

# QSPR Study of Critical Micelle Concentration of Anionic Surfactants Using Computational Molecular Descriptors<sup>†</sup>

Alan R. Katritzky,<sup>\*,‡</sup> Liliana Pacureanu,<sup>‡</sup> Dimitar Dobchev,<sup>‡,||</sup> and Mati Karelson<sup>§,||</sup>

Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, Florida 32611, Institute of Chemistry, Tallinn University of Technology, Ehitajate tee 5, Tallinn 19086, Estonia, and Chemistry Department, University of Tartu, 2 Jakobi Street, Tartu 51014, Estonia

Received October 27, 2006

A data set of 181 diverse anionic surfactants has been investigated to relate the logarithm of critical micelle concentration (cmc) to the molecular structure using Comprehensive Descriptors for Structural and Statistical Analysis (CODESSA Pro) software. A fragment approach provided superior quantitative structure–property relationship (QSPR) models in terms of statistical characteristics and predictive ability. The regression equations provided insight into the structural features of surfactants that influence cmc. The most obvious influence on cmc was manifested by hydrophobic fragments expressed by the topological and geometrical descriptors, while the hydrophilic fragment is represented by constitutional, geometrical, and charge related descriptors. Significantly important molecular descriptors in the selected QSPR models were topological, solvational, and charge-related descriptors as the driving force of the intermolecular interactions between anionic surfactants and water.

## 1. INTRODUCTION

The driving force of the absorption process is the ordering of the water molecules in close proximity, the so-called hydrophobic effect. When the concentration of a surfactant in water is increased, surfactant molecules that cannot reside at the interface form aggregates called micelles, and this particular concentration is called the critical micelle concentration.<sup>1</sup> Critical micelle concentrations (cmc) are very important from a practical point of view and depend on a series of internal and external factors including molecular structure, temperature, pH, and the presence of additives.

Anionic surfactants are important due to their wide application in industry and the domestic area. Some aspects include the following: detergency, solubilizing agents, emulsifying agents, dispersing agents, fire fighting foams, antifogging agents, antistatic agents, adhesives, coatings, corrosion inhibitors, pharmaceutical adjuvants, pseudostationary phases in micellar electrokinetic chromatography, etc.<sup>2,3</sup> A recent utilization of surfactants is the stabilization of the voltametric signal of neurotransmitter serotonin.<sup>4</sup> Anionic polymerizable surfactants and the corresponding polymers are of particular practical interest but are less suitable for modeling.<sup>5</sup>

The chemical structure of anionic surfactants includes carboxylic acid salts, esters or ethers, alkyl sulfates, alkyl ether sulfates, alkanesulfonates, alkylbenzenesulfonates, alkyl phosphates, alkylphenol ethoxylated phosphates, acylamino acids, acylpeptides, diesters such as sulfo-fatty acid esters,

etc. Among the most frequently used anionic surfactants are the salts of linear and branched alkyl sulfates and alkane-sulfonates, alpha olefin sulfonates, substituted alkane-sulfonates, heteroatom containing sulfates and sulfonates (O, N), phosphate esters, etc. The phosphate, phosphinate, and phosphonate containing surfactants are used in agrochemical emulsions together with fertilizers.<sup>6,7</sup>

The cmc of anionic hybrid surfactants follow Kleven's law<sup>7,8</sup> and relates the cmc to the fluorocarbon (*m*)/hydrocarbon (*n*) chain length as follows

$$\log \text{cmc} = A - Bm \text{ (or } n) \quad (1)$$

where *A* and *B* are constants, and *B* shows the contribution of hydrophobic CF<sub>2</sub> (CH<sub>2</sub>) to the decrease of cmc. Always CF<sub>2</sub> has a larger contribution to the decrease of cmc since the fluorocarbon chain is more hydrophobic than hydrocarbon chain. The investigation of micelle formation of hybrid phosphate type surfactants as a function of temperature shows that the CH<sub>2</sub> group hydrophobicity is considerably reduced with respect to hydrocarbon surfactants probably due to intramolecular fluorocarbon–hydrocarbon interaction that provide the contact of hydrocarbon chain with water in the micelle state.<sup>8</sup>

Kinetic studies showed that the effect of hydrophobic alkyl chains may not be accounted for on the basis of additivity of pairwise group interaction due to the sulfate group that hampers the ability of nearby methylene moieties to undergo intermolecular interactions.<sup>9</sup> On the other hand previous studies have revealed that the number of water molecules in the proximity of surfactants is dependent on the hydrophobic group volume.

An important factor that influences cmc is the temperature. The micellization of anionic surfactants is exothermic, and in the temperature range of 20–30 °C the slope of the curve

<sup>†</sup> Dedicated to Professor Nenad Trinajstić on the occasion of his 70th birthday.

<sup>\*</sup> Corresponding author phone: (352)392-0554; fax: (352)392-9199; e-mail: katritzky@chem.ufl.edu.

<sup>‡</sup> University of Florida.

<sup>§</sup> Tallinn University of Technology.

<sup>||</sup> University of Tartu.

that indicates the cmc dependence on temperature changes sign. In the case of some ionic surfactants, cmc decreases to a minimum value  $X^*\text{cmc}$ , near to 25 °C, and after that the cmc increases with an increase in temperature; the shape of the plot of cmc versus temperature is parabolic. Recently, a rigorous general equation relating cmc to the temperature  $T$  was developed

$$|X_{\text{cmc}} - X^*\text{cmc}| = \text{const } |T - T^*|^2 \quad (2)$$

where  $X^*\text{cmc}$  is the minimum cmc at temperature  $T^*$ . Eq 2 was tested on 28 different types of surfactant systems, and it fits irrespective to structural characteristics.<sup>10–12</sup>

Recently, several QSPR methodologies based on semiempirical and ab initio calculations have been employed to investigate the cmc of anionic surfactants.<sup>3,13–15</sup>

In collaboration with Huibers and Shah, and following joint studies of nonionic surfactants,<sup>13</sup> our group used 119 cmc values of anionic surfactants (sulfates and sulfonates) to develop quantitative relationships between log cmc and molecular structure.<sup>13</sup> The regression equation obtained (eq 3) contains three variables that include features of both hydrophobic and hydrophilic fragments of the surfactant and is in accordance with the structural effects on cmc (squared correlation coefficient  $R^2 = 0.940$ , Fisher criterion  $F = 597$ , standard deviation  $s^2 = 0.0472$ ).

$$\begin{aligned} \log \text{cmc} = & (1.89 \pm 0.11) - (0.314 \pm \\ & 0.010)\text{t-sum Kier \& Hall index (0th order)} - (0.034 \pm \\ & 0.003)\text{total dipole of the molecule} - (1.45 \pm \\ & 0.18)\text{h-sum-relative number of carbon atoms} \quad (3) \end{aligned}$$

In (3) the t-sum-Kier & Hall index of 0th order is the sum of the Kier & Hall index of zeroth order over all hydrophobic tails; the total dipole of the molecule is the AM1 calculated total dipole moment of the molecule; and the h-sum relative number of carbon atoms is the sum of the relative number of carbon atoms over all hydrophilic heads. A large number of molecular descriptors were explored in the initial descriptor pool, but the cmc was proved to be appropriately modeled by the use of fragment descriptors defined for the hydrophobic and hydrophilic domain of the molecule.

Roberts correlated the cmc of anionic surfactants including ether sulfates and ester sulfonates with good results ( $R^2 = 0.976$ , 133 compounds) using the hydrophobic parameter  $\pi_h$  for the hydrophobic domain of the surfactant.<sup>15</sup>

Li et al. developed a general QSPR model for 98 anionic surfactants using the ab initio method (RHF/6-31G(d)) to calculate the electronic structures. They obtained a three-parameter regression equation with good statistical characteristics ( $R^2 = 0.980$ ,  $R^2_{\text{cv}} = 0.978$ ,  $F = 1505.23$ ,  $s^2 = 0.0107$ ) which involves variables as the total number of atoms in the hydrophobic–hydrophilic segment, the maximum atomic charge on C atom, and the dipole moment.<sup>16</sup>

The QSPR analysis performed by Yuan et al. related the log cmc of 37 alkyl sulfate anionic surfactants with the selected descriptors octanol–water partition coefficient (logP), dipole moment of the molecule, and the area of the molecule ( $R^2 = 0.980$ ).<sup>14</sup>

Jalali-Heravi and Konouz developed a QSPR model for the cmc of 31 anionic surfactants (sulfates and sulfonates) using two topological indexes (Wiener and the reciprocal of

Randić index) and the reciprocal of the dipole moment ( $R^2 = 0.9907$ , SE = 0.0631). They concluded that the topology plays a major role in micelle formation and that polarity is somewhat less important.<sup>3</sup>

Recently, surfactants have become more interesting targets for theoretical studies giving considerable insight into the static and dynamic properties of these materials.<sup>17</sup> Meanwhile, the micelle–water partition coefficient ( $\log P_{\text{mw}}$ ) was successfully employed to describe the hydrophobicity of organic compounds in chemical–biological interactions.<sup>18</sup> An increase in computer power has enabled sophisticated large scale numerical simulations of the dynamics of surfactants solutions. The statistical thermodynamic theory of surfactant solution behavior has been employed to elaborate software that enables the prediction of micellar properties such as cmc and micelle size and shape.

Recently, considerable effort was devoted to all atom simulation of bilayers/membranes as models for biological membranes. The results provided insight into the organization of lipid molecules and brought about a detailed molecular level picture of the phenomena that are in agreement with experimental determinations.<sup>19–22</sup>

Performing molecular simulations that include the solvent effects and a significant number of surfactant molecules became prohibitive in terms of time at the current level of computer power and computational algorithms. Therefore, the relationship between the chemical structure and physico-chemical properties (QSPR) are of considerable importance, because it can offer faster, precise, and cheaper means to interpret and assess the structural parameters that determine the physical property and to predict the property for new compounds without the necessity to synthesize them.

The present paper summarizes the results of our study to correlate the logarithm of cmc to the molecular structure of anionic surfactants using a general QSPR approach that involves the calculated descriptors for the whole molecule and its predefined fragments. First, we examine whether the new data could fit with the previous QSPR model (eq 3), and, second, new models are created that can explain better the micellization potentials of anionic surfactants.

## 2. DATA SETS

In this study we combine two data sets of surfactants. The first data set<sup>13</sup> comprises classical anionic surfactants including linear and branched alkyl sulfates and alkanesulfonates, alkylbenzenesulfonates, hydroxyalkanesulfonates, esters of sulfo monocarboxylic acids, and dialkyl sulfosuccinates, in total 119 structures. This data set was used in our previous work for developing the QSPR model of the cmc.<sup>13</sup> The second new set of 62 anionic surfactants contains, besides alkyl (aryl) sulfates and alkanesulfonates, also diverse amino acid based anionic surfactants, hybrid fluorocarbon/hydrocarbon sulfates, sugar-based anionic surfactants, alkanecarboxylates, and p-isooctylphenol ethoxylated phosphates. Their critical micelle concentrations were collected from refs 2, 8, 10, 13, 15, 16, and 23–26. Thus, we combine both data sets so that we can increase the number (181) of the compounds and the statistical reliability of the QSPR models. By doing this it also increases the diversity of the surfactants which can lead to QSPR models with expanding prediction of wide range of surfactants.

The compounds in the whole data set (181) are all anionic surfactants that were sodium salts except for potassium alkanecarboxylates and p-isooctylphenol ethoxylated phosphates. The experimental determination of cmc for most of the anionic surfactants was at 25 °C or at 40 °C. We carried out the modeling for cmc at 40 °C because anionic surfactants with extended hydrophobic fragments have the Krafft point between 25 and 40 °C. The point at which the temperature-solubility curve of the surfactant crosses the temperature-cmc curve is called the Krafft point and represents a very important characteristic since no surface activity can be obtained below this point.<sup>27</sup> We used the ratio between cmc at 40 and that at 25 °C measured in the same experiment, as was established in our previous paper, i.e., 1.088 for sulfonates and 1.030 for sulfates.<sup>13</sup> We proceeded in this way since the range of the temperature correction is too small (3–9%) and also because it was encouraged by the quality of the previous results. The cmc of fluorinated anionic surfactants drops with an increase of temperature, and the ratio of the cmc at 25 °C to that at 40 °C is 1.259;<sup>28</sup> in the case of oxyethylenated surfactants the corresponding ratio is smaller 1.036.<sup>29</sup>

The calculation has been performed on the anion since the surfactants are dissociated in aqueous solution, and the effect of the cation on the cmc was neglected.

The data collected for anionic surfactants as well as the predicted values according to the QSPR models are given in Table 1.

### 3. METHODOLOGY

The geometry of surfactant molecules was preoptimized using the molecular mechanics force field (MM+) included in Hyperchem 7.5.<sup>30</sup> Molecular geometries were further refined using AM1 (Austin Model 1) semiempirical method<sup>31</sup> calculations. Up to 400 molecular descriptors, classified as (i) constitutional, (ii) topological, (iii) geometrical, (iv) charge-related, and (v) semiempirical, were calculated using the CODESSA Pro software.<sup>32</sup>

CODESSA Pro has been used to correlate numerous physical properties and biological activities with molecular structure and then to predict further values of inter alia boiling points,<sup>33</sup> partition coefficients (logD),<sup>34</sup> solvent scales,<sup>35</sup> viscosities,<sup>36</sup> binding energies for 1:1 complexation between organic guest molecules and  $\beta$ -cyclodextrin,<sup>37</sup> the in vitro minimum inhibitory concentration (MIC) of 3-aryloxazolidin-2-one antibacterials to inhibit growth of *Staphylococcus aureus*,<sup>38</sup> partition coefficients of drugs between human breast milk and plasma,<sup>39</sup> HIV-1 protease inhibitory activity of substituted tetrahydropyrimidinone,<sup>40</sup> and toxicities of polychlorodibenzofurans, polychlorodibenzo-1,4-dioxins, and polychlorobiphenyls.<sup>41</sup>

In this work we used fragment descriptors to build the QSPR models. Molecular fragments have been successfully included in previous structure–property correlations of nonionic and anionic surfactants.<sup>13,42</sup> Our experience has shown that the utilization of fragment descriptors improves the prediction of critical micelle concentration of surfactants. In this way there is a clear distinction between the hydrophobic and hydrophilic fragments<sup>13,42</sup> providing better statistical characteristics compared to the correlation employing solely molecular descriptors.<sup>13,42</sup>

The precise definition of the fragment descriptors is important and is indicated in Figure 3. According to the empirical equation (eq 1), an increase in the carbon atom number decreases the critical micelle concentration of anionic surfactants, while the nature of the hydrophilic group influences the micellization potential of anionic surfactants. We defined surfactant fragments for the hydrophobic and hydrophilic domains of the molecule. The hydrophobic fragments included all alkyl (hydrocarbon or fluorocarbon chains or rings) and aromatic hydrocarbon rings. The hydrophilic domain is defined to include all hydrogen bonding acceptor atoms. For the surfactants that contain multiple hydrophilic groups, e.g., sulfonate and alkyl amino or alkoxy, the hydrophobic fragment is selected so it does not include N or O atoms. In some cases the hydrophilic fragments included several carbon atoms contained in polyethyleneoxide units or methoxy groups. For structures containing multiple hydrophobic tails and/or hydrophilic fragments, we calculated the sum of the corresponding descriptors over all the hydrophobic tails and hydrophilic heads.

In our case, the calculation of fragment descriptors differs from the methodology described by Roberts in ref 15. He calculated the hydrophobic parameter  $\pi_h$  for the hydrophobic part of the molecule using logP values for the whole molecule and then subtracted the logP value for the ionic group  $\text{OSO}_3^-$  or  $\text{SO}_3^-$ , taking into account branching effects and the contribution of the other hydrophilic fragments of the molecule, e.g., oxyethylene units. Both approaches offer accurate results since there is a clear separation of hydrophobic and hydrophilic domains, while in our case branching effects are accounted for by the topological descriptors.

The heuristic algorithm approach was employed as a statistical technique for the development of the multilinear regression model over a given number of descriptors.<sup>32</sup> This technique involves the following steps: (1) The descriptors are screened in order to establish the significance, correlation degree, and missing values based on the squared correlation coefficient, Fisher F-criteria, and student t-criteria to reduce the number of significant variables. (2) The second stage is the heuristic stepwise procedure to obtain the best multiparameter linear regression equation deduced from the reduced number of variable-descriptors. (3) The final correlation equations are selected on the basis of the highest squared correlation coefficient, Fisher F-criteria, and standard errors. In this study a modified approach<sup>43</sup> for QSPR modeling is used that aims to combine the advantages of the two QSPR methods most frequently applied, i.e., (i) use of all data points to build the model and to apply as a validation method only the internal cross-validation procedures for validation or (ii) to use only a part of the available data for building the general model, keeping the remaining data points for external validation. The following procedure was used to build a QSPR model: 1. All data points were ordered in the ascending order of log cmc values. 2. The series was separated into three subsets (conditionally denoted as A, B, and C) by selection of every third point from the original data set (181) in order to obtain a similar distribution of the investigated property values for the whole set. 3. Three new data sets were constructed using all combinations of the binary sums: A+B, A+C, and B+C. 4. The standard QSPR modeling procedure including Best Multiple Linear Regres-



**Table 1.** Experimental log cmc Values from the Literature and the Predicted Values Derived from the Best QSPR Models for Anionic Surfactants

no.	surfactant code	cmc experimental			cmc predicted		no.	surfactant code	cmc experimental			cmc predicted	
		reported	ref	corrected	eq 3	eq (Table 5)			reported	ref	corrected	eq 3	eq (Table 5)
1.	C6SO4Na	-0.377 <sup>a</sup>	24	-0.364	-0.045	-0.030	76.	C10C(C)SO3Na	-1.827	13		-1.699	-1.757
2.	C7SO4Na	-0.658 <sup>a</sup>	24	-0.645	-0.351	-0.538	77.	C9C(C2)SO3Na	-1.730	13		-1.529	-1.635
3.	C9SO4Na	-1.222 <sup>a</sup>	24	-1.209	-0.972	-1.297	78.	C8C(C3)SO3Na	-1.635	13		-1.429	-1.521
4.	C8CO2K	-0.333	16		-1.177	-0.512	79.	C7C(C4)SO3Na	-1.548	13		-1.403	-1.455
5.	C10CO2K	-0.936	16		-1.817	-1.368	80.	C6C(C5)SO3Na	-1.442	13		-1.326	-1.375
6.	C12CO2K	-1.538	16		-2.460	-2.009	81.	C7C(C7)SO3Na	-2.144	13		-2.052	-2.197
7.	C14CO2K	-2.137	16		-3.106	-2.551	82.	C7PhSO3Na	-1.582	13		-1.591	-1.688
8.	C16CO2K	-2.745	16		-3.753	-3.040	83.	C8PhSO3Na	-1.907	13		-1.893	-1.978
9.	C6C(C)SO4Na	-0.745	16		-0.525	-0.487	84.	C6C(C2)PhSO3Na	-1.967	13		-2.077	-2.010
10.	C8C(C2)SO4Na	-1.539	16		-1.272	-1.457	85.	C8C(C)PhSO3Na	-2.303	13		-2.428	-2.369
11.	C5C(C5)SO4Na	-1.081	16		-1.062	-1.046	86.	C7C(C2)PhSO3Na	-2.200	13		-2.469	-2.578
12.	C11C(C)SO4Na	-2.187	16		-2.028	-2.204	87.	C5C(C4)PhSO3Na	-2.047	13		-2.310	-2.129
13.	C6C(C6)SO4Na	-1.714	16		-1.518	-1.797	88.	C9C(C)PhSO3Na	-2.721	13		-2.725	-2.617
14.	C14C(C14)SO4Na	-4.100	16		-5.147	-4.856	89.	C10C(C)PhSO3Na	-2.692	13		-3.027	-2.856
15.	C10PhSO3Na	-2.509	16		-2.507	-2.502	90.	C9C(C2)PhSO3Na	-2.606	13		-2.947	-2.798
16.	C12PhSO3Na	-2.921	16		-3.076	-2.955	91.	C8C(C3)PhSO3Na	-2.721	13		-2.881	-2.722
17.	C8C(C2)PhSO3Na	-2.500	16		-2.520	-2.360	92.	C6C(C)PhSO3Na	-2.585	13		-2.839	-2.592
18.	C7C(C3)PhSO3Na	-2.400	16		-2.632	-2.473	93.	C11C(C)PhSO3Na	-3.208	13		-3.329	-3.084
19.	C6C(C4)PhSO3Na	-2.300	16		-2.590	-2.395	94.	C13C(C)PhSO3Na	-3.577	13		-3.940	-3.522
20.	C5C(C5)PhSO3Na	-2.250	16		-2.584	-2.359	95.	C8SO4Na	-0.854	13		-0.658	-0.948
21.	C7C(C4)PhSO3Na	-2.570	16		-2.819	-2.632	96.	C10SO4Na	-1.481	13		-1.284	-1.604
22.	C10C(C2)PhSO3Na	-3.000	16		-3.247	-3.036	97.	C11SO4Na	-1.783	13		-1.599	-1.884
23.	C9C(C3)PhSO3Na	-2.900	16		-3.210	-2.974	98.	C12SO4Na	-2.066	13		-1.915	-2.145
24.	C8C(C4)PhSO3Na	-2.780	16		-3.142	-2.898	99.	C13SO4Na	-2.367	13		-2.232	-2.393
25.	C7C(C5)PhSO3Na	-2.700	16		-3.117	-2.836	100.	C14SO4Na	-2.658	13		-2.551	-2.631
26.	C6C(C6)PhSO3Na	-2.600	16		-3.118	-2.816	101.	C15SO4Na	-2.921	13		-2.869	-2.860
27.	C12C(C)PhSO3Na	-3.390	16		-3.697	-3.330	102.	C16SO4Na	-3.237	13		-3.189	-3.084
28.	C11C(C2)PhSO3Na	-3.280	16		-3.591	-3.276	103.	C18SO4Na	-3.796	13		-3.830	-3.519
29.	C10C(C3)PhSO3Na	-3.150	16		-3.507	-3.210	104.	C8C(C)SO4Na	-1.328	13		-1.123	-1.318
30.	C9C(C4)PhSO3Na	-3.050	16		-3.425	-3.137	105.	C12C(C)SO4Na	-2.481	13		-2.215	-2.397
31.	C8C(C5)PhSO3Na	-2.900	16		-3.351	-3.032	106.	C11C(C2)SO4Na	-2.367	13		-2.132	-2.332
32.	C7C(C6)PhSO3Na	-2.800	16		-3.387	-3.039	107.	C10C(C3)SO4Na	-2.288	13		-1.919	-2.215
33.	C10EO2SO4Na	-1.910 <sup>a</sup>	15	-1.9254	-2.108	-2.120	108.	C9C(C4)SO4Na	-2.171	13		-1.933	-2.182
34.	C12EO4SO4Na	-2.770 <sup>a</sup>	26	-2.7849	-3.333	-2.869	109.	C7C(C6)SO4Na	-2.013	13		-1.748	-1.967
35.	C14EO5SO4Na	-2.857 <sup>a</sup>	26	-2.8723	-3.045	-2.705	110.	C13C(C)SO4Na	-2.767	13		-2.521	-2.637
36.	C14EO2SO4Na	-3.070 <sup>a</sup>	15	-3.0854	-3.380	-2.989	111.	C12C(C2)SO4Na	-2.658	13		-2.427	-2.578
37.	C14EO4SO4Na	-3.160 <sup>a</sup>	26	-3.1753	-3.970	-3.279	112.	C10C(C4)SO4Na	-2.469	13		-2.209	-2.440
38.	C16EO2SO4Na	-3.638	10		-4.017	-3.406	113.	C7C(C7)SO4Na	-2.177	13		-2.053	-2.232
39.	C8CO2C2SO3Na	-1.312	16		-1.130	-1.288	114.	C12C(C3)SO4Na	-2.764	13		-2.498	-2.708
40.	C10CO2C2SO3Na	-1.883	16		-1.733	-1.822	115.	C10C(C5)SO4Na	-2.629	13		-2.399	-2.527
41.	C12CO2C2SO3Na	-2.523	16		-2.342	-2.299	116.	C8C(C7)SO4Na	-2.372	13		-2.288	-2.460
42.	CF4C2C(C3)SO4Na	-1.602 <sup>a</sup>	6	-1.702	-1.819	-1.876	117.	C15C(C)SO4Na	-3.310	13		-3.142	-3.087
43.	CF6C2C(C3)SO4Na	-2.398 <sup>a</sup>	6	-2.498	-2.808	-2.588	118.	C8C(C8)SO4Na	-2.629	13		-2.444	-2.646
44.	CF8C2C(C3)SO4Na	-3.328 <sup>a</sup>	6	-3.42	-3.804	-3.204	119.	C16C(C)SO4Na	-3.585	13		-3.444	-3.307
45.	CF4C2C(C5)SO4Na	-2.041 <sup>a</sup>	6	-2.141	-2.199	-2.291	120.	C14C(C3)SO4Na	-3.347	13		-3.095	-3.159
46.	CF6C2C(C5)SO4Na	-2.921 <sup>a</sup>	6	-3.021	-3.182	-2.942	121.	C12C(C5)SO4Na	-3.143	13		-2.963	-2.980
47.	CF8C2C(C5)SO4Na	-3.796 <sup>a</sup>	6	-3.896	-4.052	-3.480	122.	C9C(C9)SO4Na	-3.027	13		-2.908	-3.043
48.	CF4C2C(C7)SO4Na	-2.387 <sup>a</sup>	6	-2.487	-2.579	-2.681	123.	C14C(C4)SO4Na	-3.481	13		-3.374	-3.352
49.	CF6C2C(C7)SO4Na	-3.301 <sup>a</sup>	6	-3.401	-3.549	-3.283	124.	C13C(C)CSO4Na	-3.097	13		-2.898	-2.874
50.	CF8C2C(C7)SO4Na	-4.155 <sup>a</sup>	6	-4.255	-4.540	-3.850	125.	C12C(C2)CSO4Na	-3.046	13		-2.757	-2.809
51.	CC(C)C5PhEO10PO4H2	-1.796	23		-2.802	-1.101	126.	C11C(C3)CSO4Na	-2.959	13		-2.657	-2.743
52.	CC(C)C5PhEO15PO4H2	-1.770	23		-3.902	-1.717	127.	C10C(C4)CSO4Na	-2.824	13		-2.558	-2.677
53.	CC(C)C5PhEO20PO4H2	-1.745	23		-4.620	-2.128	128.	C9C(C5)CSO4Na	-2.699	13		-2.482	-2.615
54.	C7C(O)NGLupyrSO4Na	-1.602	10		-1.214	-1.375	129.	C8C(C6)CSO4Na	-2.638	13		-2.414	-2.561
55.	C11C(O)NGLupyrSO4Na	-2.770	10		-2.394	-2.251	130.	C7C(C7)CSO4Na	-2.523	13		-2.387	-2.529
56.	C15C(O)NGLupyrSO4Na	-3.921	10		-3.587	-3.061	131.	C12EOSO4Na	-2.396	13		-2.408	-2.374
57.	C5PhePhSO3Na	-1.854 <sup>a</sup>	2	-1.817	-2.757	-2.038	132.	C12EO2SO4Na	-2.553	13		-2.740	-2.561
58.	C8PhePhSO3Na	-2.854 <sup>a</sup>	2	-2.817	-3.289	-2.903	133.	C9C(OH)C2SO3Na	-1.606	13		-1.855	-2.051
59.	C10PhePhSO3Na	-3.481 <sup>a</sup>	2	-3.445	-3.847	-3.494	134.	C9C(OC)C2SO3Na	-2.118	13		-2.103	-2.615
60.	C12PheSO3Na	-4.201 <sup>a</sup>	2	-4.164	-3.563	-3.310	135.	C9C(OC2)C2SO3Na	-2.301	13		-2.311	-2.877
61.	C14PhePhSO3Na	-4.796 <sup>a</sup>	2	-4.759	-4.933	-4.498	136.	C9C(OC)C2SO3Na	-2.420	13		-2.513	-3.052
62.	C12AlaPhSO3Na	-3.387 <sup>a</sup>	2	-3.351	-4.035	-3.359	137.	C9C(OC(C)C)C2SO3Na	-2.458	13		-2.591	-3.036
63.	C6SO3Na	-0.496	13		-0.025	0.183	138.	C9C(OC4)C2SO3Na	-2.817	13		-2.725	-3.228
64.	C8SO3Na	-0.796	13		-0.631	-0.961	139.	C9C(OC6)C2SO3Na	-3.185	13		-3.150	-3.583
65.	C10SO3Na	-1.398	13		-1.210	-1.684	140.	C9C(OC8)C2SO3Na	-3.922	13		-3.588	-3.954
66.	C12SO3Na	-1.959	13		-1.876	-2.280	141.	C9C(OCC(C2)C4)C2SO3Na	-3.501	13		-3.637	-3.909
67.	C13SO3Na	-2.421	13		-2.193	-2.539	142.	C9C(OPh)C2SO3Na	-2.708	13		-2.791	-3.349
68.	C14SO3Na	-2.602	13		-2.509	-2.785	143.	C9C(O)C2SO3Na	-1.541	13		-1.792	-2.291
69.	C15SO3Na	-3.139	13		-2.827	-3.022	144.	C11C(OH)C2SO3Na	-2.199	13		-2.478	-2.645
70.	C16SO3Na	-3.131	13		-3.145	-3.250	145.	C11C(OC2OH)C2SO3Na	-3.445	13		-2.719	-2.660
71.	C17SO3Na	-3.635	13		-3.432	-3.459	146.	C11C(EO2)C2SO3Na	-2.922	13		-2.662	-2.688
72.	C10C=CSO3Na	-1.886	13		-1.778	-2.130	147.	C11C(OPh)C2SO3Na	-3.641	13		-3.376	-3.892
73.	C12C=CSO3Na	-2.569	13		-2.412	-2.644	148.	C11C(OPhC13)C2SO3Na	-4.787	13		-4.266	-4.487
74.	C14C=CSO3Na	-3.215	13		-3.049	-3.115	149.	C11C(N(C)2)C2SO3Na	-2.964	13		-3.037	-3.179
75.	C16C=CSO3Na	-3.745	13		-3.688	-3.561	150.	C11C(NC3)C2SO3Na	-3.200	13		-3.105	-3.447

Table 1. (Continued)

no.	surfactant code	cmc experimental			cmc predicted		no.	surfactant code	cmc experimental			cmc predicted	
		reported	ref	corrected	eq 3	eq (Table 5)			reported	ref	corrected	eq 3	eq (Table 5)
151.	C11C(NC4)C2SO3Na	-3.708	13		-3.306	-3.622	167.	C14CO2C2SO3Na	-3.046	13		-2.959	-2.746
152.	C11C(morpholino)C2SO3Na	-3.106	13		-2.733	-2.679	168.	C10C(CO2C)SO3Na	-1.986	13		-2.050	-1.697
153.	C11C(piperidino)C2SO3Na	-3.310	13		-3.435	-3.892	169.	C12C(CO2C)SO3Na	-2.583	13		-2.646	-2.378
154.	C11C(O)C2SO3Na	-2.174	13		-2.419	-2.794	170.	C14C(CO2C)SO3Na	-3.398	13		-3.253	-2.971
155.	C13C(OH)C2SO3Na	-2.839	13		-3.106	-3.162	171.	C14C(CO2C2)SO3Na	-3.509	13		-2.964	-3.373
156.	C13C(OC)C2SO3Na	-3.472	13		-3.343	-3.287	172.	C14C(CO2C3)SO3Na	-3.964	13		-3.634	-3.410
157.	C13C(OC3)C2SO3Na	-4.089	13		-3.728	-4.131	173.	C16C(CO2C)SO3Na	-4.000	13		-3.868	-3.545
158.	C13C(OC4)C2SO3Na	-4.458	13		-3.928	-4.300	174.	C16C(CO2C2)SO3Na	-4.106	13		-4.055	-3.800
159.	C13C(O)C2SO3Na	-2.735	13		-3.050	-3.257	175.	C16C(CO2C3)SO3Na	-4.899	13		-4.234	-3.967
160.	C15C(OH)C2SO3Na	-3.420	13		-3.726	-3.633	176.	C16C(CO2C(C)C)SO3Na	-4.569	13		-4.262	-3.945
161.	C10EOSO3Na	-1.787	13		-1.766	-1.828	177.	C4C(C2)CCO2CC(SO3Na)-CO2CC(C2)C4	-2.566	13		-2.789	-2.620
162.	C10C(C(OH))SO3Na	-1.787	13		-1.463	-1.442	178.	C4CO2C(SO3Na)CCO2C4	-0.663	13		-0.875	-0.860
163.	C12C(C(OH))SO3Na	-2.432	13		-2.052	-2.096	179.	C5CO2C(SO3Na)CCO2C5	-1.239	13		-1.325	-1.692
164.	C6CO2CSO3Na	-0.733	13		-0.589	-0.600	180.	C6CO2C(SO3Na)CCO2C6	-1.817	13		-1.816	-1.889
165.	C8CO2CSO3Na	-1.144	13		-1.202	-1.284	181.	C8CO2C(SO3Na)CCO2C8	-3.131	13		-2.767	-2.778
166.	C10CO2CSO3Na	-1.621	13		-1.822	-1.837							

<sup>a</sup> Measured at 25 °C.

sion Method (BMLR) was applied to those three data sets obtained in step 3. 5. The complimentary parts to each of these three data sets (respectively C, B, and A) were used as external validation data sets, by considering their consistency. 6. All the descriptors that appeared in the obtained models of step 4 were investigated to obtain a general model including all the existing compounds. 7. The general model was again validated using classical internal cross-validation procedures: leave-one-out. Thus, the above modeling procedure guarantees that the general QSPR model uses all the compounds (181) and is simultaneously externally validated.

#### 4. RESULTS AND DISCUSSION

**4.1. QSPR Development.** *4.1.1. Testing of the Previous Equation.* Before building the general model we tested the previous QSPR model developed [eq 3 in ref 13] by our group for consistency and predictivity based on the second data set (62) not used in the earlier modeling. By doing this we tested the generality of the model using this external set of compounds that consists of more or less similar structures to the training set. The plot of the predicted versus experimental  $-\log$  cmc for the new set of 62 anionic surfactants, using eq 3 in ref 13, is shown in Figure 4.

The predicted cmc values are of fair quality. The new set of (62) anionic surfactants contains several types of structures which are not included in the previous data. These structures have two or three diverse hydrophilic head groups and/or long polyoxyethylene units ( $n = 10-20$ ) or long chain alkylsulfate. Increasing the number of oxyethylene group in alkyl sulfates or alkyl sulfonates increases the minimum area per molecule at the air/aqueous solution interface. This fact suggests a behavior of these surfactants which is different from the classical alkyl sulfates and sulfonates.<sup>29</sup> In this investigation we have also tried to reproduce the results of the original study.<sup>13</sup> Comparison between the results reported in the article and the calculations performed here led to a difference in the  $R^2$  values (0.940 - original and 0.936 - current, respectively). The reason for this discrepancy is due to the small differences registered for the descriptor total dipole of the molecule, which is due to the slightly different parametrization of the PC model software that was used in the previous calculation<sup>13</sup> with respect to the MM+.

The above critical results from the test and reproduction of the previous model lead to the need that a new general model based on more diverse surfactants is necessary.

*4.1.2. General QSPR Model for 181 Anionic Surfactants.* To build a reliable QSPR model with significant predictive ability the three training subsets A+B, A+C, and B+C (containing two-thirds of the initial data set) were subjected to QSPR treatment using the heuristic algorithm encoded in CODESSA Pro. The corresponding QSPR models obtained for these three sets (A+B, A+C, and B+C) were used further to predict the cmc values for the respective complimentary set of compounds (C, B, and A, respectively). The training sets were chosen by the experimental log cmc values ordered in an ascending way as was described in the Methodology section. Thus, it guarantees a similar distribution of the data as their parental set (181). In addition, we have also examined these data sets for their normal distribution. Figure 5 shows that the A+B, A+C, and B+C possess distributions close to the normal.

After running the Heuristic method the best selected multilinear equations for the A+B, A+C, and B+C sets were analyzed. The results for the models are shown in Tables 2-4 together with their statistical characteristics.

The selected descriptors can be classified as follows: constitutional ( $D_8$ ), geometrical ( $D_6$ ,  $D_9$ ), topological ( $D_1$ ,  $D_4$ ,  $D_{10}$ ), solvational ( $D_5$ ), charge-related ( $D_{11}$ ), and quantum-chemical ( $D_2$ ,  $D_3$ ,  $D_7$ ). The respective physicochemical features can be related directly to the investigated property. The topological descriptors involved in the QSPR models include the Kier & Hall index (order 1), the Kier flexibility index, and the Kier shape index (order 2) and are defined for the hydrophobic fragment. They have a reciprocal influence on the cmc that is in accord with the empirical eq 1 and the previous QSPR model described by eq 3 and thus reflect the decrease of cmc with the increase in the number of carbon atoms in the hydrophobic fragment. The Kier shape index (order 2) describes the branching and spatial density of the molecule and has a small positive influence (see Table 4).<sup>44</sup>

All three models developed for the fitted sets include descriptors related to the dipole character of the molecule ( $D_3$ ,  $D_7$ ) and image of the Born solvation energy ( $D_5$ ); these features characterize electrostatic interactions between sur-

Code	Formula	Members of anionic surfactant class
CnSO4Na	$C_nH_{2n+1}OSO_3Na$	Linear alkyl sulfates, sodium salts (total 9) Alkyl sulfate; n = 8, 10, 11, 12, 13, 14, 15, 16, 18
CnC(Cm)SO3Na	$C_nH_{2n+1}CH(C_mH_{2m+1})OSO_3Na$	Branched alkyl sulfates, sodium salts (total 27) n+m+1=10 m=1; n+m+1=14 m=1,2,3,4,6; n+m+1=15 m=1,2,4,7 n+m+1=16 m=3,5,7; n+m+1=17 m=1,8; n+m+1=18 m=1,3,5; m+n+1=19 m=4,9
C15-nC(Cn-1)SO4Na	$C_{15-n}H_{2(15-n)+1}CH(C_{n-1}H_{2(n-1)+1})SO_4Na$	Branched hexadecyl sulfates, sodium salts n = 2, 3, 4, 5, 6, 7, 8
CnSO3Na	$C_nH_{2n+1}SO_3Na$	Linear alkanesulfonates, sodium salts (total 9) n=6, 8, 10, 12, 13, 14, 15, 16, 17
CnC=CSO3Na	$C_nH_{2n+1}C(H)=C(H)SO_3Na$	2-Alkenesulfonate, sodium salts n = 10, 12, 14, 16 (total 4)
CnC(Cm)SO3Na	$C_nH_{2n+1}CH(C_mH_{2m+1})SO_3Na$	Branched alkanesulfonates, sodium salts (total 6) m+n+1=12 m=1, 2, 3, 4, 5 m+n+1=15 m=7
CnPhSO3Na	$C_nH_{2n+1}C_6H_4SO_3Na$	Alkylbenzenesulfonates, sodium salts (total 13) 4-Alkylbenzenesulfonate, n= 7, 8
CnC(Cm)PhSO3Na	$C_nH_{2n+1}C(C_mH_{2m+1})C_6H_4SO_3Na$	4-(Alkyl)benzenesulfonate, m+n+1=9 m=2; m+n+1=10 m=1, 2, 4; m+n+1=11 m=1; m+n+1=12 m=1, 2, 3, 5; m+n+1=13 m=1; m+n+1=15 m=1
CnC(X)C2SO3Na	$C_nH_{2n+1}CH(X)(CH_2)_2SO_3Na$	3-Substituted linear alkanesulfonates, sodium salts (total 28) X = hydroxy (A), methoxy (B), ethoxy (C), propoxy (D), i-propoxy, butoxy (F), hexoxy (G), octoxy (H), 2-ethylhexoxy (I), hydroxyethoxy (J), hydroxyethoxyethoxy (K), phenoxy (L), trichlorophenoxy (M), dimethylamino (N), propylamino (O), butylamino (P), morpholino (Q), piperidino (R), oxo (x). n = 9, X = ABCDEFGHILX; n = 11, X = AJKLMNOPQRX; n = 13, X = ABDFX; n = 15 X = A.
		Other oxygen-containing sulfates and sulfonates, sodium salts (total 23)
C12(EO)nSO4Na	$C_{12}H_{25}(OC_2H_4)_nOSO_3Na$	Dodecyl[di]oxyethyl sulfate, sodium salts n = 1, 2
C10EOSO3Na	$C_{10}H_{21}(OC_2H_4)SO_3Na$	Decyloxyethylsulfonate, sodium salt
CnCO2CSO3Na	$C_nH_{2n+1}OC(O)CH_2SO_3Na$	2-Alkyloxy-2-oxoethanesulfonate, sodium salts n = 6, 8, 10
C14CO2C2SO3Na	$C_{14}H_{29}OC(O)C_2H_4SO_3Na$	3-Oxo-3(tetradecyloxy)propane-1-sulfonate, sodium salt
CnC(COH)SO3Na	$C_nH_{2n+1}CH(SO_3Na)CH_2OH$	1-Hydroxyalkane-2-sulfonate, sodium salts n = 12, 14
CnC(CO2X)SO3Na	$C_nH_{2n+1}CH(SO_3Na)C(O)OX$	1-Alkoxy-1-oxoalkane-2-sulfonate, sodium salts X = Methoxy (S), ethoxy (T), n-propoxy (U), i-propoxy (V). n+1 = 11, X = S; n+1 = 13, X = S; n+1 = 15, X = S, T, U; n+1=17, X=S,T,U,V
CnCO2C(SO3Na)CCO2Cn	$C_nH_{2n+1}OC(O)CH(SO_3Na)CH_2C(O)OC_nH_{2n+1}$	Dialkylsulfosuccinate, sodium salts n= 4, 5, 6, 8 Bis-2-ethylhexylsulfosuccinate, sodium salts (Aerosol OT) R = C <sub>4</sub> H <sub>9</sub> CH <sub>2</sub> (C <sub>2</sub> H <sub>5</sub> )

**Figure 1.** Chemical structures, names, and codes of the old set of anionic surfactants.

factant and water molecule. The plot of image of Born solvation energy against cmc shows a direct proportional relationship, while the plots of the total dipole of the molecule or the total point charge component of the molecular dipole show a reciprocal variance. This indicates that the solubility of the surfactant increases when the charge is distributed throughout the molecule and results in a higher cmc.<sup>45,46</sup>

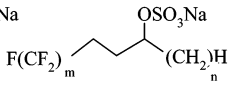
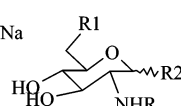
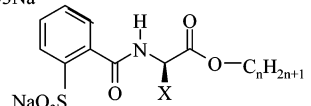
The model of Table 3 incorporates two size-related descriptors defined for hydrophilic heads of the surfactant that suggests the importance of steric interactions between hydrophilic heads in the micelle state and implicitly on critical micelle concentration. The introduction of more than one oxyethylene group into the surfactant molecule leads to the increase of the surface area of the molecule and the steric interactions;<sup>29</sup> therefore, the presence of the size-related descriptors for the hydrophilic heads is appropriate for this data set which includes a significant number of polyethoxylated surfactants (11 structures).

The effect of electrostatic repulsion between hydrophilic heads in the micelle state on the cmc is reflected by the descriptor RNCS (relative negative charged surface area) (D<sub>11</sub>) (Table 4). Both the charge on the head group and the residual charge in the surfactant tail influence the critical micelle concentration.<sup>45,46</sup> Decreasing the charge of the head group increases the solubility of the molecule.<sup>45,46</sup>

The minimum atomic state energy for an O atom (D<sub>2</sub>) describes the hydrophilic domain of the surfactant and its influence on cmc.

In conclusion, the descriptors involved in the developed models (Tables 2–4) account for all possible interactions in the micelle state that occur between surfactant tails and heads on one side and between surfactant and water molecules on the other side.

The independent variables selected in the first step (Tables 2–4) were used to derive the QSPR model for the whole set of compounds (181) using a heuristic algorithm. The

Code	Formula	Members of anionic surfactant class
CnCO2K	$C_nH_{2n+1}C(O)OK$	Linear alkanecarboxylates, potassium salts (total 5) $n=8, 10, 12, 14, 16$
CnSO4Na	$C_nH_{2n+1}SO_4Na$	Linear alkyl sulfates, sodium salts (total 3) $R=C6, 7, 9$
CnC(Cm)SO4Na	$C_nH_{2n+1}CH(C_mH_{2m+1})SO_4Na$	Branched alkyl sulfates, sodium salts (total 6) $m+n+1=8 \ m=1; m+n+1=11 \ m=2, 5; m+n+1=13 \ m=1, 6$ $m+n+1=29 \ m=14$
CnPhSO3Na	$C_nH_{2n+1}C_6H_4SO_3Na$	Linear alkylbenzenesulfonates, sodium salts (total 2) $n = 10, 12$
CnC(Cm)PhSO3Na	$C_nH_{2n+1}CH(C_mH_{2m+1})C_6H_4SO_3Na$	Branched alkylbenzenesulfonates, sodium salts (total 16) 4-(Alkyl)benzenesulfonate $m+n+1=11 \ m=2, 3, 4, 5; m+n+1=12 \ m=4$ $m+n+1=13 \ m=2, 3, 4, 5, 6 \ m+n+1=14 \ m=1, 2, 3, 4, 5, 6$
CnEOmSO4Na	$C_nH_{2n+1}(OC_2H_4)_mOSO_3Na$	Oxygen-containing sulfates and sulfonates Alkyl[m]oxyethyl sulfates, sodium salts (total 6) $n=10, m=2$ $n=12, m=4$ $n=14, m=1, 2, 4$ $n=16, m=2$
CnCO2C2SO3Na	$C_nH_{2n+1}OC(O)(CH_2)_2SO_3Na$	3-Alkyloxy-3-oxopropane-1-sulfonates, sodium salts (total 3) $R=C8, 10, 12$
CFmC2C(Cn)SO4Na		Hybrid hydrocarbon/fluorocarbon sulfates, sodium salts (total 9) (2m+1)-fluoro-(x)-alkyl-(y)yl sulfate $x=(n+m+3), y=(m+3)$ $m=4,6,8 \ n=3, 5, 7$
CC(C)C5PhEO nPO4H2	$CH_3CH_2(CH_2)_5C_6H_4(OCH_2CH_2)_nOP(O)(OH)_2$	4-(6-Methylheptyl)phenoxy n(oxy-1,2ethanediyl) dihydrogen phosphates (total 3) $n=10, 15, 20$
CnC(O)NGlupyrSO4Na		2-Acylamido-2-deoxy-6-O-sulfo-D-glucopyranosides, sodium salts (total 3) $R1=OSO_3Na, R2=OCH_3, R3=C_nH_{2n+1}CO-, n=7,11,15$
Cn(Phe)AlaPhSO3Na		2-(1-R-oxycarbonyl-2S-phenyl-ethylcarbamoyl)-benzene sulfonates, sodium salts (total 6) $X=CH_2Ph, n=5, 8, 10, 12, 14; X=CH_3, n=12$

**Figure 2.** Chemical structures, names, and codes of the new set of anionic surfactants.

variables and statistical characteristics are shown in Table 5, and the plot of experimental and predicted  $-\log \text{cmc}$  is presented in Figure 6.

The order of significance of the independent variables in Table 5 is as follows:  $D_1 > D_5 > D_{10} > D_9 > D_7$  based on the t-test criterion. The fragment and molecular descriptors are topological ( $D_1, D_{10}$ ), geometrical ( $D_9$ ), charge-related ( $D_7$ ), and solvational ( $D_5$ ).

The Kier & Hall index of first order ( $D_1$ ) is calculated according to eq 4

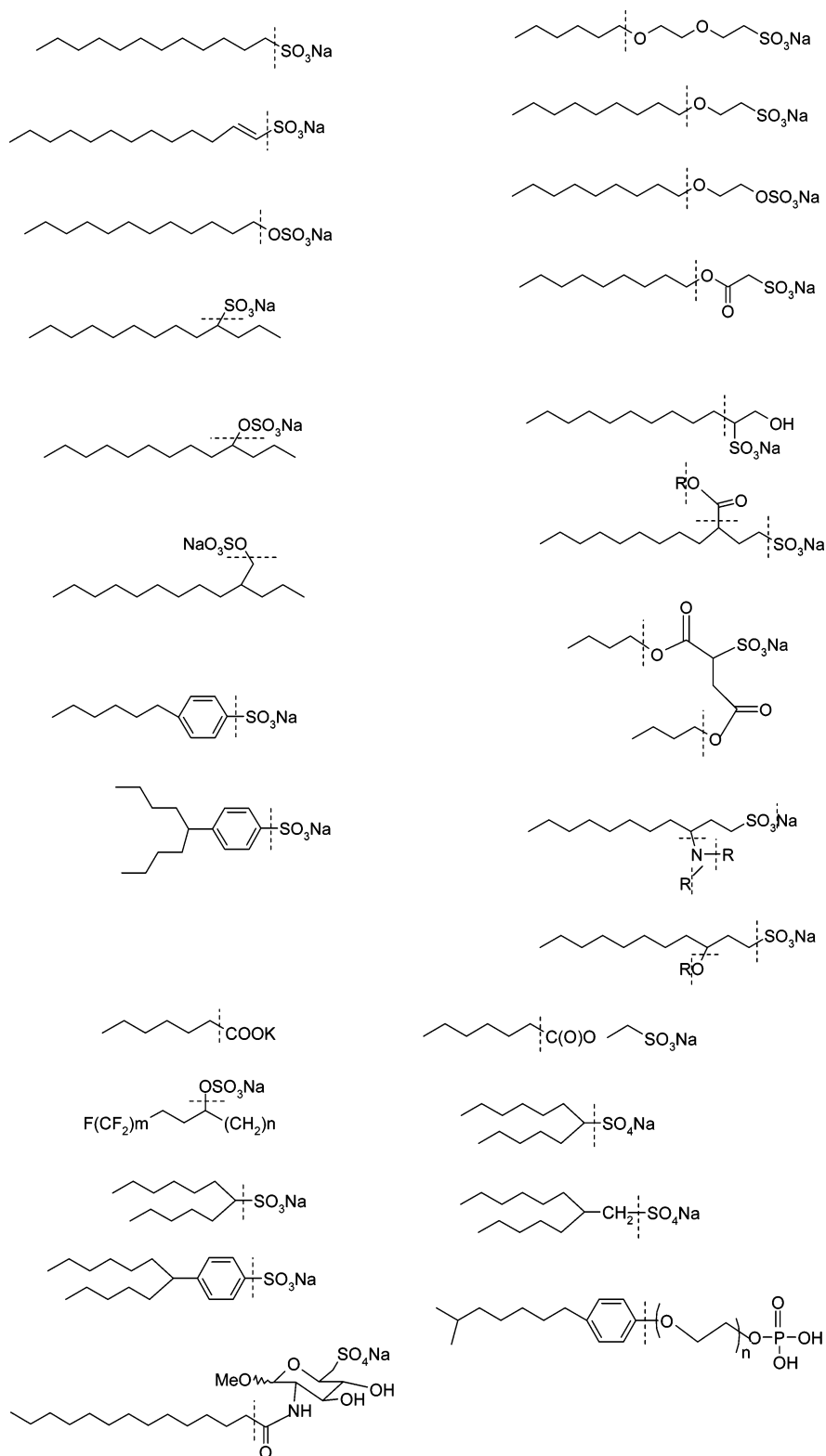
$$KH1 = \sum_{i=1}^N (\delta_i^v)^{1/2} \quad (4)$$

where

$$\delta_i^v = \frac{Z_i^v - H_i}{Z_i - Z_i^v - 1} \quad (5)$$

$Z$  represents the total number of electrons in the  $i$ th atom,  $Z_i^v$  is the number of valence electrons, and  $H_i$  is the number of valence electrons attached directly to the  $i$ th atom, the sum being taken over all the atoms except hydrogen.<sup>44</sup> In the case of the first-order index the sum is defined for all but one bond path. As previously demonstrated by eq 3 this descriptor emphasizes the importance of hydrophobic fragment topology on the critical micelle concentration.

Solvation is accompanied by electronic and orientational polarization of the solvent and dispersion interaction between the solute and solvent molecules. When anionic surfactants dissolve in water, anion-dipole interactions between the surfactant and the water molecule play an important role. An appropriate description of these interactions can be achieved by using the quantum mechanical approach. The



**Figure 3.** The 26 skeletal classes of anionic surfactants; a dashed line indicates the division of surfactants between hydrophobic and hydrophilic fragments.

solvational descriptor image of the Born solvation energy ( $D_5$ ) is defined as follows

$$E_{\text{Born}} = \frac{1 - \epsilon}{2\epsilon} \frac{Q^2}{a_o} \quad (6)$$

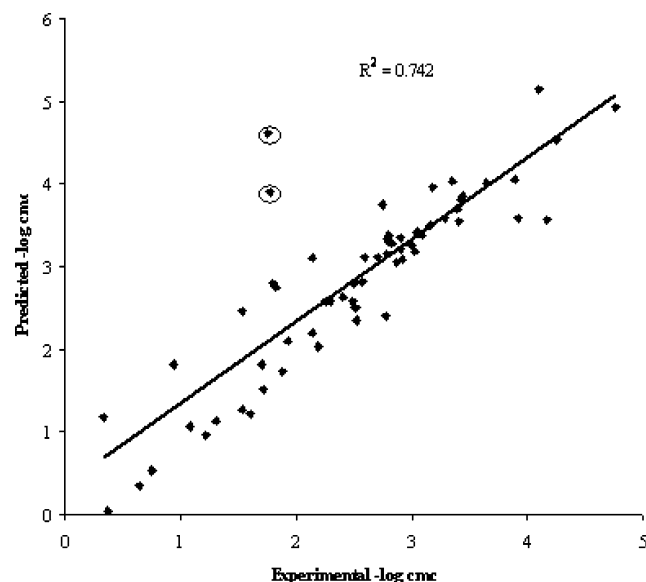
where  $Q$  is the value of the ionic charge,  $\epsilon$  is the dielectric constant of the solvent, and  $a_o$  is the Born radius of the

(molecular) ion. This equation describes the interaction of the ionic charge with the solvent.<sup>44</sup>

Critical micelle concentration is a physical phenomenon that depends on the solvation of surfactant in water, and thus the presence in the equation of Table 5 of the descriptor image of Born solvation energy ( $D_2$ ) is not surprising.

The Kier shape index (order 2) ( $D_{10}$ ) is related to the number of skeletal atoms, molecular branching, and the





**Figure 4.** Scatter plot of the experimental versus predicted (by eq 3)  $-\log \text{cmc}$  for the new set of 62 anionic surfactants.

parameter  $\alpha_i$  that is calculated as the ratio of atomic radius ( $r_i$ ) and the radius of the carbon atom in the  $\text{sp}^3$  hybridization state. This index is defined as

$$^2\kappa = (N_{\text{SA}} + \alpha - 1)(N_{\text{SA}} + \alpha - 2)^2(P + \alpha)^2 \quad (7)$$

where  $P$  is the number of paths of length  $n$  in the molecular skeleton, and  $\alpha$  is the sum of  $\alpha_i$  parameters for all skeletal atoms minus 1. This index therefore encodes the number of atoms in the molecule and the relative degree of cyclicality.

Moment of inertia  $B$  ( $D_9$ ) defined for the hydrophilic domain is derived from atomic coordinates and atomic masses and can be calculated according to the following equation

$$I_B = \sum_i m_i r_{iy}^2 \quad (8)$$

where  $m_i$  are the atomic masses, and  $r_{iy}$  represents the distance of the  $i$ th atomic nucleus from the rotational axis  $y$  of the molecule. This descriptor characterizes the mass distribution in the molecule; the presence of this descriptor in the QSPR model (Table 5) shows that the cmc is dependent on the size of the hydrophilic fragment. This dependence is more obvious in the case of the large size hydrophilic heads of polyethoxylated alkyl sulfates.

The AM1 calculated total point charge component of the molecular dipole is the fifth statistically significant descriptor in the equation of Table 5. This descriptor accounts for the dipolar character of the surfactant. The degree of ground-state charge separation can be related to  $\pi$ -conjugation, strength of donor–acceptor substituents, and solvent polarity.

The outliers for this model are entries 14, 51, 56, 60, 63, 117, 142, 143, 145, and 175 (Table 1). The structures of these outlier molecules have particular features: compound 14 is a long chain alkyl sulfate, 51 is an ethoxylated *p*-isooctylphenol, 56 is 2-acetylamido-2-deoxy-6-O-sulfate D-glucopyranoside, 60 is 2-(1-alkyloxycarbonyl-2S-phenylethylcarbamoyl)-benzene sulfonate, 63 is a short chain alkane sulfonate (6 carbon atoms), 117 is a branched alkyl sulfate,

**Table 2.** Selected Descriptors and Statistical Characteristics of the QSPR Model Derived for the A+B Subset

descriptor	symbol	b	s	t
intercept		−28.076	4.437	−6.328
Kier & Hall index (order 1) <sup>a</sup>	D <sub>1</sub>	−0.395	0.017	−23.266
minimum atomic state energy for a O atom <sup>b</sup>	D <sub>2</sub>	0.093	0.014	6.490
total dipole of the molecule <sup>b</sup>	D <sub>3</sub>	−0.021	0.002	−8.987
Kier flexibility index <sup>a</sup>	D <sub>4</sub>	−0.038	0.008	−4.436
image of the Born solvation energy <sup>b</sup>	D <sub>5</sub>	3.503	0.650	5.388

<sup>a</sup> For hydrophobic fragment. <sup>b</sup> For whole molecule.

**Table 3.** Selected Descriptors and Statistical Characteristics of the QSPR Model Derived for the A+C Subset

descriptor	symbol	b	s	t
intercept		0.389	0.299	1.301
fragment surface area <sup>a</sup>	D <sub>6</sub>	−0.005	0.0004	−12.240
total point charge component of the molecular dipole <sup>b</sup>	D <sub>7</sub>	−0.021	0.003	−6.619
image of the Born solvation energy <sup>b</sup>	D <sub>5</sub>	3.688	0.793	4.650
relative molecular weight <sup>c</sup>	D <sub>8</sub>	−0.039	0.006	−6.349
moment of inertia B <sup>c</sup>	D <sub>9</sub>	155.170	31.788	4.882

<sup>a</sup> For hydrophobic fragment. <sup>b</sup> For whole molecule. <sup>c</sup> For hydrophilic fragment.

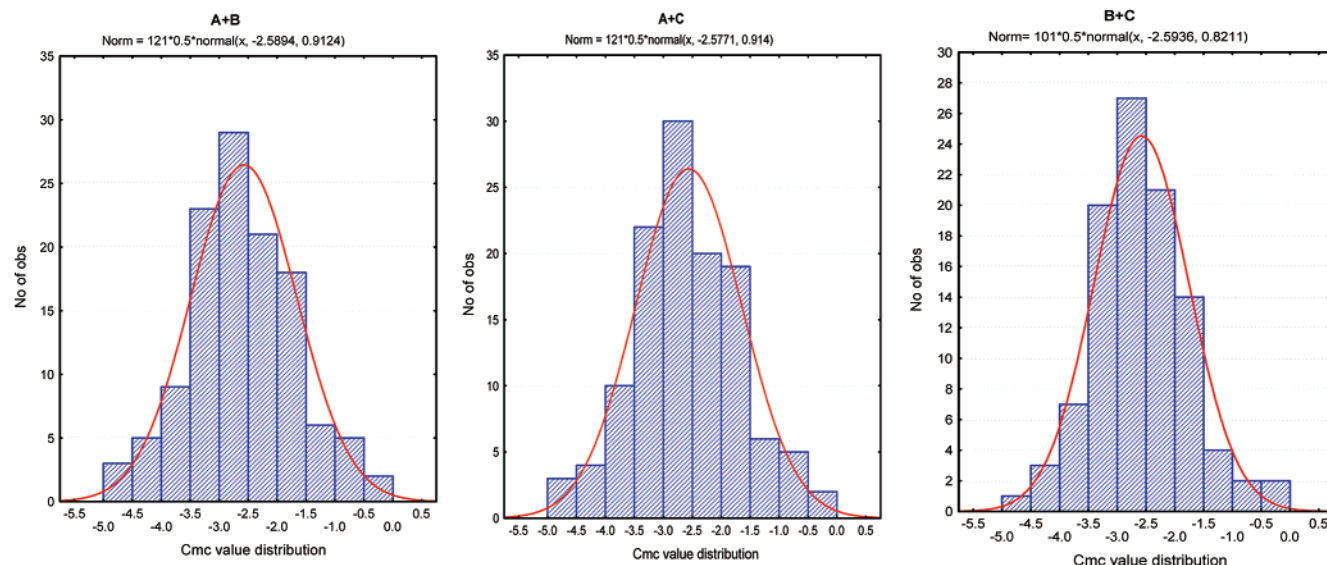
**Table 4.** Selected Descriptors and Statistical Characteristics of the QSPR Model Derived for the B+C Subset

descriptor	symbol	b	s	t
intercept		−0.701	0.195	−3.592
Kier & Hall index (order 1) <sup>a</sup>	D <sub>1</sub>	−0.375	0.018	−21.088
total point charge component of the molecular dipole <sup>b</sup>	D <sub>7</sub>	−0.026	0.002	−11.691
image of the Born solvation energy <sup>b</sup>	D <sub>5</sub>	8.124	0.874	9.296
Kier shape index (order 2) <sup>a</sup>	D <sub>10</sub>	0.001	0.0003	3.510
RNCS relative negative charged surface area <sup>b</sup>	D <sub>11</sub>	0.078	0.014	5.626

<sup>a</sup> For hydrophobic fragment. <sup>b</sup> For whole molecule.

142, 143, 145 are 3-substituted alkanesulfonates, and 175 is 1-oxo-1-propoxyoctadecane-2-sulfonate.

Compound 14 displays particular structural features such as a long alkyl chain (29 carbon atoms) with a sulfate group at position 15; it goes directly into the lamellar phase rather than the curved micelles.<sup>15</sup> Entry 10 from Table 1 has the lowest number of polyoxyethylene units and displays the lowest cmc from the ethoxylated *p*-isooctylphenols. The calculation of  $\Delta G_{\text{mic}}$  for these surfactants showed more negative values when the number of oxyethylene units decreases, indicating a lower steric inhibition of micellization.<sup>19</sup> Compound 56 shows the lowest cmc from the glucopyranoside series 10 times and respectively 100 times lower than the other congener surfactants<sup>10</sup> because this surfactant possesses the longest alkyl substituent; anyway this class of surfactants shows a particular behavior due to the stabilizing H-bonding between amide groups in the micelle state.<sup>10</sup> Compound 60 displays lower solubility and a lower cmc compared to its analogues (e.g., entry 62, Table 1) suggesting a hydrophobic contribution of the benzyl moiety.<sup>2</sup> Surfactant 63 displays a large value of cmc due to the short alkyl chain of six carbon atoms. Surfactant 117

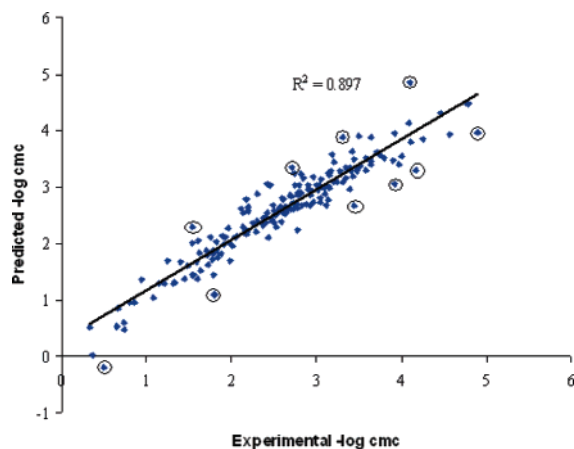


**Figure 5.** Distribution of the experimental log cmc values for the A+B, A+C, and B+C sets.

**Table 5.** Statistical Characteristics and Variables of the QSPR Model for 181 Anionic Surfactants

descriptor	symbol	b	s	t
intercept		-0.777	0.221	-3.511
Kier & Hall index (order 1) <sup>a</sup>	D <sub>1</sub>	-0.382	0.017	-22.522
total point-charge component of the molecular dipole <sup>b</sup>	D <sub>7</sub>	-0.013	0.0026	-5.060
image of the Born solvation energy <sup>b</sup>	D <sub>5</sub>	4.925	0.708	6.952
Kier shape index (order 2) <sup>a</sup>	D <sub>10</sub>	0.002	0.0003	6.618
moment of inertia B <sup>c</sup>	D <sub>9</sub>	129.640	21.594	6.004

<sup>a</sup> For hydrophobic fragment. <sup>b</sup> For whole molecule. <sup>c</sup> For hydrophilic fragment.



**Figure 6.** The plot of experimental versus predicted  $-\log$  cmc for 181 anionic surfactants resulting from the model in Table 5 ( $R^2 = 0.897$ ,  $R^2_{cv} = 0.877$ ,  $F = 303.7115$ ,  $s^2 = 0.087$ ).

belongs to the branched alkyl sulfates class. The plot of cmc against the number of carbon atoms for this type of surfactants displays a spread of cmc when the chain length increases.<sup>47</sup> Compound 142 possesses a phenoxy ring substituent that manifests steric and hydrophobic effects. Derivative 143 shows the highest cmc from its series, and this is probably due to the oxy group that provides improved solubility and an implicitly higher cmc.<sup>13</sup> Compound 145 is the only surfactant that possesses the hydroxyethoxy group that provides solubility and steric effects. Derivative 175

**Table 6.** Statistical Characteristics of the QSPR Models for Training and Predicted Subsets

training set	$R^2$	$s^2$	prediction set	$R^2$	$s^2$
A+B	0.932	0.055	C	0.824	0.152
A+C	0.913	0.076	B	0.878	0.103
B+C	0.919	0.067	A	0.858	0.132
average	0.921	0.066		0.853	0.129

shows the lowest cmc from the whole data set due to the long hydrophobic chain and *n*-propyl group that provide lower steric inhibition of micellization compared to compound 176 that possesses an isopropyl group. In conclusion, the current model has limited prediction ability for the surfactants that display lower or higher values of cmc.

Unlike the previous QSPR model (eq 3),<sup>13</sup> the current models include a geometrical descriptor defined for the hydrophilic fragment, suggesting steric interactions in the micelle phase.

The heuristic algorithm has been rerun omitting the outliers (10 structures). The regression equation involved five descriptors: the Kier & Hall index (order 1); the Kier shape index (order 2) defined for the hydrophobic fragment; moment of inertia B, calculated for the hydrophilic fragment; the total point-charge component of the molecular dipole; and the image of the Born solvation energy defined for the whole molecule. The comparison with the previous equation for 181 surfactants this equation has four identical descriptors, but the dipole of the molecule is replaced by a closely related descriptor of the total point charge component of the molecular dipole. The statistical characteristics are improved  $R^2 = 0.926$ ,  $R^2_{cv} = 0.917$ ,  $F = 411.919$ , and  $s^2 = 0.058$  showing better stability and predictive ability ( $R^2 - R^2_{cv} = 0.009$ ).

**4.2. Validation of the QSPR Models.** The most frequently used technique for the validation of the prediction of QSPR models is the leave-one-out algorithm. The cross-validated correlation coefficient ( $R^2_{cv}$ ) for the models is calculated automatically by the validation module implemented in the CODESSA Pro package. The quality of the selected model is shown by the small difference between the correlation coefficient ( $R^2$ ) of the model and the cross-validated coef-

ficient. According to the value of  $\Delta R^2 = R^2 - R^2_{cv} = 0.897 - 0.877 = 0.020$ , the selected model provides stability and good predictive ability.

The regression equations obtained for the training sets were used to predict the property values for the remaining sets (see Table 6). The averages of the correlation coefficients for the fitted and predicted sets are close to each other, the difference  $\Delta R^2 = R^2_{\text{fitt}} - R^2_{\text{predict}} = 0.068$ . This fact shows the significant prediction ability of the QSPR models obtained.

## 5. CONCLUSIONS

A data set involving diverse chemical classes of anionic surfactants has been investigated to relate log cmc to the molecular structure. QSPR modeling of log cmc was carried out using the CODESSA Pro software. A total number of 1074 descriptors were calculated based on molecular structure: constitutional, geometrical, topological, electrostatic, and quantum chemical. Fragmental approaches provided superior QSPR models in terms of statistical characteristics and predictive ability. The most stable regression equation as well as the highest value of  $R^2$  was obtained in the case of an equation (Table 5) that involves both full molecular and fragment descriptors.

The regression equations obtained provide insights into the structural features of surfactants that influence cmc. The most obvious dependence was manifested by the hydrophobic fragments. The respective interactions were expressed by the geometrical and topological descriptors, while the influence of a hydrophilic fragment was represented by geometrical and constitutional descriptors. Dominant molecular descriptors in the selected QSPR models were solvational, size, and charge-related descriptors as reflecting the intermolecular interactions between anionic surfactants themselves and the anionic surfactant and water.

The comparison between the previous QSPR model<sup>13</sup> and the current model shows that the latter includes one geometrical descriptor defined for the hydrophilic fragment. Hence, the steric interactions between the hydrophilic heads in the micelle phase may also influence cmc.

The prediction of cmc using molecular and fragment descriptors modeled by QSPR is a question of great interest from many practical points of view. The results from the current work provide a further tool for investigation of the cmc phenomenon.

## ACKNOWLEDGMENT

We acknowledge Dr. P. D. T. Huibers for helpful discussions related to the subject of the present paper.

**Supporting Information Available:** Statistical characteristics of the best QSPR model for the training set including molecular descriptors (Table 7). Plot of experimental vs predicted  $-\log$  cmc values for training and test sets according to the model in Table 7. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- Zoller, U. In *Handbook of Detergents Part A: Properties*, Broze, G., Schick, M. J., Hubbard, A., Eds.; Marcel Dekker: New York, 1999; pp 8–60.
- Baczko, K.; Larpent, C.; Lesot, P. New Amino Acid-Based Anionic Surfactants And Their Use As Enantiodiscriminating Lyotropic Liquid Crystalline NMR Solvents. *Tetrahedron: Asymmetry* **2004**, *15*, 971–982.
- Jalali-Heravi, M.; Konouz, E. Prediction of Critical Micelle Concentration of Some Anionic Surfactants Using Multiple Regression Techniques: A Quantitative Structure-Activity Relationship Study. *J. Surfactants Deterg.* **2000**, *3*, 47–52.
- Hoyer, B.; Jensen, N. Stabilization of the Voltammetric Serotonin Signal by Surfactants. *Electrochem. Commun.* **2006**, *8*, 323–328.
- Zhu, Z.; González, Y. I.; Xu, H.; Kaler, E. W.; Liu, S. Polymerization of Anionic Wormlike Micelles. *Langmuir* **2006**, *22*, 949–955.
- Miyazawa, H.; Igawa, K.; Kondo, Y.; Yoshino, N. Synthesis and Solution Properties of Sulfate-Type Hybrid Surfactants with a Benzene Ring. *J. Fluorine Chem.* **2003**, *124*, 189–196.
- Klevens, H. B. Structure and Aggregation in Dilute Solutions of Surface-Active Agents. *J. Am. Oil Chem. Soc.* **1953**, *30*, 74–80.
- Kondo, Y.; Yoshino, N. Hybrid Fluorocarbon/Hydrocarbon Surfactants. *Curr. Opin. Colloid Interface Sci.* **2005**, *10*, 88–93.
- Noordman, W. H.; Blokzijl, W.; Engberts, J. B. F. N. Kinetic Medium Effects of Amphiphilic Cosolutes Below Their Critical Micelle Concentration: The Effect of Sodium n-Alkyl Sulfates on the Neutral Hydrolysis of 1-Benzoyl-1,2,4-triazole. *J. Org. Chem.* **1993**, *58*, 7111–7114.
- Bazito, R. C.; El Seoud, O. A. Sugar-Based Anionic Surfactants: Synthesis and Micelle Formation of Sodium Methyl 2-Acylamido-2-deoxy-6-O-sulfo-D-glucopyranosides. *Carbohydr. Res.* **2001**, *332*, 95–102.
- Stasiuk, E. N. B.; Schramm, L. L. The Temperature Dependence of the Critical Micelle Concentrations of Foam-Forming Surfactants. *J. Colloid Interface Sci.* **1996**, *178*, 324–333.
- Kang, K. H.; Kim, H. U.; Lim, K. H. Effect of Temperature on Critical Micelle Concentration and Thermodynamic Potentials of Micellization of Anionic Ammonium Dodecyl Sulfate and Cationic Octadecyl Trimethyl Ammonium Chloride. *Colloids Surf., A* **2001**, *189*, 113–121.
- Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, O. D.; Karelson, M. Prediction of Critical Micelle Concentration Using a Quantitative Structure-Property Relationship Approach. 2. Anionic Surfactants. *J. Colloid Interface Sci.* **1997**, *187*, 113–120.
- Yuan, S.; Cai, Z.; Xu, G.; Jiang, Y. Quantitative Structure-Property Relationships of Surfactants: Critical Micelle Concentration of Anionic Surfactants. *J. Dispersion Sci. Technol.* **2002**, *23*, 465–472.
- Roberts, D. W. Application of Octanol/Water Partition Coefficients in Surfactant Science: A Quantitative Structure-Property Relationship for Micellization of Anionic Surfactants. *Langmuir* **2002**, *18*, 345–352.
- Li, X.; Zhang, G.; Dong, J.; Zhou, X.; Yan, X.; Luo, M. Estimation of Critical Micelle Concentration of Anionic Surfactants with QSPR Approach. *J. Mol. Struct. (Theorchem)* **2004**, *710*, 119–126.
- Kawakatsu, T.; Kawasaki, K.; Furusaka, M.; Okabayashi, H.; Kanaya, T. Theories and Computer Simulations of Self-Assembling Surfactant Solutions. *J. Phys.: Condens. Matter* **1994**, *6*, 6385–6408.
- Tanaka, A.; Nakamura, K.; Nakanishi, I.; Fujivara, H. A Novel and Useful Descriptor for Hydrophobicity, Partition Coefficient Micellar-Water, and Its Application to a QSAR Study of Antiplatelet Agents. *J. Med. Chem.* **1994**, *37*, 4563–4566.
- De Vries, A. H.; Yefimov, S.; Mark, A. E.; Marrink, S. J. Molecular Structure of the Lecitin Ripple Phase. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5392–5396.
- Knecht, V.; Müller, M.; Bonn, M.; Marrink, S. J.; Mark, A. E. Simulation Studies of Pore and Domain Formation in a Phospholipid Monolayer. *J. Chem. Phys.* **2005**, *122*, 24704.
- MacCallum, J. L.; Tieleman, D. P. Computer Simulation of the Distribution of Hexane in a Lipid Bilayer: Spatially Resolved Free Energy, Entropy, and Enthalpy Profiles. *J. Am. Chem. Soc.* **2006**, *128*, 125–130.
- Vernier, P. T.; Ziegler, M. J.; Sun, Y.; Chang, W. V.; Gundersen, M. A.; Tieleman, D. P. Nanopore Formation and Phosphatidylserine Externalization in a Phospholipid Bilayer at High Transmembrane Potential. *J. Am. Chem. Soc.* **2006**, *128*, 6288–6289.
- Al Sabagh, A. M.; Kandil, N. G.; Badawi, A. M.; El-Sharkawy, H. Surface Activity and Thermodynamic of Micellization and Adsorption for Isooctylphenol Ethoxylates, Phosphate Esters and Their Mixtures with N-Diethoxylated Perfluorooctanamide. *Colloids Surf., A* **2000**, *170*, 127–136.
- Alansson, E. A. G.; Wall, S. N.; Almgren, M.; Hoffmann, H.; Kielmann, I.; Ulbricht, W.; Zana, R.; Lang, J.; Tondre, C. Theory of the Kinetics of Micellar Equilibria and Quantitative Interpretation of Chemical Relaxation Studies of Micellar Solutions of Ionic Surfactants. *J. Phys. Chem.* **1976**, *80*, 905–922.



- (25) Nakahara, Y.; Kida, T.; Nakatsuji, Y.; Akashi, M. New Fluorescence Method for the Determination of the Critical Micelle Concentration by Photosensitive Monoazacryptand Derivatives. *Langmuir* **2005**, *21*, 6688–6695.
- (26) Rosen, M. J.; Zhu, Y.-P.; Morrall, S. W. Effect of Hard River Water on the Surface Properties of Surfactants. *J. Chem. Eng. Data* **1996**, *41*, 1160–1167.
- (27) Tsujii, K. In *Surface Activity: Principles, Phenomena and Applications*; Tanaka, T., Ed.; Academic Press: San Diego, 1998; p 84.
- (28) Kissa, E. In *Fluorinated Surfactants and Repellents*, 2nd ed.; Schick, M. J., Hubbard, A. T., Eds.; Marcel Dekker: New York, 2001; Chapter 7, pp 220–247.
- (29) Dahanayake, M.; Cohen, A. W.; Rosen, M. J. Relationship of Structure to Properties of Surfactants. 13. Surface and Thermodynamic Properties of Some Oxyethylenated Sulfates and Sulfonates. *J. Phys. Chem.* **1986**, *90*, 2413–2418.
- (30) *HyperChem, version 7.5*; HyperCube, Inc.: Gainesville, FL, 2003. <http://www.hyper.com> (accessed Jan 2006).
- (31) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (32) Katritzky, A. R.; Karelson, M.; Petrukin, R. *CODESSA PRO*; University of Florida, Gainesville, FL, 2005. <http://www.codessa-pro.com> (accessed Feb 2006).
- (33) Katritzky, A. R.; Mu, L.; Lobanov, V.; Karelson, M. Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. *J. Phys. Chem.* **1996**, *100*, 10400–10407.
- (34) Katritzky, A. R.; Tamm, K.; Kuanar, M.; Fara, D. C.; Oliferenko, A.; Oliferenko, P.; Huddleston, J. G.; Rogers, R. D. Aqueous Biphasic Systems. Partitioning of Organic Molecules: A QSPR Treatment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 136–142.
- (35) Katritzky, A. R.; Fara, D. C.; Kuanar, M.; Hür, E.; Karelson, M. The Classification of Solvents by Combining Classical QSPR Methodology with Principal Component Analysis. *J. Phys. Chem. A* **2005**, *109*, 10323–10341.
- (36) Lučić, B.; Bašić, I.; Nadramija, D.; Miličević, A.; Trinajstić, N.; Suzuki, T.; Petrukhin, R.; Karelson, M.; Katritzky, A. R. Correlation of Liquid Viscosity with Molecular Structure for Organic Compounds Using Different Variable Selection Methods. *ARKIVOC* **2002**, 45–49.
- (37) Katritzky, A. R.; Fara, D. C.; Yang, H.; Karelson, M.; Suzuki, T.; Solov'ev, V. P.; Varnek, A. Quantitative Structure-Property Relationship Modeling of  $\beta$ -Cyclodextrin Complexation Free Energies. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 529–541.
- (38) Katritzky, A. R.; Fara, D. C.; Karelson, M. QSPR of 3-Aryloxazolidin-2-one Antibacterials. *Bioorg. Med. Chem.* **2004**, *12*, 3027–3035.
- (39) Katritzky, A. R.; Dobchev, D. A.; Hür, E.; Fara, D. C.; Karelson, M. QSAR Treatment of Drugs Transfer Into Human Breast Milk. *M. Bioorg. Med. Chem.* **2005**, *13*, 1623–1632.
- (40) Katritzky, A. R.; Oliferenko, A.; Lomaka, A.; Karelson, M. Six-Membered Cyclic Ureas as HIV-1 Protease Inhibitors: a QSAR Study Based on CODESSA PRO Approach. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 3453–3457.
- (41) Beteringhe, A.; Balaban, A. T. QSAR For Toxicities of Polychlorodibenzofurans, Polychlorodibenzo-1,4-Dioxins, and Polychlorobiphenyls. *ARKIVOC* **2004**, 163–182.
- (42) Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. Prediction of Critical Micelle Concentration Using a Quantitative Structure-Property Relationship Approach. 1. Nonionic Surfactants. *Langmuir* **1996**, *12*, 1462–1470.
- (43) Katritzky, A. R.; Slavov, S. H.; Dobchev, D. A.; Karelson, M. Rapid QSPR Model Development Technique for Prediction of Vapor Pressure of Organic Compounds. Submitted to *Comput. Chem. Eng.*
- (44) Karelson, M. In *Molecular Descriptors in QSAR/QSPR*; John Wiley & Sons: New York, 2000; pp 143–187, 221–237.
- (45) Huibers, P. D. T.; Jacobs, P. T. The Effect of Polar Head Charge Delocalization on Micellar Aggregation Numbers of Decylpyridinium Salts, Revisited. *J. Colloid Interface Sci.* **1998**, *206*, 342–345.
- (46) Huibers, P. D. T. Quantum-Chemical Calculations of the Charge Distribution in Ionic Surfactants. *Langmuir* **1999**, *15*, 7546–7550.
- (47) Evans, H. C. Alkyl Sulfates. I. Critical Micelle Concentrations of the Sodium Salts. *J. Chem. Soc.* **1956**, *80*, 579–586.

CI600462D