Prediction on Critical Micelle Concentration of Nonionic Surfactants in Aqueous Solution Quantitative Structure-Property Relationship Approach

WANG Zheng-Wu[a, c]   HUANG Dong-Yang[*, a]   GONG Su-Ping[b]
LI Gan-Zuo[c]

[a] *Shantou University Medical College Shantou Guangdong 515041 China*
[b] *Shangdong Electric Power Central Hospital Jinan Shandong 250001 China*
[c] *Key Lab of Colloid and Interface Chemistry for State Education Ministry Shandong University Jinan Shandong 250001 China*

In order to predict the critical micelle concentration cmc of nonionic surfactants in aqueous solution a quantitative structure-property relationship QSPR was found for 77 nonionic surfactants belonging to eight series. The best-regressed model contained four quantum-chemical descriptors the heat of formation ΔH the molecular dipole moment D the energy of the lowest unoccupied molecular orbital E_{LUMO} and the energy of the highest occupied molecular orbital E_{HOMO} of the surfactant molecule two constitutional descriptors the molecular weight of surfactant M and the number of oxygen and nitrogen atoms n_{ON} of the hydrophilic fragment of surfactant molecule and one topological descriptor the Kier & Hall index of zero order $KH0$ of the hydrophobic fragment of the surfactant. The established general QSPR between lg cmc and the descriptors produced a relevant coefficient of multiple determination $R^2 = 0.986$. When cross terms were considered the corresponding best model contained five descriptors E_{LUMO} D $KH0$ M and a cross term n_{ON} $KH0$ which also produced the same coefficient as the seven-parameter model.

Keywords   quantitative structure-property relationship critical micelle concentration nonionic surfactant

Introduction

Critical micelle concentration cmc of surfactants in aqueous solution is one of the most useful parameters for characterizing the properties of surfactants. Over a very narrow concentration range around the cmc transitions of the existence of surfactants occur from monomer premicellar to micellar. And companying these transitions many other important properties of surfactant solution such as surface tension interfacial tension conductivity osmotic pressure detergency emulsification foaming and so on also change sharply at the point.[1,2] The value of cmc of a surfactant is firmly related to its structure i.e. the con-

tributions from both the characteristics of the hydrophobic domain tail and the hydrophilic domain head of a surfactant.

Nonionic surfactant is one of mostly used surfactants and some empirical relationships between the structural feature of surfactants and cmc have been developed for some commonly used surfactants based on experimental data. But so far they are mainly limited to the homologous series of the linear alkyl ethoxylates C_mE_n .[3-6] Though a series of models have also been presented for diverse nonionic surfactants very little literature has been published on the relationship between the structure of surfactant molecule and the observed cmc from the motion state as well as the motion level of electrons with the quantum chemistry method.[7-9] Therefore it is important to employ the quantitative structure-property relationship QSPR techniques to expand our understanding in this area.

In our previous papers we have successfully used QSPR method to predict respectively the critical micelle concentration[10] and the corresponding surface tension[11] at this concentration of anionic surfactants in aqueous solution. For nonionic surfactants we have established the relationship between molecule structure and observed cmc for two series [12] i.e. the linear alkyl ethoxylates and the alkyl phenyl polyethoxylates at a given temperature. Based on molecular topology and quantum chemistry a multiple regression model Eq. 1 between lg cmc and three descriptors was founded.

$$\text{lg cmc} = 1.930 - 0.7846KH0 - 8.871 \times 10^{-5}E_T + 0.04938D \quad R^2 = 0.995 \qquad 1$$

where E_T D $KH0$ and R^2 are the total molecular energy the molecular dipole moment and the Kier & Hall

molecular connectivity index of zero order for the hydrophobic fragment and the squared correlation coefficient or coefficient of multiple determination respectively. For only the linear alkyl ethoxylates had a multiple regression model with the squared correlation coefficient being 0.997 which is more even superior to Becher's empirical model $R_2 = 0.996$.[7, 13]

The present work is based on the observed cmc for much more divers of molecular structures using much more molecular descriptors such as quantum-chemical descriptors ΔH E_T D E_{LUMO} and E_{HOMO} of surfactant constitutional descriptors M and n_{ON} of the hydrophilic fragment in surfactant molecule and topological descriptor $KH0$ of the hydrophobic fragment in surfactant molecule to establish the QSPR between cmc and these descriptors for nonionic surfactants.

Mathematical foundation of QSPR

The mathematical foundation of QSPR is based on the principle of polylinearity.[11] According to the theory a continuous and singular dependence on a property P_i which is experimentally measurable and some intrinsic structural factors of molecule x_j is assumed to be linear in a certain domain of these factors x_j . Under this assumption the experimental property P_i may depend linearly on one or more structure factors x_j and by using the multilinear least squares method the corresponding linear multi-parameter regression equation can be found as

$$P_i = P_0 + \sum_{j=1}^{n} a_j x_{ij} \qquad 2$$

In some cases nonlinear functions for x_j have to be used for the description of the property such as square x_j^2 exponential function b^{x_j} logarithm function x_j. Even in some cases P_i is not only simply dependent on single descriptors but also on the combination of a few descriptors because of the existence of interaction among them such as the interaction between x_j and x_k with a cross terms $x_j x_k$ and so on. The principle of polylinearity is still valid by introducing the following nonlinear transformations Eq. 3 into Eq. 2 .

$$x_j = x_j^2 \quad b^{x_j} \quad \ln x_j \quad x_j x_k \quad \dots \qquad 3$$

Data and methodology

Data

The chosen data set of cmc has contained 77 nonionic surfactants at 25.0 ℃ in aqueous solution with no extra salt which belongs to eight classes linear dodecyl polyoxyethylene polyoxypropylene ethers linear alkyl ethoxylates octylphenol ethoxylates branched alkyl ethoxylates

alkanediols alkyl-mono and disaccharide ethers and esters ethoxylated alkyl amines and amides fluorinated linear ethoxylates and amides *etc*. Table 1 in which the values for linear dodecyl polyoxyethylene polyoxypropylene ethers $C_{12}H_{25}$ OC_2H_4 $_m$ OC_3H_6 $_n$OH $m = 3$ $n = 6$ $m = 4$ $n = 5$ $m = 5$ $n = 4$ were obtained from our unpublished data. These materials were supplied by Sigma Chemical Co. USA purity > 99% and their cmc data were obtained from the sharp break points in γ-lg$c$ molar concentration curves in aqueous solution with the Du-Nouy ring method in our laboratory. The other data were chosen from Rosen's textbook.[1] Since most nonionic surfactants consisting of the ethylene oxide oligomers in the hydrophilic domain of the molecule often contain a distribution of the polyethylene oxide chain lengths rather than a constant number of units only the monomerically pure surfactants were chosen for the present study.

Computation of descriptors

The semi-empirical molecular orbital method in quantum chemistry was used for the computation of the quantitative chemical descriptors of the surfactants ΔH_f E_T E_{LUMO} E_{HOMO} and D. The parameters for computation were obtained from Ref. 13. Before the calculation the molecular structures were optimized with the MINDO-PM3 method. In the real solution surfactants may curl form non-regular clew or wrap together among them. But in the optimization of the molecular structures these phenomena are neglected. All computations were carried out on a PIII 850 PC computer with the MNDO-MOPAC7.0 software.

The topological descriptor $KH0$ which represents the size of the hydrophobic segment and contains group contributions from all nonhydrogen atoms in the fragment is defined as[6, 14]

$$KH0 = \sum_{j=1}^{N} \delta_i^\nu {}^{-1/2} \quad \text{where} \quad \delta_i^\nu = \frac{Z_i^\nu - H_i}{Z_i - Z_i^\nu - 1} \qquad 4$$

Here Z_i is the total number of electrons in the ith atom Z_i^ν is the number of valance electrons and H_i is the number of hydrogen directly attached to the ith atom. Valence contributions are summed for all atoms in the fragment with the exception of the hydrogen atoms $N = N_{total} - N_H$.

The number of oxygen and nitrogen atoms n_{ON} which captures the size of the hydrophilic fragment and accounts for the influence of the fragment on cmc due to the formation of hydrogen bounds between surfactant and solvent molecules is directly obtained by accounting the number of oxygen and nitrogen atoms in the fragment. The divisions of hydrophobic and hydrophilic segments of the nonionic surfactants the observed cmc and the values of some of descriptors are listed in Table 1.

Table 1 Values of descriptors the observed and the calculated logarithm of cmc for 77 nonionic surfactants 25 ℃

Structure Hydrophobic-hydrophilic segment	n_{NO}	$KH0$	ΔH_f kJ	E_{LUMO} eV	E_{HOMO} eV	D Debye	M	lg cmc ob.	lg cmc cal.
Linear dodecyl polyoxyethylene polyoxypropylene ether									
$C_{12}H_{25}$ OC_2H_4 ₃ OC_3H_6 ₆OH	10	8.778	− 2057	1.991	10.57	6.599	667.1	− 5.148	− 4.894
$C_{12}H_{25}$ OC_2H_4 ₄ OC_3H_6 ₅OH	10	8.778	− 2034	1.937	10.57	6.602	653.0	− 4.796	− 4.838
$C_{12}H_{25}$ OC_2H_4 ₅ OC_3H_6 ₄OH	10	8.778	− 2012	1.897	10.58	4.604	639.3	− 4.762	− 4.560
Linear alkyl ethoxylate									
C_4H_9 OC_2H_4 ₁₀H	2	3.121	− 445.4	2.544	10.54	0.360	118.2	− 0.009	− 0.018
C_4H_9 OC_2H_4 ₆OH	7	3.121	− 1254	1.891	10.58	1.410	338.4	− 0.110	0.126
C_6H_{13} OC_2H_4 ₃OH	4	4.534	− 814.4	2.078	10.57	0.376	234.3	− 1.000	− 1.015
C_6H_{13} OC_2H_4 ₆OH	7	4.534	− 1299	1.890	10.58	1.406	366.4	− 1.164	− 0.991
C_8H_{17} OC_2H_4 ₁₀H	2	5.951	− 536.3	2.524	10.55	0.371	174.2	− 2.310	− 2.217
C_8H_{17} OC_2H_4 ₃OH	4	5.951	− 859.8	2.077	10.57	0.379	262.3	− 2.125	− 2.027
C_8H_{17} OC_2H_4 ₆OH	7	5.951	− 1345	1.890	10.58	1.405	394.5	− 2.004	− 2.002
C_8H_{17} OC_2H_4 ₉OH	10	5.951	− 1830	1.829	10.58	0.376	526.7	− 1.886	− 1.722
$C_{10}H_{21}$ OC_2H_4 ₃OH	4	7.364	− 905.3	2.076	10.57	0.381	290.4	− 3.222	− 3.196
$C_{10}H_{21}$ OC_2H_4 ₄OH	5	7.364	− 1067	1.985	10.58	1.404	334.4	− 3.167	− 3.127
$C_{10}H_{21}$ OC_2H_4 ₆OH	7	7.364	− 1390	1.889	10.58	1.396	422.6	− 3.046	− 3.013
$C_{10}H_{21}$ OC_2H_4 ₈OH	9	7.364	− 1714	1.844	10.58	1.410	510.7	− 3.000	− 2.922
$C_{10}H_{21}$ OC_2H_4 ₉OH	10	7.364	− 1875	1.829	10.58	0.372	554.7	− 2.886	− 2.737
$C_{11}H_{23}$ OC_2H_4 ₈OH	9	8.071	− 1743	1.822	10.59	1.657	524.7	− 3.523	− 3.469
$C_{12}H_{25}$ OC_2H_4 ₂OH	3	8.778	− 789.2	2.232	10.56	1.403	274.4	− 4.481	− 4.276
$C_{12}H_{25}$ OC_2H_4 ₃OH	4	8.778	− 950.8	2.076	10.57	0.382	318.4	− 4.284	− 4.058
$C_{12}H_{25}$ OC_2H_4 ₄OH	5	8.778	− 1112	1.985	10.58	1.403	362.5	− 4.194	− 4.149
$C_{12}H_{25}$ OC_2H_4 ₅OH	6	8.778	− 1274	1.927	10.58	0.383	406.0	− 4.194	− 3.941
$C_{12}H_{25}$ OC_2H_4 ₆OH	7	8.778	− 1436	1.889	10.58	1.403	450.6	− 4.060	− 4.033
$C_{12}H_{25}$ OC_2H_4 ₇OH	8	8.778	− 1597	1.863	10.59	0.384	494.7	− 4.086	− 3.846
$C_{12}H_{25}$ OC_2H_4 ₈OH	9	8.778	− 1759	1.843	10.58	1.404	538.7	− 4.000	− 3.984
$C_{12}H_{25}$ OC_2H_4 ₉OH	10	8.778	− 1921	1.829	10.58	0.384	582.8	− 4.000	− 3.754
$C_{12}H_{25}$ OC_2H_4 ₁₂OH	13	8.778	− 2413	1.791	10.59	1.579	714.9	− 3.854	− 3.788
$C_{13}H_{27}$ OC_2H_4 ₈OH	9	9.485	− 1782	1.844	10.58	1.392	552.7	− 4.569	− 4.447
$C_{14}H_{29}$ OC_2H_4 ₆OH	7	10.19	− 1481	1.890	10.58	1.387	478.7	− 5.000	− 5.042
$C_{14}H_{29}$ OC_2H_4 ₈OH	9	10.19	− 1804	1.844	10.58	1.407	566.8	− 5.046	− 4.952
$C_{15}H_{31}$ OC_2H_4 ₈OH	9	10.90	− 1827	1.843	10.58	1.390	580.8	− 5.456	− 5.457
$C_{16}H_{33}$ OC_2H_4 ₆OH	7	11.61	− 1527	1.889	10.58	1.404	506.7	− 5.780	− 6.062
$C_{16}H_{33}$ OC_2H_4 ₇OH	8	11.61	− 1688	1.862	10.58	0.385	550.8	− 5.770	− 5.874
$C_{16}H_{33}$ OC_2H_4 ₉OH	10	11.61	− 2012	1.829	10.59	0.387	638.9	− 5.678	− 5.789
$C_{16}H_{33}$ OC_2H_4 ₁₂OH	13	11.61	− 2504	1.792	10.59	1.667	771.1	− 5.638	− 5.833
Octylphenol ethoxylate									
$C_8H_{17}C_6H_4$ OC_2H_4 ₁₀H	2	8.425	− 430.2	0.2018	9.255	0.179	250.3	− 4.305	− 3.910
$C_8H_{17}C_6H_4$ OC_2H_4 ₂OH	3	8.425	− 591.9	0.1765	9.288	1.607	294.4	− 4.116	− 4.077
$C_8H_{17}C_6H_4$ OC_2H_4 ₃OH	4	8.425	− 753.6	0.1715	9.296	0.186	338.4	− 4.013	− 3.845
$C_8H_{17}C_6H_4$ OC_2H_4 ₄OH	5	8.425	− 915.3	0.1679	9.302	1.627	382.5	− 3.886	− 4.002
$C_8H_{17}C_6H_4$ OC_2H_4 ₅OH	6	8.425	− 1077	0.1663	9.303	0.198	426.5	− 3.824	− 3.765
$C_8H_{17}C_6H_4$ OC_2H_4 ₆OH	7	8.425	− 1234	0.1918	9.229	1.637	470.6	− 3.678	− 3.893
$C_8H_{17}C_6H_4$ OC_2H_4 ₇OH	8	8.425	− 1396	0.1927	9.232	0.230	514.6	− 3.602	− 3.660
$C_8H_{17}C_6H_4$ OC_2H_4 ₈OH	9	8.425	− 1558	0.1918	9.234	1.628	558.7	− 3.553	− 3.813
$C_8H_{17}C_6H_4$ OC_2H_4 ₉OH	10	8.425	− 1720	0.1912	9.245	0.220	602.8	− 3.523	− 3.583
$C_8H_{17}C_6H_4$ OC_2H_4 ₁₀OH	11	8.425	− 1881	0.1906	9.236	1.629	646.8	− 3.481	− 3.732

Continued

Structure Hydrophobic-hydrophilic segment	n_{NO}	KH0	ΔH_f kJ	E_{LUMO} eV	E_{HOMO} eV	D Debye	M	lg cmc ob.	lg cmc cal.
Branched alkyl ethoxylate									
CH$_3$ $_2$CH$_2$CH$_2$ OC$_2$H$_4$ $_6$OH	7	3.285	− 1252	1.892	10.56	1.370	338.4	− 0.049	− 0.070
C$_2$H$_5$ $_2$CH$_2$CH$_2$ OC$_2$H$_4$ $_6$OH	7	4.699	− 1289	1.891	10.56	1.326	366.4	− 1.016	− 1.079
C$_3$H$_7$ $_2$CH$_2$CH$_2$ OC$_2$H$_4$ $_6$OH	7	6.113	− 1334	1.891	10.56	1.319	394.5	− 1.670	− 1.690
C$_4$H$_9$ $_2$CH$_2$CH$_2$ OC$_2$H$_4$ $_6$OH	7	7.527	− 1382	1.891	10.51	1.320	422.6	− 2.547	− 2.582
C$_4$H$_9$ $_2$CH$_2$CH$_2$ OC$_2$H$_4$ $_9$OH	10	7.527	− 1867	1.829	10.51	0.397	554.7	− 2.526	− 2.816
Alkanediol									
C$_8$H$_{17}$ OCH$_2$CH OH CH$_2$OH	3	5.950	− 724.9	2.498	10.51	2.340	204.3	− 2.237	− 2.464
C$_8$H$_{17}$ CH OH CH$_2$OH	2	5.950	− 585.8	2.840	10.97	1.845	174.2	− 2.638	− 2.701
C$_8$H$_{17}$ CH OH CH$_2$CH$_2$OH	2	5.950	− 605.1	2.755	11.03	1.955	188.3	− 2.638	− 2.765
C$_{10}$H$_{21}$ CH OH CH$_2$OH	2	7.364	− 631.3	2.838	10.97	1.846	202.3	− 3.745	− 3.711
C$_{12}$H$_{25}$ CH OH CH$_2$CH$_2$OH	2	8.778	− 696.1	2.753	11.04	1.956	244.4	− 4.886	− 4.801
Alkyl-mono and disaccharide ether and ester									
C$_8$H$_{17}$ O C$_6$H$_{11}$O$_5$	6	5.950	− 1261	1.883	10.93	2.113	292.3	− 1.602	− 1.760
C$_{10}$H$_{21}$ O C$_6$H$_{11}$O$_5$	6	7.364	− 1307	1.882	10.93	2.126	320.4	− 2.658	− 2.893
C$_{12}$H$_{25}$ O C$_6$H$_{11}$O$_5$	6	8.778	− 1352	1.882	10.93	2.123	348.4	− 3.721	− 3.707
C$_{12}$H$_{25}$ NH C$_6$H$_{12}$O$_4$ O C$_6$H$_{11}$O$_5$	11	8.778	− 2104	1.467	8.812	5.938	511.6	− 3.222	− 3.047
C$_{12}$H$_{25}$ O C$_6$H$_{10}$O$_4$ O C$_6$H$_{11}$O$_5$	11	8.778	− 2239	1.538	10.55	2.15	510.6	− 3.620	− 3.376
C$_{11}$H$_{23}$ COO C$_6$H$_{10}$O$_4$ O C$_6$H$_{11}$O$_5$	12	8.071	− 2398	0.8515	10.66	4.984	524.6	− 3.469	− 3.240
C$_8$H$_{17}$CHCH CH$_2$ $_7$ COO C$_6$H$_{10}$O$_4$ O-C$_6$H$_{11}$O$_5$	12	12.05	− 2423	0.8567	9.801	4.811	606.7	− 5.292	− 5.471
Ethoxylated alkyl amine and amide									
C$_{11}$H$_{23}$ CON C$_2$H$_4$OH $_2$	4	8.071	− 816.2	0.8562	9.657	2.333	287.4	− 3.585	− 3.509
C$_9$H$_{19}$ CON C$_2$H$_4$O $_3$CH$_3$ $_2$	5	6.657	− 1385	0.8962	9.611	3.438	463.6	− 2.299	− 2.476
C$_9$H$_{19}$ CON C$_2$H$_4$O $_4$CH$_3$ $_2$	6	6.657	− 1709	0.8900	9.619	3.168	551.7	− 2.193	− 2.248
C$_{11}$H$_{23}$ CON C$_2$H$_4$O $_2$CH$_3$ $_2$	4	8.071	− 1097	0.9051	9.643	3.121	403.6	− 3.398	− 3.420
C$_{11}$H$_{23}$ CON C$_2$H$_4$O $_3$CH$_3$ $_2$	5	8.071	− 1421	0.8915	9.658	2.666	491.7	− 3.292	− 3.408
C$_{11}$H$_{23}$ CON C$_2$H$_4$O $_4$CH$_3$ $_2$	6	8.071	− 1743	0.9089	9.651	3.256	579.8	− 3.611	− 3.539
C$_{12}$H$_{23}$ NHCH CH$_3$ COO C$_2$H$_4$O $_4$H	6	8.778	− 1298	0.9434	9.478	2.043	433.6	− 3.413	− 3.299
C$_{12}$H$_{23}$ NHCH$_2$COO C$_2$H$_4$O $_4$H	6	8.778	− 1285	0.8932	9.582	1.693	419.6	− 3.474	− 3.246
C$_{12}$H$_{23}$ N CH$_3$ CH$_2$COO C$_2$H$_4$O $_4$H	6	8.778	− 1290	0.9690	9.441	1.821	433.6	− 3.533	− 3.259
Fluorinated linear ethoxylate and amide									
C$_6$F$_{13}$C$_2$H$_4$SC$_2$H$_4$ OC$_2$H$_4$ $_2$OH	3	11.97	− 3237	− 1.324	9.399	2.87	512.3	− 4.602	− 4.828
C$_6$F$_{13}$C$_2$H$_4$SC$_2$H$_4$ OC$_2$H$_4$ $_3$OH	4	11.97	− 3398	− 1.326	9.402	2.774	556.3	− 4.553	− 4.775
C$_6$F$_{13}$C$_2$H$_4$SC$_2$H$_4$ OC$_2$H$_4$ $_5$OH	6	11.97	− 3722	− 1.326	9.405	2.779	644.4	− 4.432	− 4.696
C$_6$F$_{13}$C$_2$H$_4$SC$_2$H$_4$ OC$_2$H$_4$ $_7$OH	8	11.97	− 4045	− 1.326	9.406	2.776	732.5	− 4.319	− 4.614
C$_6$F$_{13}$CH$_2$ CON C$_2$H$_4$O $_3$CH$_3$ $_2$	5	8.621	− 3859	− 1.492	10.09	4.743	669.4	− 3.260	− 3.220
C$_8$F$_{17}$CH$_2$ CON C$_2$H$_4$O $_3$CH$_3$ $_2$	5	11.13	− 4665	− 1.765	10.06	2.996	769.4	− 4.921	− 4.956
C$_{10}$F$_{21}$CH$_2$ CON C$_2$H$_4$O $_3$CH$_3$ $_2$	5	13.64	− 5470	− 1.866	10.07	3.16	869.5	− 6.523	− 6.523

Correlation analysis

Descriptor analysis and subsequent regression analysis were carried out with SPSS 10.0 statistical software under the Microsoft Windows operating system. This software can provide available statistical analysis techniques including principle component analysis best multylinear regression analysis and a heuristic method.

Results and discussion

Selection of descriptors

Up to now there are more than 300 different molecular

descriptors having been set for the study of QSPR for different properties of materials.[10] However for the aqueous solution of nonionic surfactants most of the descriptors will not be relation with cmc. The formation of micelle is a process that the total free energy of system automatically reduces as the concentration of surfactant increasing. Below the critical micelle concentration the surfactant added to the solution remains in monomer form or adsorbs on the solution surface but above that essentially all additional surfactants form micelles. It is well known that accompanying the formation of micelle some other properties such as the enthalpy and the entropy of solution as well as the dipole moment of surfactant molecule will be changed.[2,3] The capacity of micelle formation is also firmly related to the structure of surfactant molecule. The surfactant molecule with a larger volume of hydrophobic segment and a smaller volume of hydrophilic head is provided with that it is thermodynamically favorable for the molecule to leave the aqueous solution and form micelle. Therefore the original structure properties of a surfactant i.e. ΔH_f E_{HOMO} E_{LUMO} D E_T M n_{NO} $KH0$ and so on could be expected to use as descriptors. The correlation analysis between lg cmc and each of the above-mentioned descriptors is illustrated in Table 2.

In this table pcc and p are the values of the pearson correlation coefficients and the one-tailed significance tests for the linear correlation analysis between lg cmc and each descriptor respectively. It can be seen that the values of lg cmc of the nonionic surfactants are somehow linearly related to the above selected descriptors with the highest behavior for $KH0$ -0.916 and the lowest for E_{HOMO} 0.141 which are separately shown in Figs. 1 and 2.
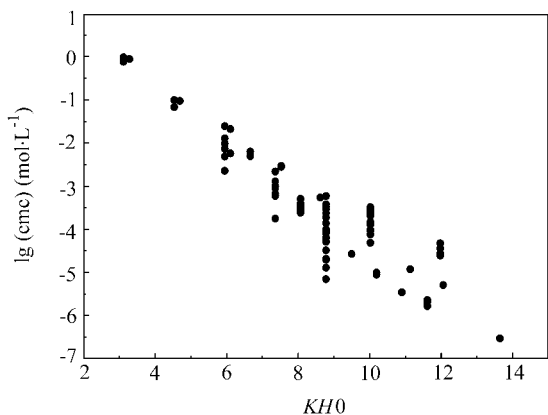

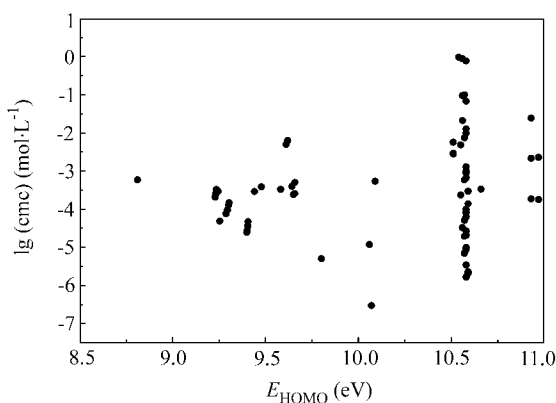**Fig. 1** Scatter plot of the calculated lg cmc vs. $KH0$ of hydrophobic fragments for 77 nonionic surfactants.

Fig. 2 Scatter plot of the calculated lg cmc vs. E_{HOMO} for 77 nonionic surfactants.

Multi-descriptor linear correlation analysis

By combining the principal component analysis with the best multilinear regression analysis and a heuristic method the multiple linear regression analysis between the surfactant molecular descriptors and lg cmc was made. Tables 3 and 4 illustrate the best correlation models of lg cmc created with the stepwise multi-descriptor linear correlation analysis method. From Table 3 it can be seen that the most important predictor is the Kier & Hall index of zero order for the hydrophobic fragment which gives the squared correlation coefficient high up to 0.840. This result fits the empirical conclusion well that cmc of nonionic surfactants is mainly determined by the volume of the hydrophobic fragment.

The secondly most impotent factor is the lowest unoccupied molecular orbital E_{LUMO} which can immediately improve the model by enhancing the squared correlation coefficient from 0.840 to 0.961 Model 2 . Since there are hydrogen bonds may formed in the solution with the lone electron pair on the oxygen atom in water molecule as the donor and the lowest unoccupied molecular orbital of the nonionic surfactant as the acceptor a lower E_{LUMO} of surfactant will be favorable for the formation of hydrogen bonds. Also since the hydrogen bonds between the surfactants and the water molecules will result the formation of micelle being unfavorable the values of lg cmc s increase as the values of E_{LUMO} go up. E_{HOMO} can just slightly improve the model Model 4 which indicates that another kind hydrogen bond on the contrary by using the electron pair in the highest occupied molecular orbital of surfactant as the donor and the hydrogen atom of water molecule as the acceptor is more difficult.

Table 2 Pearson correlation coefficient pcc between lg cmc 25 ℃ and each descriptors for 77 nonionic surfactants[a]

Descriptor	$KH0$	M	E_T	ΔH_f	E_{LUMO}	D	n_{ON}	E_{HOMO}
pcc	-0.916	-0.550	0.514	0.426	0.309	-0.211	-0.207	0.141
p	0.035	0.000	0.000	0.000	0.111	0.003	0.033	0.000

[a] Dependent variable lg cmc .

Table 3 Best correlation models and their statistical characteristics for 77 nonionic surfactants[a]

Number of parameter	Predictor							R^2	F	s^2
1	Constant	$KH0$						0.840	392.7	0.5453
2	Constant	$KH0$	E_{LUMO}					0.961	448.2	0.3785
3	Constant	$KH0$	E_{LUMO}	D				0.968	367.6	0.3439
4	Constant	$KH0$	E_{LUMO}	D	E_{HOMO}			0.974	331.0	0.3156
5	Constant	$KH0$	E_{LUMO}	D	E_{HOMO}	n_{ON}		0.979	320.6	0.2882
6	Constant	$KH0$	E_{LUMO}	D	E_{HOMO}	n_{ON}	M	0.983	329.0	0.2608
7	Constant	$KH0$	E_{LUMO}	D	E_{HOMO}	n_{ON}	M f	0.986	357.5	0.2325

[a] Dependent variable lg cmc .

Table 4 Correlations of the best models and their statistical characteristics with different parameters for 77 nonionic surfactants[a]

Model	Operator	Coefficient	Std. error	t-Test	Sig.
1	Constant	1.204	0.246	4.889	0.000
	$KH0$	−0.557	0.028	−19.817	0.000
2	Constant	2.857	0.250	11.409	0.000
	$KH0$	−0.687	0.024	−28.351	0.000
	E_{LUMO}	−0.430	0.048	−9.034	0.000
3	Constant	3.098	0.235	13.180	0.000
	$KH0$	−0.685	0.022	−31.111	0.000
	E_{LUMO}	−0.464	0.044	−10.538	0.000
	D	−0.114	0.028	−4.081	0.000
4	Constant	6.406	0.890	7.194	0.000
	$KH0$	−0.683	0.020	−33.815	0.000
	E_{LUMO}	−0.328	0.054	−6.096	0.000
	D	−0.109	0.026	−4.254	0.000
	E_{HOMO}	−0.343	0.090	−3.829	0.000
5	Constant	6.296	0.814	7.738	0.000
	$KH0$	−0.707	0.019	−36.403	0.000
	E_{LUMO}	−0.368	0.050	−7.334	0.000
	D	−0.130	0.024	−5.408	0.000
	E_{HOMO}	−0.336	0.082	−4.103	0.000
	n_{ON}	4.855E-02	0.012	3.917	0.000
6	Constant	4.879	0.814	5.993	0.000
	$KH0$	−0.669	0.020	−33.762	0.000
	$E\ LUMO$	−0.544	0.063	−8.690	0.000
	D	−0.104	0.023	−4.588	0.000
	E_{HOMO}	−0.164	0.085	−1.928	0.058
	n_{ON}	0.123	0.021	5.747	0.000
	M	−2.132E-03	0.001	−4.084	0.000
7	Constant	8.004	1.019	7.856	0.000
	$KH0$	−0.660	0.018	−37.113	0.000
	E_{LUMO}	−0.267	0.084	−3.161	0.002
	D	−0.138	0.022	−6.357	0.000
	E_{HOMO}	−0.515	0.110	−4.661	0.000
	n_{ON}	0.129	0.019	6.756	0.000
	M	−3.596E-03	0.001	−6.272	0.000
	ΔH_f	−4.350E-04	0.000	−4.370	0.000

[a] Dependent variable lg cmc .

The best seven-parameter model produces the highest squared correlation coefficient to 0.986 which is almost equal to 1. This result indicates that it is practical to use these simple descriptors to predict cmc of divers nonionic surfactants just from their molecular structures. The correspondent scatter plot of calculated lg cmc cal. vs. observed lg cmc ob. for Model 7 in Table 4 is shown in Fig. 3 which produces a linear regression equation between calculated lg cmc and observed lg cmc as

$$\text{lg cmc cal.} = 0.02655 + 1.0045 \text{ lg cmc ob.}$$
$$r = 0.993 \quad s^2 = 0.16315 \quad n = 77 \quad p < 0.0001 \qquad 4$$

In Table 2 the correlation coefficient between lg cmc and E_T is the third largest value 0.514 but E_T does not appear in the optimistic models in Table 3. By further examining the correlation coefficient between E_T and ΔH_f which is high up to 0.969 the possible reason may be that both ΔH_f and E_T represent the energy properties and have the same influence on cmc. Fig. 4 shows this tendency of this change.
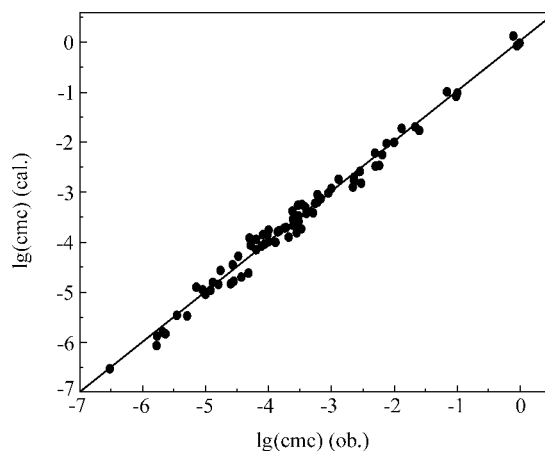
Fig. 3 Scatter plot of the calculated lg cmc vs. the observed lg cmc for 77 nonionic surfactants.

Besides the single operators the correlation coefficients on cross terms of them i.e. the product of $x_i x_j$ where $x_i x_j = n_{NO}$ $KH0$ E_T ΔH_f E_{HOMO} E_{LUMO} D and M have been calculated and only the term $KH0$ n_{ON}

shows a significant relation with lg cmc （ -0.566 ）. By combining both the single and the cross terms together，the multilinear regression analysis with the stepwise heuristic method produces the best model with only five parameters （Table 5）. The squared correlation coefficient is also 0.986，which equals that of the best Model 7 in Table 4.
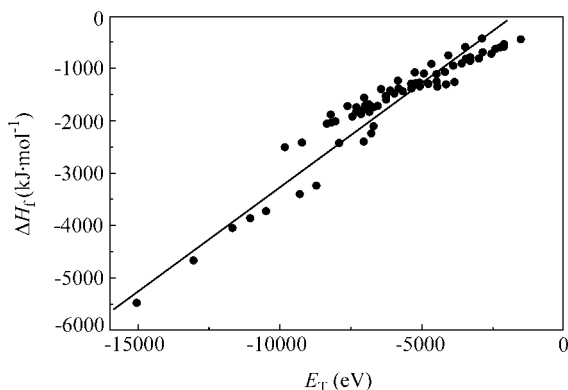

Fig. 4   Scatter plot of ΔH_f vs. E_T for 77 nonionic surfactants （ $\Delta H_f = 478.8 + 0.3511 E_T$，$r = 0.969$，$s^2 = 295.1$，$p < 0.000$ ）.

Table 5   Correlations of the best model and their statistical characteristics containing the cross term n_{ON}. $KH0$ for 77 nonionic surfactants （ $R^2 = 0.986$ $F = 425.9$ $s^2 = 0.000$ stepwise method ）[a]

Operator	Coefficient	Std. error	t-Test	Sig.
Constant	4.324	0.232	18.598	0.000
$KH0$	-0.774	0.021	-36.361	0.000
E_{LUMO}	-0.671	0.042	-16.106	0.000
D	-0.104	0.022	-4.783	0.000
n_{ON}. $KH0$	1.611E-02	0.002	8.097	0.000
M	-2.546E-03	0.000	-6.202	0.000

[a] Dependent variable：lg cmc .

Conclusion

1 A successful example has been illustrated for the prediction of surfactant properties with a general QSPR methodology. The established best models with different numbers of descriptors for predetermining the cmc on eight series of nonionic surfactant contain just the easily computed descriptors but most of them can produce correlation equations with correlation coefficient larger than 0.96 and with the highest up to 0.986.

2 The structure feature that mostly influences cmc is the size of the hydrophobic segment of nonionic surfactants. This phenomenon is in qualitative agreement with the knowledge about the nature of water around solutes as well as Huibers' remarks.[10] In order to form cavities around the surfactant molecule，the molecules of water have to rearrange regularly and this rearrangement distorts their bonding network. Companying this rearrangement both the enthalpy and entropy of the solution will vary. These changes are also influenced by the other aspects of surfactant such as the energies of the lowest unoccupied and highest occupied molecular orbitals，the dipole moment of surfactant and so on. Since these selected descriptors can capture both the hydrophobic and the hydrophilic properties of surfactants for different molecular structures （for instance，the branch of tail，the position of the head group and its structure diversity），they can successfully predict the properties of surfactants.

3 Because of the complexity of the solution of surfactants，almost no work has been carried out in the prediction of surface activity for surfactant by using the theory of the quantum chemistry so far. This work shows a good example to study of the QSPR of surfactant in colloid chemistry field by using the quantum chemistry and also provides to gain insight into the structural aspects of surfactant as well as to allow for estimation of cmc of the nonionic surfactants that have not yet been synthesized.

References

1 Rosen M. J. *Surfactants and Interfacial Phenomena*，John Wiley and Sons Inc. New York **1987**，pp. 122—132.

2 Elimelech M. Gregory J. Jis X. Williams R. A. *Particle Deposition and Aggregation：Measurement，Modeling and Simulation*，Butterworth Oxford **1995**，pp. 121—256.

3 Rosen M. J. *J. Colloid Interface Sci*. **1976**，*56*，320.

4 Meguro K. Takasawa Y. Kawahashi N. Tabata Y. Ueno M. *J. Colloid Interface Sci*. **1981**，*83*，50.

5 Ravey J. C. Fherbi A. Stebe M. J. Prog J. *Colloid Polym. Sci*. **1988**，*76*，234.

6 Becher P. *J. Dispersion Sic. Technol*. **1984**，*5*，81.

7 Puvada S. Blankschtein D. *J. Chem. Phys*. **1990**，*92*，3710.

8 Nagarajan R. Ruckenstein E. *Langmuir* **1991**，*7*，2934.

9 Huibers P. D. T. Lobanov V. S. Katritzky A. R. Shah D. O. Karelson M. *Langmuir* **1996**，*12*，462.

10 Wang Z.-W. Li G.-Z. Zhang X.-Y. Li L. *Acta Chim. Sinica* **2002**，*60*，1548（in Chinese）.

11 Wang Z.-W. Li G.-Z. Mu J.-H. Zhang X.-Y. Lou A.-J. *Chin. Chem. Lett*. **2002**，*13*（4），363.

12 Wang Z. W. Li G. Z. Zhang X. Y. Wang R. K. Lou A. *J. Colloids and Surfaces A：Physicochem. and Engineering Aspects* **2002**，*197*，37.

13 Steward J. J. *J. Comp. Chem*. **1989**，*10*，209.

14 Alan R. K. Victor S. L. Mati K. *Chem. Soc. Rev*. **1995**，*24*，279.