

Prediction of Critical Micelle Concentration Using a Quantitative Structure–Property Relationship Approach

2. Anionic Surfactants

P. D. T. HUIBERS,* V. S. LOBANOV,§ A. R. KATRITZKY,§¹ D. O. SHAH,* AND M. KARELSON‡

*Department of Chemical Engineering, Center for Surface Science and Engineering, University of Florida, Gainesville, Florida 32611-6005;

§Department of Chemistry, Center for Heterocyclic Chemistry, University of Florida, Gainesville, Florida 32611-7200;

and ‡Department of Chemistry, University of Tartu, 2 Jakobi Strasse, Tartu EE 2400, Estonia

Received June 10, 1996; accepted November 5, 1996

Relationships between the molecular structure and the critical micelle concentration (cmc) of anionic surfactants were investigated using a quantitative structure–property relationship approach. Measured cmc values for 119 anionic structures, representing sodium alkyl sulfates and sodium sulfonates with a wide variety of hydrophobic and hydrophilic structures, were considered. The best multiple linear regression model involved three terms (descriptors) and had a correlation coefficient of $R^2 = 0.940$. Very good correlations ($R^2 = 0.988$) were obtained using three descriptors for a subset of 68 structures, with structural variation only in the hydrophobic domain. From the descriptors used in these regressions, one can conclude that the cmc is primarily dependent on the size (volume or surface area) of the hydrophobic domain and to a lesser extent on the structural complexity of the surfactant molecule. © 1997 Academic Press

Key Words: critical micelle concentration; anionic surfactants; quantitative structure–property relationships; molecular connectivity; dipole moment; topology; property prediction; MOPAC; CODESSA.

INTRODUCTION

The critical micelle concentration (cmc) is the single most useful quantity for characterizing surfactants. This parameter alone captures much of the surface activity of the molecule. The cmc, above which micelles are present in solution, is also key to the industrial use or biological activity of the surfactant. Many applications of surfactants, such as detergency, require that surfactant molecules organize into micelles in order to facilitate the desired process. Qualitatively it is well known that contributions from both the size of the hydrophobic domain (tail) and the size of the hydrophilic

domain (head) of the surfactant influence the cmc. The two contributions are counteracting, with a lower cmc for a larger hydrophobic domain and a higher cmc for a larger hydrophilic domain. The current study attempts to define quantitative measures for these two counteracting contributions that will apply over a wide range of anionic surfactant structures.

Previous correlations of the cmc. Linear relationships between the logarithm of the cmc and the size of a homologous series of surfactants have been known for decades. Examples of such dependencies as a function of alkane carbon number are summarized in Table 1. The limitation of applicability of these relationships is that the coefficients must be recalculated for each homologous series. More general relationships would be of value in establishing specific quantitative aspects of molecular structure that influence cmc, as well as allowing prediction of cmc for molecules not yet synthesized.

A previous quantitative structure–property relationship (QSPR) study of the cmc of nonionic surfactants (1) produced a multiple regression (Eq. [1]) between log cmc and three descriptors based on molecular topology and constitution.

$$\begin{aligned} \log_{10}\text{cmc} = & -1.80 - 0.567t\text{-KH0} \\ & + 1.054t\text{-ASIC2} + 7.51\text{RNNO} \\ R^2 = & 0.983, F = 1433, s^2 = 0.0313, N = 77 \quad [1] \end{aligned}$$

In this regression, $t\text{-KH0}$ is the zeroth-order Kier and Hall molecular connectivity index for the hydrophobic fragment (surfactant tail), which correlates highly with both molecular volume ($r = 0.979$) and surface area ($r = 0.971$). $t\text{-ASIC2}$ is the second-order average information content in-

¹ To whom correspondence should be addressed. E-mail: katritzk@chem.ufl.edu.

TABLE 1

Constants for the Relationship between cmc (mol/L) and Carbon Number: $\log_{10}\text{cmc} = A - B(\text{C}\#)$

Surfactant series (Na ⁺ alkyl sulf[on]ates)	Temp (°C)	A	B	Ref.
C _n SO ₃ , C _n SO ₄	25	1.51	0.30	(14)
C _n SO ₃	40	1.59	0.29	(15)
C _n SO ₃	55	1.15	0.26	(16)
C _n SO ₄	45	1.42	0.30	(15)
C _n SO ₄	60	1.35	0.28	(14)
2-C _n SO ₄	55	1.28	0.27	(16)
C _n (C ₆ H ₄)SO ₃	55	1.68	0.28	(16)
C _n (C ₆ H ₄)SO ₃	70	1.33	0.27	(17)
C _n SO ₃	40	1.44 ± 0.14	0.293 ± 0.013	This study
C _n SO ₄	40	1.46 ± 0.02	0.293 ± 0.002	This study

dex for the hydrophobic fragment, which captures some of the information on the complexity (branching and unsaturation) of the hydrophobic tail. Finally, RNN0 is the relative number of nitrogen and oxygen atoms, representing the contribution of the hydrophilic head group. This study established that surfactant properties could be predicted, based only on molecular structure, for sets of structures much larger and diverse than those considered in the regressions of Table 1. It was also established that a QSPR methodology could be applied to surfactants, using molecular descriptors calculated for fragments of the molecule (in this case either the hydrophobic or the hydrophilic domain) rather than the traditional approach of using descriptors calculated for the entire molecule.

Consideration of the nature of the descriptors in the correlation model (Eq. [1]) gave insight into the molecular features that determine cmc. The dominant descriptor of the three was *t*-KH0, which correlates highly with the hydrocarbon fragment surface area. The process of micellization is driven by both enthalpic and entropic considerations. In this process, the changes in enthalpy and entropy of a surfactant molecule must be considered, but the dominant changes are due to the nature of water self-association. Micelles form not because the surfactant tails attract each other, but because the water molecules highly favor self-association over association with any solute that cannot engage in dipole or hydrogen bonding interactions. The nature of this water self-association via hydrogen bonding can explain qualitatively the influence of molecular surface area on the cmc, as solutes such as the hydrophobic tails of surfactants force the water to form a hydrogen bonding "cage" around the solute, and the extent of the distorted hydrogen bonds (enthalpically disfavored) and the forced order of the water around the solute (entropically disfavored) are both expected to be proportional to the surface area of the solute. Thus, for the micellization process, where the influences of the hydropho-

bic tail outweigh those of the hydrophilic head group, it is the surface area of the tail that makes the dominant contribution to cmc.

DATA AND METHODOLOGY

Data sources. The cmc values for the set of 119 anionic surfactant structures (Table 2) were taken from the compilations by van Os (2), Rosen (3), and Mukerjee and Mysels (4). All cmc values were measured at 40°C, for the sodium salts of the surfactants. A wide variety of surfactant structures were included (Fig. 1). The diverse hydrophobic fragments included linear, branched, phenyl substituted, and unsaturated hydrocarbon structures. Double-tailed surfactants were also considered, together with both dialkyl sodium sulfosuccinates and 3-substituted alkyl sodium sulfonates. The hydrophilic groups included sulfonate and sulfate polar head groups, in addition to various ester, ether, and amine linkages.

The dividing line between the two fragments can best be defined by considering that the hydrophobic group contains a contiguous carbon backbone and associated hydrogen atoms, while the hydrophilic group contains all of the heteroatoms that participate in hydrogen bonding or dipole interactions (oxygen, nitrogen, sulfur). Small carbon fragments of one or two atoms, such as in ethylene oxide polymers, are considered part of the hydrophilic fragment, as well as carbons that are bonded to two oxygen atoms, such as in an ester structure.

Temperature dependence of cmc. Correlations were performed on cmc data at 40°C. The majority of cmc values in the literature have been measured at either 40 or 25°C. A number of structures, especially those with larger hydrophobic domains, have Krafft points between 25 and 40°C. For these structures, micelles would not be formed at 25°C and thus the cmc would be meaningless. For structures for which data were available at 25 but not 40°C, scaling rules were used to extrapolate to 40°C. The temperature dependence of cmc of anionic surfactants has been shown to be parabolic in a number of cases, with a minimum cmc at approximately 25°C (2, 5, 6). The minimum temperature and the shape of the parabola appear to depend on the size of the surfactant, but this relationship is not clear. Using cmc data from research in which measurements were taken at least at four temperatures, it was established that the ratio between cmc at 40 and that at 25°C is approximately constant. For sulfonates and sulfates, this ratio is 1.088 ± 0.030 and 1.030 ± 0.030 , respectively. The magnitude of this temperature correction is small (3–9%), probably smaller than the experimental error in many cases (>10%). The cmc values used in this study are summarized in Table 2, with values scaled from

TABLE 2
Critical Micelle Concentrations (mol/L) for 119 Anionic Surfactants

Structure name	Literature log(cmc) (40°C)	Estimated log(cmc) Eq. [2]	Estimated log(cmc) Eq. [3]	Ref.	Structure name	Literature log(cmc) (40°C)	Estimated log(cmc) Eq. [2]	Estimated log(cmc) Eq. [3]	Ref.
c06so3	-0.496	-0.063	-0.378	(18)	10c19so4	-3.027	-3.305	-3.135	(30)
c08so3	-0.796	-0.678	-0.953	(19)	5c19so4	-3.481	-3.529	-3.616	(30)
c10so3	-1.398	-1.273	-1.527	(20)	2c15cso4	-3.097	-3.061	-3.035	(33)
c12so3	-1.959	-1.886	-2.102	(20)	3c15cso4	-3.046	-2.943	-2.927	(33)
c13so3 ^a	-2.421	-2.195	-2.390	(21)	4c15cso4	-2.959	-2.859	-2.831	(33)
c14so3	-2.602	-2.505	-2.676	(19)	5c15cso4	-2.824	-2.782	-2.735	(33)
c15so3 ^a	-3.139	-2.813	-2.964	(21)	6c16cso4	-2.699	-2.728	-2.638	(33)
c16so3	-3.131	-3.125	-3.250	(22)	7c15cso4	-2.638	-2.677	-2.542	(33)
c17so3 ^a	-3.635	-3.435	-3.539	(21)	8c15cso4	-2.523	-2.672	-2.446	(33)
c122zso3	-1.886	-1.752	-1.909	(22)	c12e1so4 ^a	-2.396	-2.534		(20)
c142zso3	-2.569	-2.366	-2.483	(22)	c12e2so4	-2.553	-2.850		(20)
c162zso3	-3.215	-2.985	-3.058	(22)	c123aso3	-1.606	-1.795		(22)
c182zso3	-3.745	-3.605	-3.632	(22)	c123bso3 ^a	-2.118	-2.129		(36)
2c12so3	-1.827	-1.815	-1.817	(23)	c123cso3 ^a	-2.301	-2.339		(36)
3c12so3	-1.730	-1.710	-1.718	(23)	c123dso3 ^a	-2.420	-2.564		(36)
4c12so3	-1.635	-1.632	-1.622	(23)	c123eso3 ^a	-2.458	-2.602		(36)
5c12so3	-1.548	-1.575	-1.526	(23)	c123fso3 ^a	-2.817	-2.796		(36)
6c12so3	-1.442	-1.539	-1.430	(23)	c123gso3 ^a	-3.185	-3.266		(36)
8c15so3 ^a	-2.144	-2.318	-2.095	(24)	c123hso3 ^a	-3.922	-3.753		(36)
c07bso3	-1.582	-1.650	-1.607	(25)	c123iso3 ^a	-3.501	-3.761		(36)
c08bso3	-1.907	-1.950	-1.889	(26)	c123lso3 ^a	-2.708	-2.770		(36)
3c09bso3	-1.967	-2.155	-1.929	(6)	c123xso3	-1.541	-1.794		(22)
2c10bso3	-2.303	-2.501	-2.294	(5)	c143aso3	-2.199	-2.338		(22)
3c10bso3	-2.200	-2.437	-2.209	(5)	c142jso3 ^a	-3.445	-2.693		(36)
5c10bso3	-2.047	-2.350	-2.017	(5)	c143kso3 ^a	-2.922	-2.744		(36)
2c11bso3	-2.721	-2.804	-2.575	(27)	c143lso3 ^a	-3.641	-3.388		(36)
2c12bso3 ^a	-2.692	-3.099	-2.857	(16)	c143mso3 ^a	-4.787	-4.240		(36)
3c12bso3	-2.606	-3.018	-2.771	(6)	c143nso3 ^a	-2.964	-2.949		(37)
4c12bso3 ^a	-2.721	-2.955	-2.675	(28)	c143oso3 ^a	-3.200	-3.099		(37)
6c12bso3	-2.585	-2.868	-2.482	(29)	c143pso3 ^a	-3.708	-3.331		(37)
2c13bso3	-3.208	-3.397	-3.138	(27)	c143qso3 ^a	-3.106	-2.798		(37)
2c15bso3	-3.577	-3.992	-3.703	(27)	c143rso3 ^a	-3.310	-3.442		(37)
					c143xso3	-2.174	-2.405		(22)
c08so4	-0.854	-0.782	-0.931	(30)	c163aso3	-2.839	-2.938		(22)
c10so4	-1.481	-1.396	-1.505	(30)	c163bso3 ^a	-3.472	-3.231		(36)
c11so4 ^a	-1.783	-1.718	-1.791	(31)	c153dso3 ^a	-4.089	-3.651		(36)
c12so4	-2.066	-2.001	-2.080	(32)	c163fso3 ^a	-4.458	-3.867		(36)
c13so4	-2.367	-2.308	-2.365	(33)	c163xso3	-2.735	-3.020		(22)
c14so4	-2.658	-2.610	-2.654	(34)	c183aso3	-3.420	-3.631		(22)
c15so4	-2.921	-2.927	-2.940	(33)					
c16so4	-3.237	-3.238	-3.228	(30)	c10eso3 ^a	-1.787	-1.810		(20)
c18so4	-3.796	-3.856	-3.803	(35)	c12ohso3 ^a	-1.787	-1.572		(38)
2c10so4 ^a	-1.328	-1.257	-1.332	(32)	c14ohso3 ^a	-2.432	-2.152		(38)
2c14so4	-2.481	-2.449	-2.481	(30)	c06aso3 ^a	-0.733	-0.629		(39)
3c14so4	-2.367	-2.324	-2.374	(30)	c08aso3 ^a	-1.144	-1.229		(39)
4c14so4	-2.288	-2.218	-2.278	(30)	c10aso3 ^a	-1.621	-1.834		(39)
5c14so4	-2.171	-2.141	-2.181	(30)	c14pso3	-3.046	-3.105		(40)
7c14so4	-2.013	-2.053	-1.989	(30)	c111sso3 ^a	-1.986	-2.290		(41)
2c15so4	-2.767	-2.737	-2.770	(30)	c131sso3 ^a	-2.583	-2.872		(41)
3c15so4	-2.658	-2.619	-2.660	(30)	c151sso3 ^a	-3.398	-3.474		(42)
5c15so4	-2.469	-2.421	-2.467	(30)	c151tso3 ^a	-3.509	-3.680		(42)
8c15so4	-2.177	-2.298	-2.179	(30)	c151uso3 ^a	-3.964	-3.881		(42)
4c16so4	-2.764	-2.797	-2.852	(30)	c171sso3 ^a	-4.000	-4.071		(42)
6c16so4	-2.629	-2.590	-2.659	(30)	c171tso3 ^a	-4.106	-4.268		(42)
8c16so4	-2.372	-2.541	-2.467	(30)	c171uso3 ^a	-4.899	-4.475		(42)
2c17so4	-3.310	-3.292	-3.344	(30)	c171vso3 ^a	-4.569	-4.504		(42)
9c17so4	-2.629	-2.771	-2.657	(30)	aot ^a	-2.566	-3.114		(43)
2c18so4	-3.585	-3.595	-3.630	(30)	c4c4succ ^a	-0.663	-1.131		(43)
4c18so4	-3.347	-3.365	-3.426	(30)	c5c5succ ^a	-1.239	-1.658		(43)
6c18so4	-3.143	-3.158	-3.233	(30)	c6c6succ ^a	-1.817	-2.145		(39)
					c8c8succ ^a	-3.131	-3.135		(39)

Note. See Table 3 for structural formulas.

^a cmc values at 40°C estimated from values at 25°C (see text).

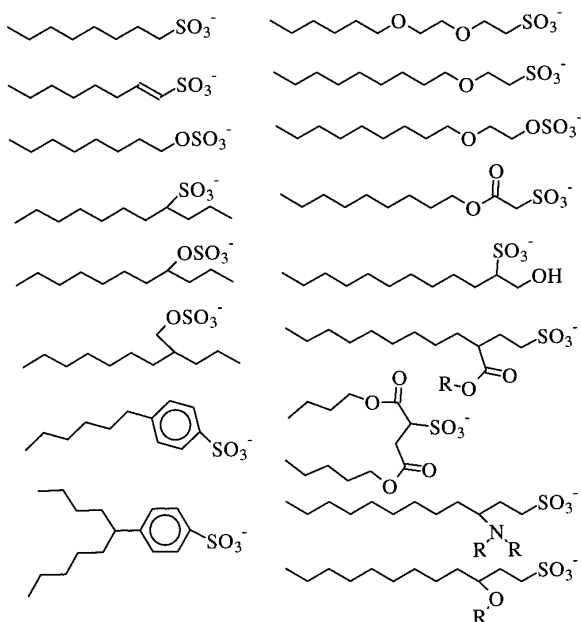


FIG. 1. Representative structures of the anionic surfactants, showing the diversity of the hydrophobic and hydrophilic domains.

25°C marked with an asterisk. The naming scheme for the surfactant structures is summarized in Table 3.

Computational methods. The quantitative structure–property relationships were developed using CODESSA (7), QSPR software based on the Microsoft Windows environment. This program performs the calculation of descriptors, statistical analysis using several multilinear regression techniques, and property prediction. The methodology of searching for the best statistical relationships between calculated descriptors and experimental property data is described in previous papers (1, 8). Briefly, the three-dimensional molecular structures of the surfactant molecules were drawn and preoptimized using a molecular–mechanics-based program (9). The structures were submitted for one SCF calculation using MOPAC 6.0 (10) to generate quantum–chemical wave function parameters using the AM1 Hamiltonian. The MOPAC output files were supplied to CODESSA to calculate five types of molecular descriptors: constitutional, topological, geometrical, electrostatic, and quantum–chemical (8). The CODESSA program has been modified for automated calculation of hydrophobic and hydrophilic fragment descriptors, shown to be successful in our previous study of surfactant cmc correlation (1). Due to the presence of structures with multiple hydrophobic and/or hydrophilic fragments, the sums of corresponding descriptors over all fragments from either the hydrophobic or the hydrophilic domain were calculated. Both the entire molecule descriptors and the fragment descriptors were used in correlation analysis.

RESULTS AND DISCUSSION

General model for anionic surfactants. In the search for quantitative structure–property relationships using a large set of molecular descriptors, there is no certainty that the descriptors chosen to construct the statistically best regression will make clear physical sense with regard to the effect of molecular structure on the property of interest. Using the approach outlined previously (1) an optimal multilinear relationship was developed (Eq. [2]) for predicting cmc based on the diverse set of 119 anionic surfactants (Fig. 2). The correlation equation contains quantitative contributions from both the hydrophobic and the hydrophilic domains of the surfactant molecules, as well as a descriptor for the entire molecule. The nature of the selected descriptors is in agreement with what is known qualitatively about structural effects on cmc. Finding a good correlation with only three descriptors can be considered a success. Although correlation coefficients can always be improved by the addition of more terms, a point of diminishing returns is soon reached, where the additional descriptors start to fit experimental error and contribute nothing to the physical understanding of the property.

$$\begin{aligned} \log_{10}\text{cmc} = & (1.89 \pm 0.11) \\ & - (0.314 \pm 0.010)t\text{-sum-KH0} \\ & - (0.034 \pm 0.003)\text{TDIP} \\ & - (1.45 \pm 0.18)h\text{-sum-RNC} \end{aligned}$$

$$R^2 = 0.940, F = 597, s^2 = 0.0472, N = 119 \quad [2]$$

The most significant descriptor in the above correlation is *t*-sum-KH0, the sum of Kier and Hall molecular connectivity indices of zeroth order (11, 12) over all hydrophobic fragments (surfactant tails). Although this descriptor correlates highly with both molecular volume and surface area, it performs better than either of these two other parameters. It is also notable that the same descriptor is most significant in the correlation model reported for the nonionic surfactants (1).

The second descriptor, the total dipole of the molecule (TDIP), is calculated by MOPAC from the quantum chemical charge distribution in the molecule (10). Analysis of the variation of the total dipole moments in the data set revealed that for the linear surfactants with the head group attached to the first carbon, the dipole moment increases with the size of the tail (Fig. 3). For molecules of a given size, the dipole moment decreases as the head group moves toward the center of the molecule. This second dependency has less influence on the cmc. Finally, for the double-tailed structures, the dipole moment is primarily determined by the longest tail. These dependencies can be used for the prediction of

TABLE 3
Structural Formulas for 119 Anionic Surfactants by Class

Name	Formula	Members of anionic surfactant class
Linear alkylsulfates (total 9)		
c##so4	$R\text{-OSO}_3$	Alkylsulfate; $R = \text{C8, 10, 11, 12, 13, 14, 15, 16, 18}$
Branched alkylsulfates (total 27)		
#c##so4	$\text{C}_x\text{CH}(\text{C}_y)\text{-OSO}_3$	#-Alkylsulfate, $x + y + 1 = \##$ #C## = 2C10, 2C14, 2C15, 2C17, 2C18, 3C14, 3C15, 4C14, 4C16, 4C18, 5C14, 5C15, 5C19, 6C16, 6C18, 7C14, 8C15, 8C16, 9C17, 10C19
#c15cso4	$\text{C}_{15-n}\text{CH}(\text{C}_{n-1})\text{CH}_2\text{-OSO}_3$	Pentadecyl-#-methylsulfate, # = 2, 3, 4, 5, 6, 7, 8
Linear alkylsulfonates (total 9)		
c##so3	$R\text{-SO}_3$	Alkylsulfonate, $R = \text{C6, 8, 10, 12, 13, 14, 15, 16, 17}$
c##2zso3		2-Alkenesulfonate, ## = 12, 14, 16, 18
Branched alkylsulfonates (total 6)		
#c##so3	$\text{C}_x\text{CH}(\text{C}_y)\text{-SO}_3$	#-Alkylsulfonate, $x + y + 1 = \##$ #C## = 2C12, 3C12, 4C12, 5C12, 6C12, 8C15
Alkylbenzenesulfonates (total 13)		
c##bso3	$R\text{-C}_6\text{H}_4\text{-SO}_3$	4-Alkylbenzenesulfonate, $R = \text{C7, 8}$
#c##bso3	$R\text{-C}_6\text{H}_4\text{-SO}_3$	4-(#-Alkyl)benzenesulfonate, $R = \text{2C10, 2C11, 2C12, 2C13, 2C15, 3C9, 3C10, 3C12, 4C12, 5C10, 6C12}$
3-Substituted linear alkylsulfonates (total 24)		
c##3Xso3	$\text{C}_{\##-3}\text{C}(\text{X})\text{C}_2\text{-SO}_3$	3-(X)-Alkylsulfonate $X = \text{hydroxy (A), methoxy (B), ethoxy (C), propoxy (D), } i\text{-propoxy (E), butoxy (F), hexoxy (G), octoxy (H), 2-ethylhexoxy (I), hydroxyethoxy (J), hydroxyethoxyethoxy (K), phenoxy (L), trichlorophenoxy (M), dimethylamino (N), propylamino (O), butylamino (P), morpholino (Q), piperidino (R), oxo(X). ## = 12, X = ABCDEFGHILX; ## = 14, X = AJKLMNOPQRX; ## = 16, X = ABDFX; ## = 18, X = A}$
Other oxygen-containing sulfates and sulfonates (total 23)		
c12e#so4	$\text{C}_{12}\text{-(OC}_2\text{H}_4)\#\text{-OSO}_3$	Dodecyl[di]oxyethylsulfate, # = 1, 2
c10eso3	$\text{C}_{10}\text{-(OC}_2\text{H}_4)\text{-SO}_3$	Decyloxyethylsulfonate
c##aso3	$R\text{-OC(O)-CH}_2\text{-SO}_3$	Alkyl acetate 2-sulfonate, ## = 6, 8, 10
c14pso3	$\text{C}_{14}\text{-OC(O)-C}_2\text{H}_4\text{-SO}_3$	Tetradecyl propionate 3-sulfonate
c##ohso3	$R\text{-CH(SO}_3)\text{-CH}_2\text{OH}$	1-Hydroxy-2-alkylsulfonate, ## = 12, 14
c##1Xso3	$R\text{-(SO}_3)\text{-C(O)O-X}$	1-Carboalkoxyalkylsulfonate, $X = \text{Methoxy (S), ethoxy (T), } n\text{-propoxy (U), } i\text{-propoxy (V). ## = 11, X = S; ## = 13, X = S; ## = 15, X = S, T, U; ## = 17, X = S, T, U, V}$
c#c#succ	$R\text{-OC(O)CH(SO}_3)\text{CH}_2\text{C(O)O-R}$	Dialkylsulfosuccinate, $R = \text{C4, 5, 6, 8}$
aot		Bis-2-ethylhexylsulfosuccinate (Aerosol OT), $R = i\text{-C8}$

cmc, similar to the cmc dependence on carbon number. Table 4 summarizes coefficients for certain classes of anionic surfactants that can be used to predict cmc from dipole moment. Although it appears that TDIP accounts for similar information concerning the size of the molecule as the KH0 term, the two terms are actually rather poorly correlated ($r = 0.40$)

and thus account for significantly different structural aspects of the molecule.

The third descriptor, h -sum-RNC, is the sum of the relative numbers of carbon atoms over all hydrophilic fragments (heads). This parameter accounts for the diversity of the head group structures, i.e., when a sulfonate

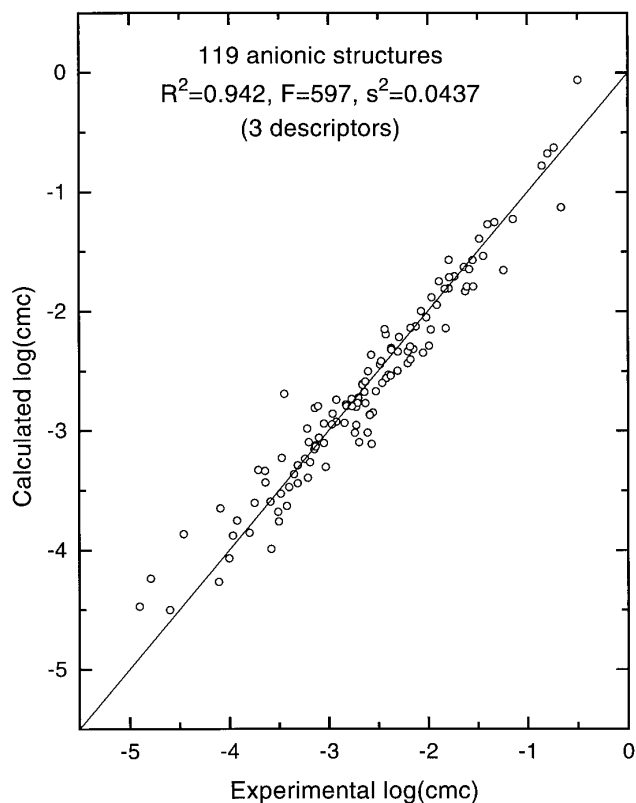


FIG. 2. Scatter plot showing the correlation between calculated and experimental cmc for a diverse set of 119 anionic surfactant structures.

group is attached to an ethoxy group/chain or when a surfactant has a second hydrophilic group other than sulfate or sulfonate. Note that this descriptor does not include contributions related to the total size of the molecule, as does the RNNO term in the nonionic cmc equation (Eq. [1]). The range and average values for all descriptors are summarized in Table 5.

Model for simple sulfates and sulfonates. Another useful multiple linear relationship (Eq. [3]) was developed for a subset of the anionic surfactants that had only the sulfate or sulfonate group as the hydrophilic domain (Fig. 4). Thus, all of the variation in this set of structures is in the hydrophobic domain. Moreover, due to the minimal variation in head fragments, there is no need to use fragment descriptors. This regression is of special interest because it provides an estimate of cmc with a significantly lower error than the general model, while still covering quite a diverse set of surfactant structures.

$$\log_{10}\text{cmc} = (2.42 \pm 0.07) - (0.537 \pm 0.009)\text{KH1} \\ - (0.019 \pm 0.002)\text{KS3} + (0.096 \pm 0.005)\text{HGP} \\ R^2 = 0.988, F = 1691, s^2 = 0.0068, N = 68 \quad [3]$$

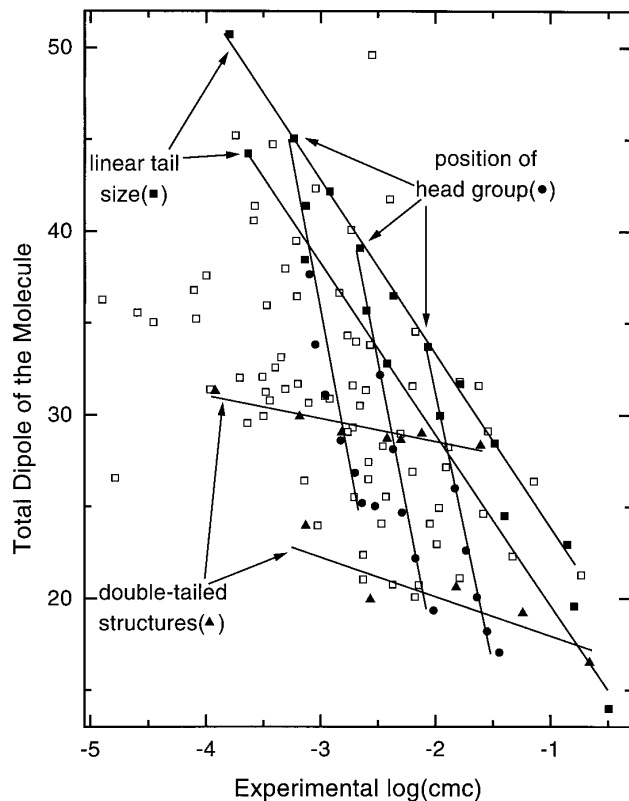


FIG. 3. Scatter plot of the total molecular dipole vs the experimental values of $\log(\text{cmc})$, showing trends in the "linear tail size" with molecular structure. Note that some structures in the "linear tail size" set are also included in the "head group position" correlations.

Since the variation in the head group is minimal, it is no surprise that the three descriptors in this regression describe various aspects of the hydrophobic group. KH1 is the first-order Kier and Hall molecular connectivity index. This descriptor, as the zeroth-order index used in the previous regression, correlates highly with both molecular surface area and volume. KH1 is defined as

TABLE 4
Constants for the Relationship between cmc (mol/L) and Total Dipole Moment: $\log_{10}\text{cmc} = A - B(\text{TDIP})$

Surfactant series	Excursion on:	A	B	r^2
$C_n\text{SO}_3$	Chain length	-2.63	2.74	0.999
$C_n\text{SO}_4$	Chain length	0.873	2.76	0.999
$_n\text{C}_{12}\text{SO}_3$	Head group position	29.7	-2.23	0.963
$_n\text{C}_{14}\text{SO}_4$	Head group position	36.0	-2.54	0.946
$_n\text{C}_{15}\text{CSO}_4$	Head group position	40.3	-2.12	0.943
$C_{12}(3-X)\text{SO}_3$	Side chain length	27.9	0.368	0.770
$C_n\text{C}_n\text{-succinate}$	Chain length (two tails)	9.72	1.80	0.984

$$\text{KH1} = \sum_{j=1}^N \sum_{i=1}^N (\delta_i^y \delta_j^y)^{-1/2}, \quad i \neq j,$$

$$\text{where } \delta_i^y = \frac{Z_i^y - H_i}{Z_i - Z_i^y - 1}, \quad [4]$$

where Z_i is the total number of electrons in the i th atom, Z_i^y is the number of valence electrons, and H_i is the number of hydrogens directly attached to the i th atom. KS3 is the third-order Kier shape index (13) and is defined as

$$\begin{aligned} \text{KS3} &= (A - 1)(A - 3)^2 / ({}^3P_i)^2, \quad \text{when } A \text{ is odd} \\ \text{KS3} &= (A - 3)(A - 2)^2 / ({}^3P_i)^2, \quad \text{when } A \text{ is even,} \end{aligned} \quad [5]$$

where $({}^3P_i)$ is the count of possible paths of three contiguous bonds in the molecular graph, and A is the number of atoms in the molecule. This index contains information on the extent of branching of the molecule and has higher values for linear molecules and lower values for highly branched molecules. HGP is the head group position on the longest linear chain. This index is simply the number of the carbon atom in the chain attached to the head group. For branched alkylbenzenesulfonates, it is the linear alkyl carbon chain number of the attachment of the benzenesulfonate group. This descriptor accounts for the observation that the cmc increases as the head group moves farther from the alpha carbon position. No significant intercorrelations exist between these three descriptors, the largest being between KH1 and KS3 ($r = 0.45$).

CONCLUSIONS

A general QSPR methodology that is useful for the prediction of surfactant properties has been established. The correlation equations presented in this paper provide insight into the structural aspects of anionic surfactants that influence cmc and also allow estimation of cmc for anionic surfactants that have not yet been synthesized. The structural feature most influencing cmc is the size of the hydrophobic domain

TABLE 5
Minimum, Maximum, and Average Values
for Descriptors Used in Eqs. [2] and [3]

Descriptor	No. of structures	Min value	Max value	Avg value
Tail-KHO	119	4.54	16.62	10.86
TDIP	119	13.99	50.72	30.19
Head-RNC	119	0	0.33	0.054
KH1	68	5.07	11.47	9.18
KS3	68	5.24	21.95	13.92
HGP	68	1	10	3.2

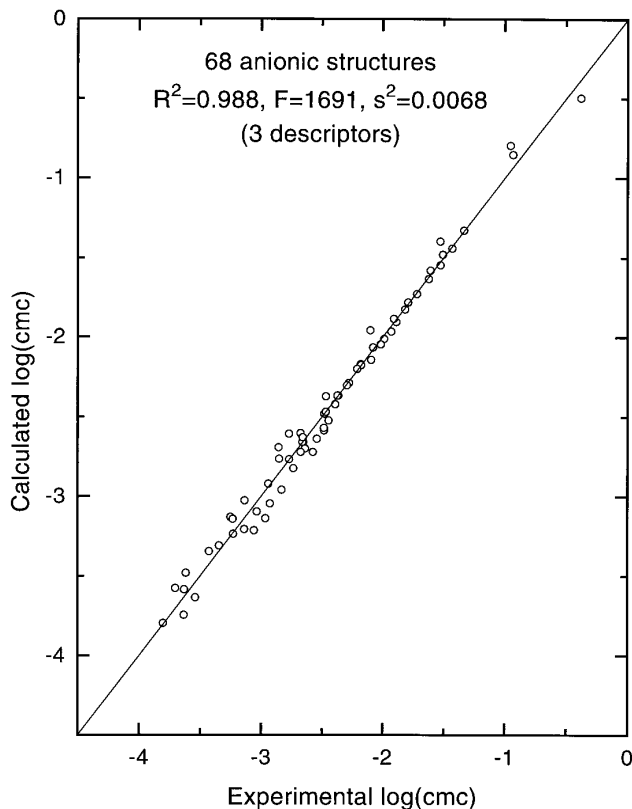


FIG. 4. Scatter plot showing the correlation between calculated and experimental cmc, for a set of 68 anionic surfactant structures containing only sulfate and sulfonate hydrophilic groups.

as represented by the Kier and Hall indices, which are highly correlated with both hydrophobic fragment surface area and volume. This is in agreement with the qualitative knowledge about the nature of water around solutes. Water is known to become more ordered and to distort its hydrogen bonding network, in order to form cavities around solutes with which it cannot form any specific interactions. Other aspects of surfactant molecular structure, such as tail branching, position of the head group, and structural diversity of the head groups, influence cmc as well.

REFERENCES

- Huibers, P. D. T., Lobanov, V. S., Katritzky, A. R., Shah, D. O., and Karelson, M., *Langmuir* **12**, 1462–1470 (1996).
- van Os, N. M., Haak, J. R., and Rupert, L. A. M., “Physico-chemical Properties of Selected Anionic Cationic and Nonionic Surfactants.” Elsevier, Amsterdam, 1993.
- Rosen, M. J., “Surfactants and Interfacial Phenomena,” 2nd ed. Wiley, New York, 1989.
- Mukerjee, P., and Mysels, K. J., “Critical Micelle Concentrations of Aqueous Surfactant Systems.” NSRDS-NBS 36, U.S. Dept. of Commerce, Washington DC, 1971.
- van Os, N. M., Daane, G. J., and Bolsman, T. A. B. M., *J. Colloid Interface Sci.* **115**, 402 (1987).

6. van Os, N. M., Daane, G. J., and Bolsman, T. A. B. M., *J. Colloid Interface Sci.* **123**, 267 (1988).
7. Katritzky, A. R., Lobanov, V. S., and Karelson, M., "CODESSA Software," University of Florida, Gainesville, 1994. CODESSA is distributed by SemiChem, Inc., 7128 Summit, Shawnee, KS 66216.
8. Katritzky, A. R., Lobanov, V. S., and Karelson, M., *Chem. Soc. Rev.* **24**, 279 (1995).
9. "PCMODEL Molecular Modeling Software," 1987–1992. Serena Software, Box 3076, Bloomington, IN 47402.
10. Stewart, J. J. P., "MOPAC 6.0," QCPE No. 455, 1989.
11. Kier, L. B., and Hall, L. H., "Molecular Connectivity in Chemistry and Drug Research." Academic Press, New York, 1976.
12. Kier, L. B., and Hall, L. H., "Molecular Connectivity in Structure-Activity Analysis." Wiley, New York, 1986.
13. Kier, L. B., in "Computational Chemical Graph Theory" (D. H. Rouvray, Ed.), Chap. 6. Nova Science Publishers, New York, 1990.
14. Rosen, M. J., *J. Colloid Interface Sci.* **56**, 320 (1976).
15. Klevens, H. B., *J. Am. Oil Chem. Soc.* **30**, 74 (1953).
16. Schick, M. J., and Fowkes, F. M., *J. Phys. Chem.* **61**, 1062 (1957).
17. Lange, H., "Proceedings of the 4th International Congress on Surface Active Substances," Vol. 2, p. 497. Brussels, Belgium, 1964.
18. Kovtunencko, L. I., Smirnov, N. I., and Titova, N. P., *Zh. Prikl. Khimii* **48**, 325 (1975).
19. Klevens, H. B., *J. Phys. Colloid Chem.* **52**, 130 (1948).
20. Dahanayake, M., Cohen, A. W., and Rosen, M. J., *J. Phys. Chem.* **90**, 2413 (1986).
21. Weil, J. K., Stirton, A. J., Smith, A. J., and Bistline, R. G., U.S. Patent 3,228,980. 11/1/1966.
22. Püschel, F., *Tenside* **4**, 320 (1967).
23. Besserman, M., Ph.D. thesis., University of Washington, Seattle, Washington, 1953.
24. Fanghänel, E., Ortman, W., Behrmann, K., Willscher, S., Turro, N. J., and Gould, I. R., *J. Phys. Chem.* **91**, 3700 (1987).
25. Rouviere, J., Faucompre, B., Lindheimer, M., Partyka, S., and Brun, B., *J. Chim. Phys.* **80**, 309 (1983).
26. Griess, W., *Fette Seifen Anstrichmittel* **57**, 24 (1955).
27. Gershman, J. W., *J. Phys. Chem.* **61**, 581 (1957).
28. Lianos, P., and Lang, J., *J. Colloid Interface Sci.* **96**, 222 (1983).
29. Bauernschmitt, D., Hoffmann, H., and Platz, G., *Ber. Bunsenges. Phys. Chem.* **85**, 203 (1981).
30. Evans, H. C., *J. Chem. Soc.*, 579 (1956).
31. Aniansson, E. A. G., Wall, S. N., Almgren, M., Hoffmann, H., Kielman, I., Ulbricht, W., Zana, R., Lang, J., and Tondre, C., *J. Phys. Chem.* **80**, 905 (1976).
32. Flockhart, B. D., *J. Colloid Sci.* **16**, 484 (1961).
33. Götte, E., and Schwuger, M. J., *Tenside* **3**, 131 (1969).
34. Lange, H., and Schwuger, M. J., *Kolloid Z. Z. Polym.* **223**, 145 (1968).
35. Huisman, H. F., *Proc. Kon. Ned. Akad. Wetensch. B* **67**, 388 (1964).
36. Püschel, F., and Kraatz, K. H., "IV. Intern. Tagung Grenzfl. Stoffe," Band 1, p. 202. Berlin, 1974.
37. Püschel, F., "IV. Intern. Tagung Grenzfl. Stoffe," Band 1, p. 235. Berlin, 1974.
38. Weil, J. K., Smith, F. D., Stirton, A. J., and Bistline, R. G., *J. Am. Oil Chem. Soc.* **40**, 538 (1963).
39. Jobe, D. J., and Reinsborough, V. C., *Can. J. Chem.* **62**, 280 (1984).
40. Hikota, T., Morohara, K., and Meguro, K., *Bull. Chem. Soc. Jpn.* **43**, 3913 (1970).
41. Schwuger, M. J., *Fette Seifen Anstrichmittel* **72**, 565 (1970).
42. Weil, J. K., and Stirton, A. J., *J. Phys. Chem.* **60**, 899 (1956).
43. Williams, E. F., Woodbury, N. T., and Dixon, J. K., *J. Colloid Sci.* **12**, 452 (1957).