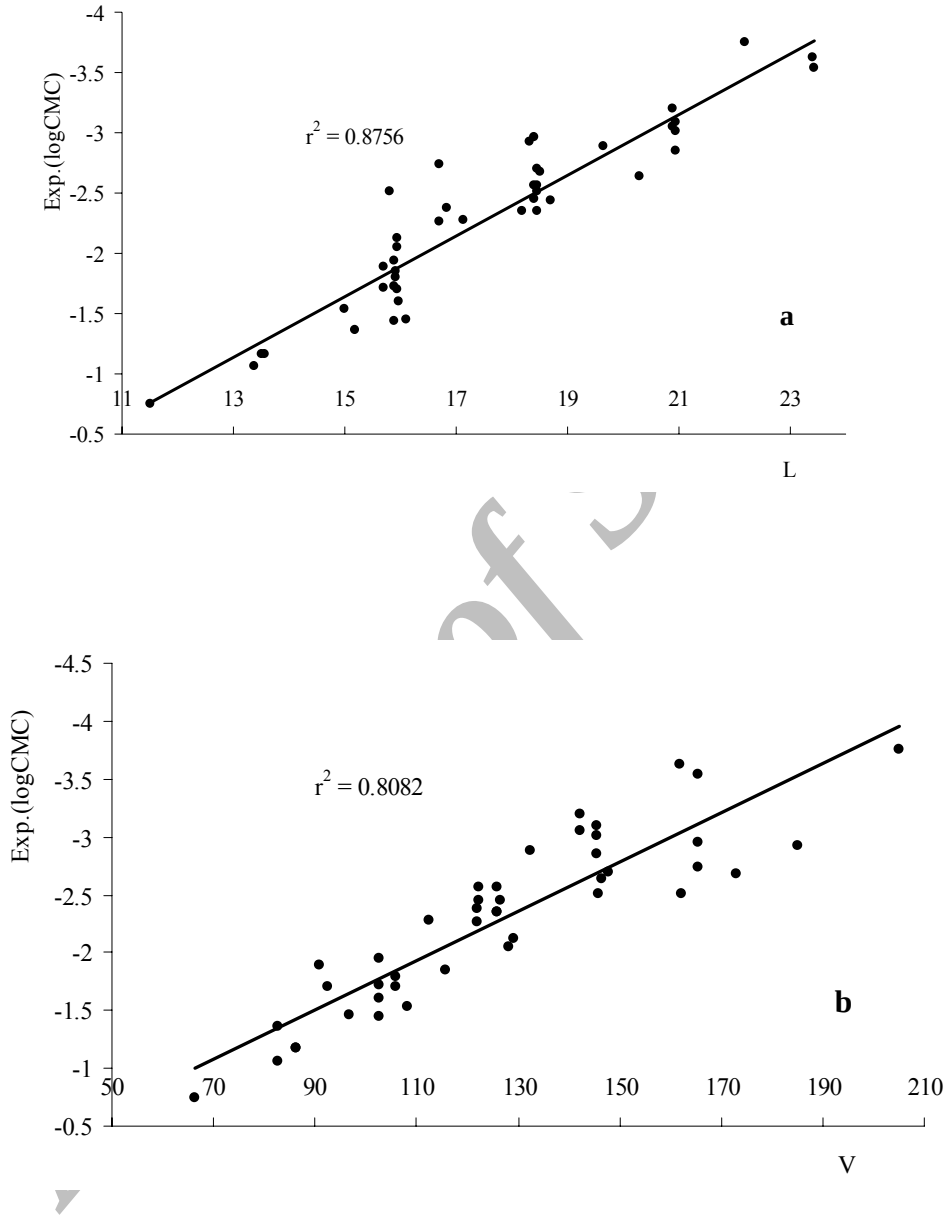

Fig 1- Theexperimental logCMC values cationic surfactants correlate well with the (a) hydrophobic chain length [L_C] and (b) hydrophobic volume [V_H].

QSMR models were building using a training set of 20 and 10 samples for model (I) and (II), respectively (Table 1(a-I),(a-II)). The stability and validity of models were tested by prediction of the response values for the prediction set. These models were applied for the prediction of the logCMC of some cationic surfactants, which were not used in the modeling procedure.Test sets containing of five alkylpyridinium and nine tetra-alkyl ammonium surfactants with regularly distributed *logCMC* values were used to assess the predictive ability of the produced QSMR models. The results of prediction set are shown in Table 3. Clearly, the calculated values of CMC are consistent with the experimental values. The scatter plots of predicted *logCMC* vs. experimental *logCMC* with high square correlation coefficients and the residuals versus experimental *logCMC* for above models are presented in figures 2, 3 and 4. Random distribution of residuals about zero mean indicated that no systematic error exists in the development of the OLS models.The average relative errors ($\overline{RE}\%$) of prediction are 0.4127%, 0.3234% and 2.1124% for model (I), model (II) and simultaneous model, respectively. However, the appearance of fewer descriptors (four parameters) in simultaneous model and its statistics illustrates that this equation (4) can be selected as a good model for predicting the logCMC of all tetra-alkylammonium and alkylpyridinium salts.

Table 3- Comparison between experimental and calculated logCMC values for external prediction set.

			LogCMC(Exp.)		
	LogCMC(Pred.)				
No.	Compound			Model (I)	Model
(II)	simultaneous Model				
1	$C_{16}H_{33}N^{+}(CH_3)_3Br^{-}$	(b-I)	-3.0088		-2.9667
	-3.0723				
2	$C_{14}H_{29}N^{+}(nBu)_3Br^{-}$	(b-I)	-2.9208		-3.0188
	-2.9933				
3	$C_{14}H_{29}N^{+}(CH_2Ar)(Me)_2Cl^{-}$	(b-I)	-2.6990		-2.5599
	-2.5832				
4	$C_{14}H_{29}N^{+}(CH_3)_3NO_3^{-}$	(b-I)	-2.5686		-2.4828
	-2.5746				
5	$C_{14}H_{29}N^{+}(CH_3)_3Br^{-}$	(b-I)	-2.4437		-2.4409
	-2.5206				
6	$C_{12}H_{25}N^{+}Bu(Me)_2Br^{-}$	(b-I)	-2.1249		-2.0757
	-2.0809				
7	$C_{12}H_{25}N^{+}(CH_3)_3Br^{-}$	(b-I)	-1.7959		-1.8214
	-1.8639				
8	$C_{12}H_{25}NH_2^{+}CH_2CH_2OHCl^{-}$	(b-I)	-1.4559		-1.4347
	-1.4533				
9	$C_{10}H_{21}N^{+}(CH_3)_3Cl^{-}$	(b-I)	-1.1675		-1.1986
	-1.1984				
10	$C_{16}H_{33}PY^{+}Br^{-}$	(b-II)	-3.1938	-3.1559	-
	3.0221				
11	$C_{15}H_{31}PY^{+}Br^{-}$	(b-II)	-2.8861	-2.8270	-
	2.7202				

Tabel 3 countinue

12 $C_{10}H_{21}CHMePY^{+}MeI^{-}$	(b-II)	-2.3799	-2.3946	-2.3307
13 $C_6H_{13}C\equiv C\text{-}CPY^{+}MeI^{-}$	(b-II)	-1.8861	-1.8469	-1.8535
14 $C_{10}H_{21}PY^{+}Br^{-}$	(b-II)	-1.3565	-1.5000	-1.5192

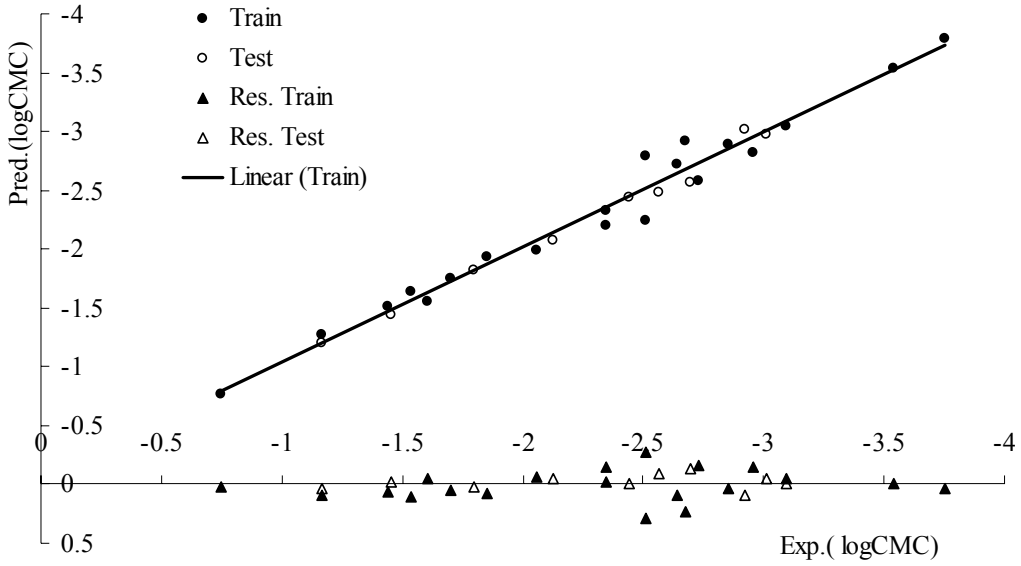

Fig 2- Plots of predicted logCMC and residuals logCMC versus experimental logCMC using model (I) for tetra-alkyl ammonium surfactants

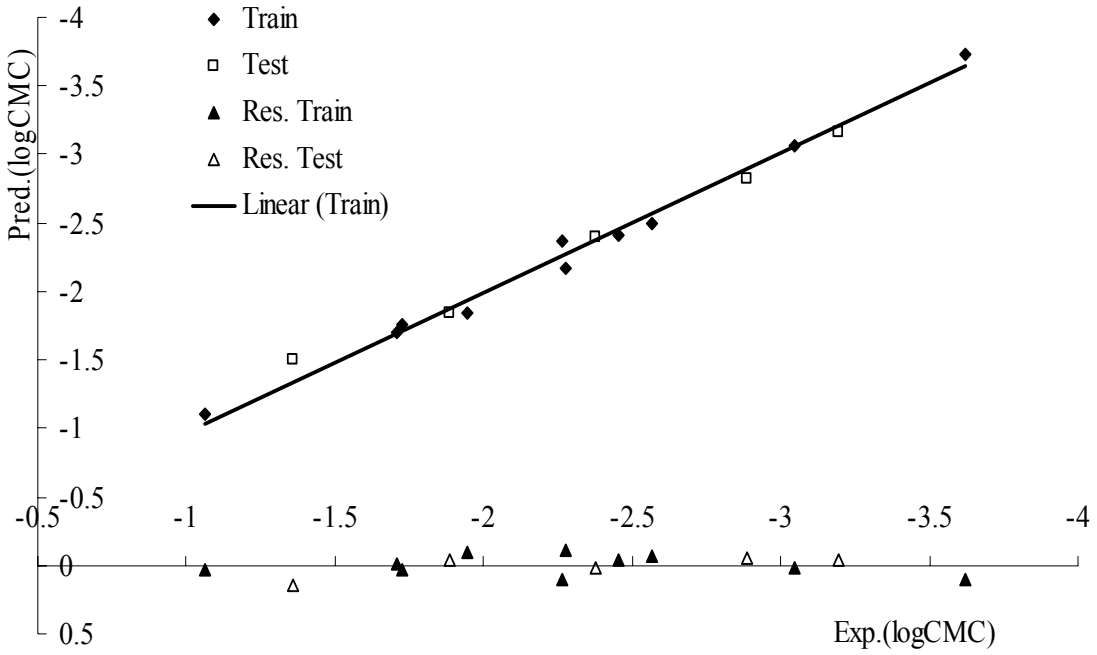

Fig 3- Plots of predicted logCMC and residuals logCMC versus experimental logCMC using model (II) for alkylpyridinium surfactants

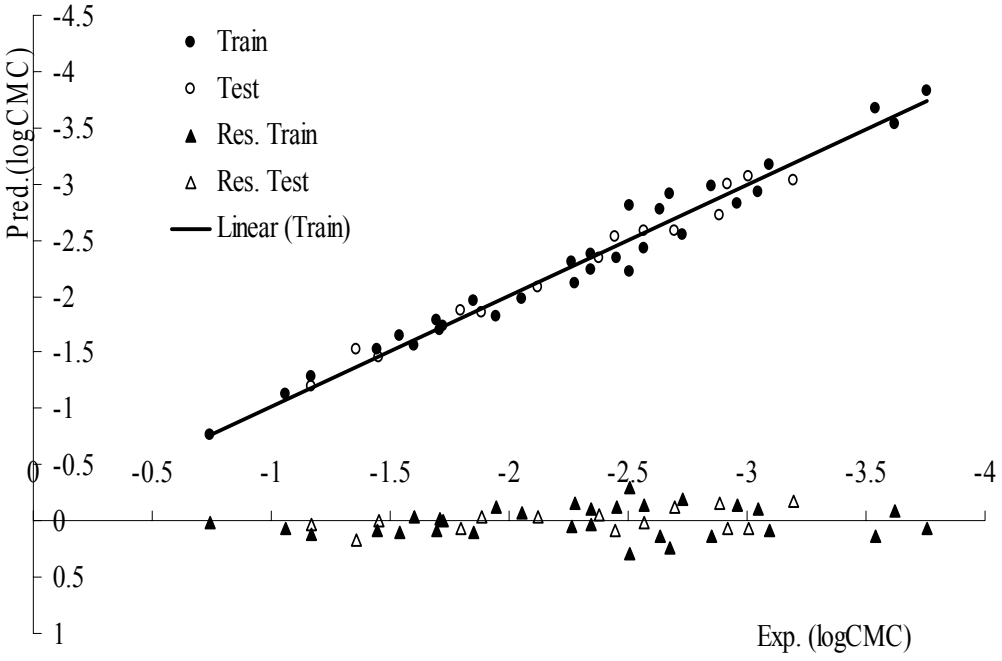

Fig 4- Plots of predicted logCMC and residuals logCMC versus experimental logCMC using simultaneous model for test set of cationic surfactants

Root mean square errors of prediction (RMSEP) values were then calculated for the prediction sets, and the results of RMSEP are 0.0688, 0.0738 and 0.0938 for model (I), model (II) and simultaneous model, respectively.

$$RMSEP = \left[\frac{1}{n} \sum_{i=1}^{n} (y_{\exp} - \hat{y}_{pred})^2 \right]^{0.5} \qquad (5)$$

Interpretation of descriptors

The QSMR models developed indicated that hydrophobic portions longest chain length (L_C), hydrophobic volume (V_H), area of slice through hydrophilic portion (A_{HP}) and the size of the hydrated counter ions (R_{HCl}) of the molecules significantly influence cationic surfactants micellization. Micelle formation of a surfactant in solution is induced by the hydrophobic interaction between hydrocarbon parts of the surfactant molecules balanced by their hydration and electrostatic repulsive effects. In aqueous media, the CMC of ionic surfactants is much higher than nonionic surfactants containing a similar hydrophobic tail. In general, the CMC decreases as the hydrophobic character of the surfactant increases, i.e. as the number of carbon atoms in the hydrophobic tail increases. A common rule is that for ionic surfactants the CMC is halved by the addition of one methylene group to a straight-chain hydrophobic group attached to a single terminal hydrophilic group. The presence of polar groups in the hydrophobic tail causes a significant increase in the CMC. The replacement of a hydrogen group by an ionic group also significantly increases the CMC. As seen in Eqs. 2, 3 and 4, hydrophobic portions longest chain length and hydrophobic volume descriptors appearing in models had negative coefficients. This means that as these parameters increase, the CMC of the cationic surfactants decrease. This is in agreement with the experimental values showing that the CMC of cationic surfactants decreases as the hydrophobicity of the molecule increases. Comparison of the mean effects of the descriptors appearing in models shows that the hydrophobic portions longest chain length had the largest effect on the micellization of cationic surfactants. These models show that as the L_C and V_H increases, the *logCMC* decreases. Inspection of models also indicates that for all molecules the lengths of the tails of the molecules plays an important role in micellization. This is also in agreement with the experimental values showing that, for all types of surfactants, as the length of a side chain increases, the CMC decreases. Increase in the length and volume of the hydrophobic group decreases the solubility of the surfactant in water and increases its solubility in organic solvents.

From geometric considerations, the aggregation numbers of micelles in aqueous media should increase rapidly with increase in the length of the hydrophobic group (L_C) of the surfactant molecule and decrease with increase in the area of the hydrophilic group (A_{HP}). Surfactants with bulky hydrophilic groups have larger CMC values. When the hydrophilic group is moved from a terminal position to a more central position, the CMC increases. It has been found that the CMC is higher when the charge on an ionic hydrophilic group is closer to the α-carbon atom of the (alkyl) hydrophobic group. This is explained as being due to an increase in electrostatic self-potential of the surfactant ion when the ionic head group moves from the bulk water to the vicinity of the nonpolar micellar core during the process of micellization; work is required to move an electric charge closer to a medium of lower dielectric constant. The area of slice through hydrophilic portion shows a positive coefficient in model (I) and simultaneous model. This means that as this parameter increases, the CMC of the cationic surfactant increases.

The effect of electrolyte on the CMC in aqueous media is very pronounced for ionic surfactants, and less so for zwitterionics and nonionics.The reduction in CMC is due mainly to the decrease in electrical repulsion between the ionic head-groups. The CMC in an aqueous solution is influenced by the degree of binding of counter ion to the micelle. For aqueous

systems, the increased binding of the counter ions to surfactant causes a decrease in the CMC and an increase in the aggregation number. The extent of binding of the counter ion increases with an increase in the polarizability and valence of counter ions and decrease with an increase in its hydrated radius. [43, 44]

The *logCMC* increases with the increase the hydrated radius of counter ions (decrease in the binding of the counter ions to the micelle) and the increase in the size of the hydrophilic group in cationic surfactants. The addition of neutral electrolyte to solutions of ionic surfactants in aqueous solution causes an increase in the aggregation number, presumably because of compression of the electrical double layer surrounding the ionic heads. The resulting reduction of their mutual repulsion in the micelle permits closer packing of the head groups (A_{HP} is reduced), with a consequent increase in aggregation number. For similar cationic surfactants with different counter ions, the CMC increases with increases the size of the hydrated counter ions. Increased binding of the less hydrated counter ion, in aqueous systems, causes a decrease in the CMC of the surfactant. The extent of binding of the counter ion decreases with increase in its hydrated radius. The larger the hydrated radius of the counter ion, the weaker the degree of binding on surface micelle. In a number of series of cationic surfactants, has shown that the degree of binding is related to the surface area per head group, in the ionic micelle, with the degree of binding increasing as the surface area per head group decreases. The CMC in aqueous solution reflects the degree of binding of the counter ion to the micelle.The increase in binding strength of the counter ion on surface of micelle in aqueous systems causes a decrease in the CMC of the surfactant. On the other hand, when comparing surfactants of different structural types, the value of the CMC does not always decrease with increase in the degree of binding of the counter ion. The decrease in the CMC is due mainly to the increased hydrophobicity of the surfactant. Within a group of ions of similar charge, those ions with a smaller hydrated radius or those that are more polarizable bind more strongly. Again, ions of higher charge and smaller hydrated radius bind more strongly than ions with a lower charge and a larger hydrated radius.

In this respect, inspection of equation 4, especially for all molecules, reveals the agreement between the results obtained and the experimental values. Therefore, the geometric descriptors and the types of counter ions in the molecules play a major role in micelle formation of the cationic surfactants salts. The delicate balance of opposing forces forms the micelles: the attractive tail–tail hydrophobic interaction provides the driving force for the aggregation of the surfactant molecules, while the electrostatic repulsion between the cationic head groups limits the size that a micelle can attain. The properties of ionic surfactants depend on the nature of counter ions, as their association at the micellar surface compensates electrostatic repulsion between the charged heads, thus stabilizing the micellar phase.

From the models presented in this work, one may conclude that the development of linear equations between log CMC and molecular descriptors might be of some help in designing new surfactants, with emphasis on geometric parameters for both the tetra-alkylammonium and alkylpyridinium salts. The results showed that the predicted values from this technique are close to the true values. High correlation coefficients for OLS, and low prediction errors demonstrated the ability of the selected structural descriptors for quantitative characterization of CMC of cationic surfactants. As a result, the small changes in chemical structure of the surfactant molecule easily control the characteristics of these aggregates. Quantitative studies offer two major advantages. One of them is that it provides a way for estimating the property for novel compounds with a fast and an acceptable degree of precision. The other is that it makes structural interpretations of the physicochemical property possible.

Conclusions

OLS modeling method was used to QSMR study of the CMC of cationic surfactants using structural descriptors. The results showed that the predicted values from this technique are close to the true values. High correlation coefficients for OLS, and low prediction errors demonstrated the ability of the selected structural descriptors for quantitative characterization of CMC of cationic surfactants. Qualitatively it is well known that contributions from both the size of the hydrophobic domain (tail) (L and V), the size of the hydrophilic domain (head) (A) and the counter ion nature (charge and size)(CI) of the surfactant influence the CMC values. The two contributions are contrary, with a lower CMC for a larger length and volume of hydrophobic domain and a higher CMC for a bigger surface area of hydrophilic fragment (size of hydrophilic fragment). The results indicate that the CMC decreases as the hydrophobic character (L and V) increases and CMC increases as the hydrophilic character (A) of the surfactant increases. The delicate balance of opposing forces forms the micelles: the attractive tail–tail hydrophobic interaction provides the driving force for the aggregation of the surfactant molecules, while the electrostatic repulsion between the cationic head groups limits the size that a micelle can attain. The properties of ionic surfactants depend on the nature of counter ions, as their association at the micellar surface compensates electrostatic repulsion between the charged heads, thus stabilizing the micellar phase. In this study with increasing the size of counter ion, as less extent CMC decreases. As a result, the small changes in chemical structure of the surfactant molecule easily control the characteristics of these aggregates.

References

1. Holmberg, K., Shah, D.O., Schwuger, M.J., *Handbook of applied surface and colloid chemistry*, John Wiley & Sons, New York (2002).
2. Mukerjee, P., *Solution chemistry of surfactants*. Plenum Press, New York (1979) .
3. Vollhardt, D., The *A. C. polarographic method for the simple direct determination of the critical micelle concentration.Tenside Deterg*, **12**, 255 (1975) .
4. Brooks, S.H., Berthod, A., Kirsch, B., Dorsey, J.G., *Anal Chim Acta* , **209**,111 (1988).
5. Chang, C.Y., Wang, S.J., Liu, I.J., Chiu, Y.C., *J. Chin. Chem. Soc., (Taipei)*, **34**, 243 (1987) .
6. Nishikido, N., Shinozaki, M., Sugihara, G., Tanaka, M., Kaneshina, S.A., *J. Colloid Interface Sci*., 74, 474.(1980) .
7. Corno, C., Platone, E., Ghelli, S., *J. Colloid polym Sci*., **262**, 667 (1984).
8. Dominguez, Am, Fernandez, Am, Gonzalezm N., Iglesias, E., Montenegro, L., *J. Chem. Educ.*, 74, 1227 (1997).
9. Rakesh, S., Dharmesh, V., Pratap, B., *J. Dispersion Sci Technol*, **24**, 53 (2003)
10. Lin, C.E., *J. Chromatogr A*, **1037**, 467 (2004).
11. Schwartz, A.M., Perry, J.W., *Surface active agents: their chemistry and technology*/ Robert E. Krieger Publishing Company, New York (1978).
12. Salager, J.L., *Surfactants Types and Uses.,* FiRP, Venezuela (2002)
13. Holmberg, K., Shah, D.O., Schwuger, M.J., *Handbook of applied surface and colloid chemistry*, John Wiley & Sons, New York (2002).
14. Oros, G., Ciserhati, T., Fpga, E., *Chemom Intell Lab Syst*., **47**, 149 (1999).
15. Katritzky, A.R., Maran, U., Lobanov, V.S., Karelson, M., *J. Chem Inf Comp Sci.,* **40**, 1 (2000).
16. Heravi, M.J., Konouz, E., *Internet Electron J. Mol. Des*., **1**, 410 (2002).
17. Young, D.C., *Computational chemistry: a practical guide for applying techniques to real-world problems*, Wiley, New York (2001)
18. Karelson, M., Lobanov, V.S., Katritzky, A.R., *Quantum-chemical descriptors in QSAR/QSPR studies. Chem Rev*., **96**, 1027 (1996) .
19. Todeschini, R., Consonni, V., *Handbook of molecular descriptors*, Wiley-VCH, Germany (2000).
20. Li, X., Zhang, G., Dong, J., Zhou, X., Yan, X., Luo, M., *J. Mol. Struct. (Theochem)*, **710**,119 (2004) .
21. Huibers, P.D.T., Lobanov, V.S., Katritzky, A.R., Shah, D.O., Karelson, M., *J. Colloid Interface Sci.*, 187, 113(1997) .
22. Jalali-Heravi, M., Konouz, E., *J. Surfact Deterg*, **3**, 47 (2000) .
23. Cheng, J.S., Chen, Y.P., *Fluid Phase Equilib*, **232**, 37 (2005) .
24. Flores, M.V., Voutsas, E.C., Spiliotis, N., Eccleston, G.M., Bell, G., Tassios DP, Halling, P.J., *J. Colloid Interface Sci.,* **240**, 277 (2001).
25. Huibers, P.D.T., Lobanov, V.S., Katritzky, A.R., Shah, D.O., Karelson, M., *1. Nonionic surfactants. Langmuir*, **12**, 1462 (1996).
26. Kardanpour, Z., Hemmateenejad, B., Khayamian, T., *Anal. Chim. Acta*., 531, 285 (2005).
27. Absalan, G., Hemmateenejad, B., Soleimani, M., Akhond, M., Miri, R., *QSAR Comb Sci.*, **23**: 416(2004).
28. Ghasemi, J., Saaidpour, S., Brown, S.D., *J. Mol. Struct. (Theochem),* **805**, 27 (2007).
29. Ghasemi, J., Saaidpour, S., *Chem Pharm Bull*, **55**,669 (2007).
30. Ghasemi, J., Shahmirani, S., Farahani, E,V., *Anal. Chim*., **96**, 327(2006).

31. Ghasemi, J., Ahmadi, Sh., *Anal, Chim.*, **97**, 69(2007) .
32. Ghasemi, J., Asadpour, S., Abdolmaleki, A., *Anal Chim Acta*., **588**, 200(2007).
33. Vanos N M, Haak J R, Rupert LAM (1993) Physico- Chemical properties of selected anionic, cationic and nonionic surfactants. Elsevier, Amsterdam
34. Rosen, M.J., *Surfactants and Interfacial Phenomena*, John Wiley & Sons, New York (2004).
35. ChemOffice, *CambridgeSoft Corporation*, http://www.cambridgesoft.com/(2005).
36. Web: http://www.mysiteinc.com/winmopac/
37. Web: http://www.chemsw.com/
38. Web: http://www.xlstat.com/
39. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., Stewart, J.J.P., *J. Am. Chem. Soc*., **107**, 3902(1985).
40. Darlington, R.B., *Regression and Linear Models*, McGraw-Hill, New York (1990).
41. Massart, D.L.; Vandeginste, B.G.M.; Buydens, L.M.C.; Jong, S.D.; Lewi, P.J.; Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics: Part A*, Elsevier, Amsterdam (1997.)
42. Dempster, A.P., *Elements of Continuous Multivariate Analysis*, Addison-Wesley Publishing Company, Massachusetts (1969).
43. Mukerjee, P., *Adv Colloid Interface Sci.*, **1**, 242(1967).
44. Shinoda, K., *Colloidal Surfactants: Some Physicochemical Properties*, Academic Press, NewYork (1963).