JOURNAL OF COLLOID AND INTERFACE SCIENCE **193,** 132–136 (1997)
ARTICLE NO. CS975053

NOTE

Predicting Surfactant Cloud Point from Molecular Structure

A general empirical relationship has been developed for estimating the cloud point of pure nonionic surfactants of the alkyl ethoxylate class. For a set of 62 structures, composed of linear alkyl, branched alkyl, cyclic alkyl, and alkylphenyl ethoxylates, cloud points can be estimated to an accuracy of ±6.3°C (3.7°C median error) using the logarithm of the number of ethylene oxide residues and three topological descriptors that account for hydrophobic domain variation.

Key Words: **cloud point; nonionic surfactant; topological indices; information-theoretical indices; QSPR; quantitative structure–property relationships; CODESSA.**

INTRODUCTION

The cloud point is an important property of nonionic surfactants; below this temperature a single phase of molecular or micellar solution exists, above it the surfactant loses sufficient water solubility and a cloudy dispersion results (1). The aggregates in the cloudy dispersion are much larger than micelles, and in some situations the solution phase separates into a surfactant rich phase and a surfactant poor phase. Above the cloud point the surfactant ceases to perform some or all of its normal functions as a surfactant. This can either limit the choice of nonionic surfactant applied to certain processes, or it can be used advantageously, as in cloud point separation (surfactant mediated phase separation) of organic compounds (2, 3) and proteins (4).

Several trends in cloud point with surfactant molecular structure are commonly known: cloud point increases with the relative ethylene oxide content and decreases with increasing alkyl carbon chain length. It is worth noting that the cloud point of aqueous surfactant solutions can be strongly influenced by the presence of other materials. For example, cloud point can be decreased by the introduction of polar compounds, anions that are water structure formers (hard bases, F^-, OH^-, SO_4^{2-}, Cl^-, and PO_4^{3-}), and certain cations (NH_4^+, alkali metal ions except for Li^+). Cloud point can be increased by addition of long chain nonpolar material, anions that are water structure breakers (large, polarizable anions, soft bases, SCN^-, I^-), and certain cations (polyvalent cations, H^+, Li^+) (5).

Several empirical relationships between structure and cloud point can be found in the literature. Gu and Sjöblom (6) have demonstrated a linear relationship between the cloud point and the logarithm of the ethylene oxide number (EO#) for alkyl ethoxylates, alkylphenyl ethoxylates, and methyl capped alkyl ethoxylate esters, as well as a linear relationship between the cloud point and alkyl carbon number (C#) for linear alkyl ethoxylates. For the first two classes of nonionic surfactant mentioned above, they derived an overall empirical relationship,

$$\mathrm{CP} = A \log(\mathrm{EO\#}) - 5.5\ \mathrm{C\#} - B, \qquad [1]$$

where A and B are empirical constants depending on the surfactant class. While these relationships are useful for the surfactant classes for which they have been determined, a more general empirical relationship is desirable.

The validity of predicting properties from molecular structure using a quantitative structure–property approach (QSPR) has been well established and applied to a variety of classes of compounds and properties, using topological, geometrical, electrostatic, and quantum–chemical descriptors (7). One important aspect of the resulting regressions is that properties can be predicted directly from the molecular structure, for molecules which have never been synthesized. The application of topological indices to surfactant property prediction has been established, predicting the critical micelle concentration from surfactant structure for the nonionic (8) and anionic (9) surfactants.

Topological descriptors are an application of the mathematical discipline of graph theory to chemistry, where the structural formula of a molecule is equivalent to the corresponding molecular graph. Topological indices are formed from different combinations and weightings of the vertices (atoms) and edges (bonds) of the molecular graph (chemical structure). The application of topological descriptors to chemical property prediction was initiated by Weiner (10), and greatly expanded by Randic (11), Kier and Hall (12), and others in the 1970s. By the 1980s, the utility of topological descriptors in the prediction of chemical and physical properties was well established (13, 14). Topological descriptors have certain advantages over quantum–chemical or geometrical descriptors, in that they are much easier to calculate, can be precisely calculated (they are not approximations and do not contain empirical constants), and are based only on the two-dimensional structure (graph) of the molecule (15).

The use of topological descriptors should be limited to certain types of properties, where size and shape effects dominate. They have been successfully applied to a number of alkane properties where there are only nonspecific molecular interactions. One aspect of their utility is the ability to handle isomers well. In indices with order higher than zero, the neighboring atoms or bonds are considered in the weighting of each atom or bond. This compares favorably to group contribution methods, which are basically ''zero'' order, and cannot account for differences in isomers. Given these limitations and advantages, it should be possible to account for the cloud point of nonionic surfactants. Cloud point data are available for a large number of nonionic surfactants, where the primary variation is in the hydrophobic surfactant tail, with a variety of alkane branching and ring structures. This is perfectly suited to the application of topological indices.

DATA AND METHODS

Cloud point data from many literature sources have been compiled by van Os (16) and Rosen (5) (Table 1). This set of data represents a wide variety of structure in the hydrophobic domain, including linear alkyl, branched alkyl, cyclic alkyl, linear alkylphenyl, and branched alkylphenyl ethoxylates (Fig. 1). Cloud point values were found for 62 different structures. Note that the cloud point is less often reported than other properties, such as the cmc.

For structures with several reported cloud point values in the literature, the average has been used in this study. These multiple reported values can give some insight into the approximate experimental error of the data. Twenty-one structures with multiple reported values had an average standard deviation of ±2.7°C around their mean cloud points. The lower limit of error approached by a predictive method is the experimental error, for if the model error is less than the experimental error, the model is fitting some systematic aspect of the experimental error, and not the physical property of interest.

0021-9797/97 $25.00

TABLE 1
Cloud Points (Literature Values and Estimated Values) of 62 Nonionic Surfactants

Structure	EO No.	CP_{lit} (°C)	CP_{est} (°C)	$Cp_{lit} - CP_{est}$	Structure	EO No.	CP_{lit} (°C)	CP_{est} (°C)	$Cp_{lit} - CP_{est}$
C6E3[a]	3	40.5	36.0	4.5	C13E8	8	72.5	68.7	3.8
C6E4[a]	4	63.8	60.7	3.1	C14E5	5	20	26.5	−6.5
C6E5	5	75	80.0	−5.0	C14E6[a]	6	42.3	42.2	0.1
C6E6	6	83	95.6	−12.6	C14E7	7	57.6	55.4	2.2
C8E3[a]	3	7	11.1	−4.1	C14E8	8	70.5	66.9	3.6
C8E4[a]	4	38.5	35.8	2.7	C15E6	6	37.5	40.1	−2.6
C8E5[a]	5	58.6	55.0	3.6	C15E8	8	66	64.9	1.1
C8E6[a]	6	72.5	70.7	1.8	C16E6[a]	6	35.5	39.3	−3.8
C8E8	8	96	95.5	0.5	C16E7[a]	7	54.0	52.6	1.4
C9E4	4	32	27.1	4.9	C16E8[a]	8	65.0	64.1	0.9
C9E5	5	55	46.3	8.7	C16E9	9	75	74.2	0.8
C9E6	6	75	62.0	13.0	C16E12	12	92	99.0	−7.0
C10E4[a]	4	19.7	21.4	−1.7	IC6E6	6	78	77.5	0.5
C10E5[a]	5	41.6	40.6	1.0	IC10E6	6	27	26.7	0.3
C10E6[a]	6	60.3	56.3	4.0	C12E9A	9.4	84	86.1	−2.1
C10E8	8	84.5	81.1	3.4	XC12E9	9.2	75	80.1	−5.1
C10E10	10	95	100.3	−5.3	C13E8A	8.9	79	77.9	1.1
C11E4	4	10.5	15.9	−5.4	IC13E9	9.2	35	59.0	−24.0
C11E5	5	37	35.1	2.1	TC13E9	9.2	34	37.7	−3.7
C11E6	6	57.5	50.8	6.7	C16E12A	12.2	97	100.4	−3.4
C11E8	8	82	75.5	6.5	TC16E12	12.0	48	47.3	0.7
C12E4[a]	4	6.0	12.5	−6.5	XC16E11	11.9	80	84.0	−4.0
C12E5[a]	5	28.9	31.7	−2.8	TC8PE9	9.5	64.3	70.4	−6.1
C12E6[a]	6	51.0	47.4	3.6	NC09PE8	8	34	49.8	−15.8
C12E7[a]	7	64.7	60.7	4.0	NC09PE9	9.2	56	61.9	−5.9
C12E8[a]	8	77.9	72.2	5.7	NC09PE10	10	75	69.1	5.9
C12E9[a]	9	87.8	82.4	5.4	NC09PE12	12.4	87	87.6	−0.6
C12E10[a]	10	95.5	91.4	4.1	NC09PE13	13	89	91.6	−2.6
C12E11	11	100.3	99.6	0.7	NC12PE9	9	33	27.7	5.3
C13E5	5	27	28.3	−1.3	NC12PE11	11.1	50	45.7	4.3
C13E6	6	42	44.0	−2.0	NC12PE15	15	90	71.6	18.4

[a] Cloud point values are an average of several literature sources, as compiled by van Os *et al.* (16).

The cloud point is assumed to be for the pure surfactant. For certain cases where pure compound data are not available, data for material with an ethylene oxide size distribution is used. These cloud points are known to be lower than the cloud points of pure compounds. Three known comparisons between pure and mixed surfactants were available (5), with a difference of 3.5°C for $C_{12}E_7$, 6°C for $C_{12}E_8$, and 7°C for $C_{12}E_{10}$. A clear relationship between the two cannot be established with such few data points, but given the magnitude of the difference and the size of the estimated error in cloud point measurement, we choose to include the cloud point data for surfactants with EO# distributions without alteration. Additionally, three cloud point values (C_4E_1 at 46.6°C, C_5E_2 at 36°C, C_8E_{12} at 106°C) were clear outliers in most regressions, and were rejected as inaccurate measurements.

CALCULATIONS

The calculation of topological descriptors for the hydrophobic domain, the search for optimum structure–property relationships, and the calculation of the statistics of the resulting regressions were performed using the CODESSA program (17). The specific details of the methodology have been described elsewhere (7). Briefly, the three dimensional structure of the molecules are entered in computer readable form. The structures were read into CODESSA, and topological descriptors were calculated for the hydrophobic domain, which is considered to be the contiguous carbon backbone up to and including the carbon attached to the first oxygen atom of the polyethylene oxide chain. The program provides different means of searching for an optimum regression between the properties of interest and any number of descriptors and will calculate the statistical parameters of any regression. The topological descriptors included in the analysis were the Randic indices, Kier and Hall molecular connectivity indices, Kier shape indices, Kier flexibility index, and information content indices.

RESULTS AND DISCUSSION

The empirical relationship developed by Gu and Sjöblom (6) has been recalculated for the larger data set considered here. The relationship for the linear alkyl ethoxylates is

$$CP = (87.1 \pm 3.3)\log EO\# - (5.78 \pm 0.38)C\# - (40.7 \pm 5.2)$$
$$R^2 = 0.943,\ F = 355,\ s^2 = 40.4,\ N = 46. \quad [2]$$

The high correlation coefficient and low standard error of ±6.4°C for this regression is quite satisfactory for the subset of surfactants that it predicts. However, if we use these descriptors for all alkyl ethoxylates, the result is less satisfactory,

$$CP = (76.8 \pm 5.8)\log EO\# - (5.66 \pm 0.68)C\# - (21.9 \pm 8.6)$$
$$R^2 = 0.749,\ F = 88,\ s^2 = 162,\ N = 62. \quad [3]$$

From the low correlation coefficient of Eq. [3] it can be seen that the carbon number is an inadequate descriptor for modeling the cloud point for other than linear alkyl ethoxylates. The topological descriptors were developed precisely for this reason, to allow one to account for branching and ring structure in alkanes (14). Thus it is reasonable to explore multiple linear regressions of a variety of topological descriptors in order to predict the cloud point of the structures considered in this paper.

TOPOLOGICAL DESCRIPTORS

A number of descriptors have been developed over the years to account for structural variation in hydrocarbons, based on the connectivity or topology of the molecule (15, 18). These can be applied to the prediction of cloud point, to account for the influence of variations in hydrophobic structures. Regressions were calculated using different combinations of 41 topological descriptors. The best regression resulted in (Fig. 2)

$$CP = (-264. \pm 17.) + (86.1 \pm 3.0)\log \text{EO\#} + (8.02 \pm 0.78)^3\kappa - (1284 \pm 86)^0\text{ABIC} - (14.26 \pm 0.73)^1\text{SIC}$$
$$R^2 = 0.937,\ F = 211,\ s^2 = 42.3,\ N = 62, \qquad [4]$$

where EO# is the number of ethylene oxide residues, $^3\kappa$ is the third order Kier shape index for the hydrophobic tail, ^{0}ABIC is the zeroth order average bonding information content of the tail, and ^{1}SIC is the first order structural information content of the tail. Although the standard error (s) of 6.3°C appears quite large, we believe it is quite reasonable. Considering a statistical analysis of the Gu and Sjöblom relationships for homologous series of C_8E_m through $C_{16}E_m$, where limited data is available (only five to eight structures per homologous series), the standard deviation is in the range of 3.1 to 4.3°C. This seems quite large, especially in light of the fact that cloud point for a given sample can be measured accurately and repeatably to a fraction of a degree. The variability that we see in the published data can be attributed to sample purity and variation between experimenters. The standard error of 6.3°C in our analysis is indeed skewed by a few points with high error, and the median error is actually just 3.7°C.

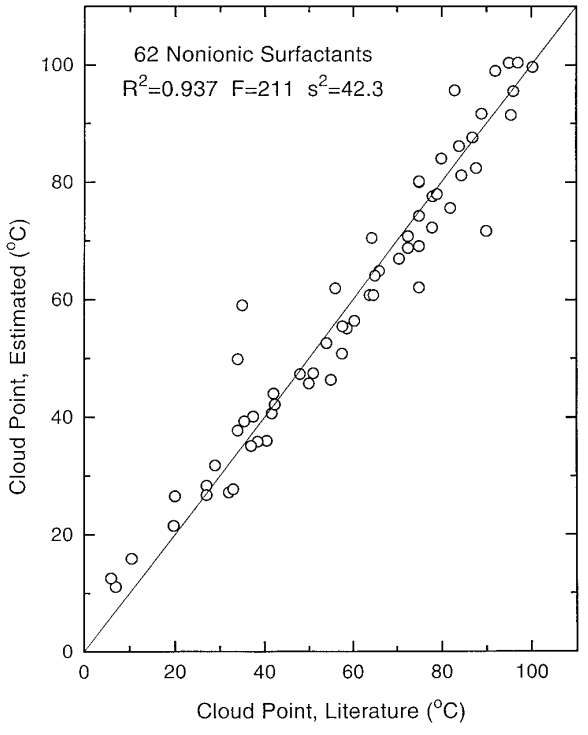

FIG. 2. Scatter plot for the calculated (Eq. 4) vs literature values of the nonionic surfactant cloud point.

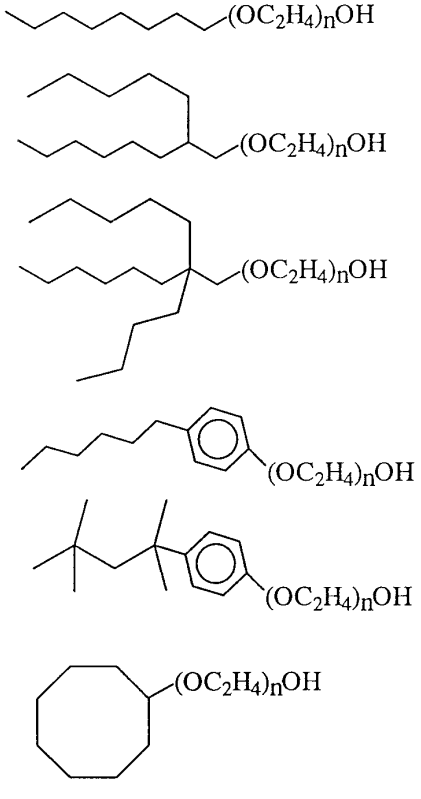

FIG. 1. The variety of nonionic ethoxylate structures in the cloud point data set.

The logarithm of the EO# was used successfully by Gu and Sjöblom (6) and proved to work well for regressions of both alkyl and alkylphenyl ethoxylates. Other nonlinear terms involving EO# were investigated, along with cross-terms between EO# and C# terms, but no significant improvement over the logarithm function was discovered.

The topological descriptors model various aspects of the hydrophobic tail structure. The third order Kier shape index ($^3\kappa$) increases with the size of the hydrophobe, from 1.33 for C_6 to 9.95 for C_{16}. $^3\kappa$ is lower for branched hydrophobes with the same carbon number, with a value of 7.10 for C_{13}, 5.33 for i-C_{13}, and 3.32 for t-C_{13} (Fig. 3). Cyclic structures result in even lower values for the same carbon number (Table 2). $^3\kappa$ is calculated by (19)

$$^3\kappa = (N_{SA} + \alpha - 1)(N_{SA} + \alpha - 3)^2(^3P + \alpha)^2 \quad \text{if } N_{SA} \text{ is odd}$$
$$^3\kappa = (N_{SA} + \alpha - 3)(N_{SA} + \alpha - 2)^2(^3P + \alpha)^2 \quad \text{if } N_{SA} \text{ is even}, \qquad [5]$$

where N_{SA} is the number of skeleton atoms (hydrogen excluded), α is one less than the sum of ratios of atomic radii to carbon atom atomic radii for all skeletal atoms, and ^3P is the number of paths of length three (three contiguous bonds) in the skeleton.

The information content indices are derived from the average information content (^{k}IC) as defined in Shannon information theory, and calculated by

$$^k\text{IC} = -\sum_i \frac{n_i}{n} \log_2 \frac{n_i}{n}, \qquad [6]$$

where n is the total number of atoms, n_i is the number of atoms in the ith class, and k is the number of neighboring atoms considered in the formation of a class, where a class is a unique set of $k + 1$ atoms (20, 21).

The zeroth order average bonding information content (^{0}ABIC) decreases with increasing carbon number of the molecule, from 0.216 for C_6 to 0.163 for C_{16}. ^{0}ABIC will increase with increasing variety of types of atoms (atom and hybridization). ^{0}ABIC increases over the linear hydrophobes for the cyclic and alkylphenyl structures (Table 2), but does not increase for the branched hydrophobes (Fig. 3). ^{0}ABIC is calculated by (22),

$$^0\mathrm{ABIC} = {}^0\mathrm{IC}/\log_2 q, \qquad [7]$$

where q is the number of edges (bonds) in the structural graph of the molecule.

The first order structural information content (^{1}SIC) increases with carbon number, from 5.79 for C_6 to 9.86 for C_{16}. The value of this descriptor increases with increasing branching or ring structure of the molecule (Fig. 3). ^{1}SIC is calculated by (22)

$$^1SIC = n\,{}^1\mathrm{IC}/\log_2 n. \qquad [8]$$

The combination of these three topological descriptors, which increase or decrease with carbon number and type of branching in different nonlinear ways, provides the best fit to the variation in the cloud point of the structures considered here.

CONCLUSION

An empirical relationship has been developed for predicting the cloud point of alkyl ethoxylate surfactants. The cloud points for surfactants with a wide range of hydrophobic domain structural variation is predicted by applying topological descriptors, developed from the application of graph theory to chemistry, thus allowing one equation to be used for all surfactant structures considered in this effort. Using the logarithm of the ethylene oxide count (EO#) and three topological terms, we have estimated the effect of diverse hydrocarbon tail structures on the cloud point of nonionic surfactants, with a standard error of just 6.5°C (median error of 3.7°C).

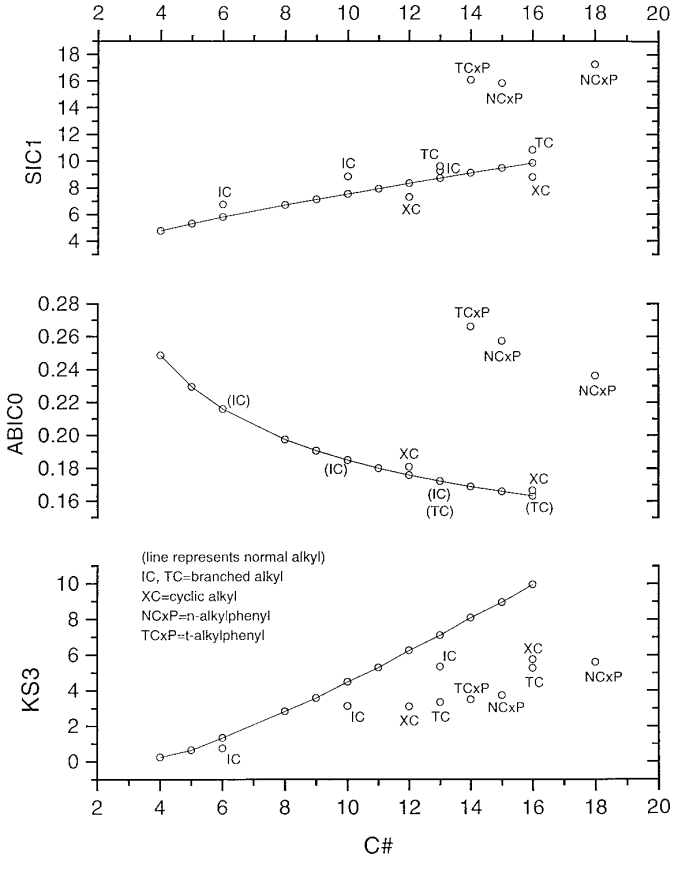

FIG. 3. Topological descriptor values vs carbon number for the different hydrophobic structures.

TABLE 2
Topological Index Values for the Hydrophobic Structures

Label	Structure	$^3\kappa$	^{0}ABIC	^{1}SIC
c4	C_4H_9-	0.250	0.248	4.750
c5	$C_5H_{11}-$	0.640	0.229	5.298
c6	$C_6H_{13}-$	1.333	0.216	5.792
c8	$C_8H_{17}-$	2.813	0.197	6.697
c9	$C_9H_{19}-$	3.556	0.191	7.123
c10	$C_{10}H_{21}-$	4.480	0.185	7.537
c11	$C_{11}H_{23}-$	5.289	0.180	7.942
c12	$C_{12}H_{25}-$	6.250	0.176	8.338
c13	$C_{13}H_{27}-$	7.101	0.172	8.726
c14	$C_{14}H_{29}-$	8.082	0.169	9.109
c15	$C_{15}H_{31}-$	8.960	0.166	9.486
c16	$C_{16}H_{33}-$	9.953	0.163	9.859
ic6	$(C_2H_5)_2CHCH_2-$	0.750	0.216	6.734
ic10	$(C_4H_9)_2CHCH_2-$	3.111	0.185	8.847
ic13	$(C_6H_{13})_2CH-$	5.333	0.172	9.259
tc13	$(C_4H_9)_3CH-$	3.324	0.172	9.623
tc16	$(C_5H_{11})_3CH-$	5.265	0.163	10.843
xc12	*c*-$(C_{12}H_{24})-$	3.083	0.181	7.297
xc16	*c*-$(C_{16}H_{32})-$	5.748	0.167	8.800
tc8p	*t*-$C_8H_{17}-C_6H_4-$	3.483	0.266	16.070
nc09p	*n*-$C_9H_{19}-C_6H_4-$	3.721	0.257	15.827
nc12p	*n*-$C_{12}H_{25}-C_6H_4-$	5.591	0.236	17.238

Note. $^3\kappa$, Kier shape index of third order; ^{0}ABIC, average bond information content of zeroth order; ^{1}SIC, structural information content of first order.

ACKNOWLEDGMENTS

The authors thank Professor M. Karelson of Tartu University, Estonia, and Dr. V. S. Lobanov of 3D Pharmaceuticals for their help and advice on the use of the CODESSA program and the application of topological indices. P.D.T.H. thanks Dr. N. M. van Os of the Koninklijke Shell Laboratorium Amsterdam for helpful discussions on the prediction of surfactant properties.

REFERENCES

1. Myers, D., ''Surfactant Science and Technology,'' 2nd ed. VCH, New York, 1992.
2. Cordero, B. M., Pavon, J. L. P., Pinto, C. G., and Laespada, M. E. F., *Talanta* **40,** 1703 (1993).
3. Akita, S., and Takeuchi, H., *Sep. Sci. Technol.* **30,** 833 (1995).
4. Saitoh, T., and Hinze, W. L., *Talanta* **42,** 119 (1995).
5. Rosen, M. J., ''Surfactants and Interfacial Phenomena,'' 2nd ed. Wiley, New York, 1989.
6. Gu, T., and Sjöblom, J., *Colloids Surf.* **64,** 39 (1992).
7. Katritzky, A. R., Lobanov, V. S., and Karelson, M., *Chem. Soc. Rev.* **24,** 279 (1995).
8. Huibers, P. D. T., Lobanov, V. S., Katritzky, A. R., Shah, D. O., and Karelson, M., *Langmuir* **12,** 1462 (1996).

9. Huibers, P. D. T., Lobanov, V. S., Katritzky, A. R., Shah, D. O., and Karelson, M., *J. Colloid Interface Sci.* **187,** 113 (1997).
10. Weiner, H., *J. Am. Chem. Soc.* **69,** 17 (1947).
11. Randic, M., *J. Am. Chem. Soc.* **97,** 6609 (1975).
12. Kier, L. B., and Hall, L. H., ''Molecular Connectivity in Chemistry and Drug Research.'' Academic Press, New York, 1976.
13. Rouvray, D. H., *Sci. Am.* **254,** 40 (1986).
14. Rouvray, D. H., *J. Comput. Chem.* **8,** 470 (1987).
15. Stankevich, M. I., Stankevich, I. V., and Zefirov, N. S., *Russ. Chem. Rev.* **57,** 191 (1988).
16. 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.
17. Katritzky, A. R., Lobanov, V. S., and Karelson, M., *CODESSA software*, University of Florida, 1994. CODESSA is distributed by Semi-Chem, Inc., 7128 Summit, Shawnee, KS 66216.
18. Rouvray, D. H., *J. Mol. Struct.* **185,** 187 (1989).
19. Kier, L. B., *in* ''Computational Chemical Graph Theory'' (D. H. Rouvray, Ed.), pp. 151. Nova Science Publishers, New York, 1990.
20. Kier, L. B., *J. Pharm. Sci.* **69,** 807–810 (1980).
21. Bonchev, D., ''Information Theoretic Indices for Characterization of Chemical Structure.'' Wiley, New York, 1983.
22. Basak, S. C., Harriss, D. K., and Magnuson, V. R., *J. Pharm. Sci.* **73,** 429 (1984).

Paul D. T. Huibers[1]
Dinesh O. Shah

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

Alan R. Katritzky

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

Received October 28, 1996; accepted June 24, 1997

[1] To whom correspondence should be addressed at Department of Chemical Engineering, Massachusetts Institute of Technology, #66-317, Cambridge, MA 02139-4307. Fax: (617) 253-8723. E-mail: huibers@mit.edu.