(15) E. N. da C. Andrade, "Viscosity and Plasticity," Heffer & Sons, Cambridge, England, 1947, Lecture I.
(16) H. Green, "Industrial Rheology and Rheological Structures," Wiley, New York, N. Y., 1949, pp. 19, 20.
(17) A. C. Zettlemoyer and G. W. Lower, *J. Colloid Sci.*, **10**, 29(1955).
(18) H. Schott, *Kolloid-Z. Z. Polym.*, **199**, 158(1964).

ACKNOWLEDGMENTS AND ADDRESSES

Received August 29, 1969, from *Temple University School of Pharmacy, Philadelphia, PA 19140*

Accepted for publication April 30, 1970.

The author is indebted to Mr. J. F. Wichmann of the U. S. Forest Products Laboratory, Madison, Wis., for measuring the mechanical properties of the films.

NOTES

Cautions Regarding the Physical Interpretation of Statistically Based Structure–Activity Relationships

ARTHUR CAMMARATA*, RICHARD C. ALLEN*, J. K. SEYDEL†, and E. WEMPE†

Abstract □ The distinction between the use of multiple regression analysis as a predictive tool and as a means of investigating controlling physical characteristics in structure–activity studies often is unrecognized. Three examples of complications that can arise with either of these goals in mind are discussed. The first is an illustration of a "false" parabolic dependence of activity on lipophilicity; the second deals with unrecognized interrelationships between certain physical parameters; and the third is a situation where a number of statistically significant correlations can be presented, each of which may be given a differing physical interpretation.

Keyphrases □ Structure–activity relationships—precautions concerning interpretation □ Multiple regression analysis—use, misuse

Within the past few years, multiple regression analysis has been exploited as a statistical tool for the evaluation of structure–activity data. One goal of these analyses is the derivation of a regression equation which will provide estimates of the biological potencies for additional structural entities within a series. A second goal is the determination of the physical and chemical properties of a given series of compounds which are most influential in affecting the observed biological potencies. It is often not recognized that each of these goals represents a separate problem, because the multiple regression approach usually makes use of physically meaningful parameters. As a consequence, any regression equation that correlates structure–activity data can be given a physical interpretation. At times, however, a quirk within a set of structure–activity data can lead to a statistically significant regression equation which provides a poor, if not erroneous, reflection of the physical factors affecting biological potency. At other times, a physical interpretation becomes difficult because a number of correlations can be presented for the same data, each of which involves one or more parameters that could be given a differing physical interpretation. Three examples to illustrate these complications are discussed in this report.

EXAMPLE 1

The minimum inhibitory concentrations against *Escherichia coli* for a variety of congeneric sulfanilamides have been correlated linearly with the Hammett σ-value or with the pKa for the compounds (1, 2). Other congeneric sulfanilamide series have been found (3) which require the addition of π or of π and π^2 terms in a multiple regression model in order to gain a correlation with their bacteriostatic activities. A π term in combination with σ or pKa in a regression equation indicates that lipophilic and electronic factors, respectively, are controlling biological potency. When both π and π^2 appear in a regression equation, the biological activities are parabolically related to the lipophilicities of the compounds; *i.e.*, there is an optimal lipophilicity to observe a maximum biological response within the series.

A reasonable approach to follow if regression equations are to be used as a guide to further syntheses is first to synthesize and test a relatively few compounds which vary over a wide range with respect to their expected electronic (σ) or lipophilic (π) characteristics. If the subsequent regression analysis requires the addition of a physical parameter other than that chosen as an initial criterion to correlate the activity data, it could be said that the additional parameter (or parameters) reflects a real physical requirement for the system under study. Following this approach, N^1-benzoylsulfanilamides having substituents covering extremes in Hammett σ-values (Compounds 1–7; Table I) are found to have their bacteriostatic potencies correlated by the equation

$$\log(1/C) = \underset{(-4.99)}{-0.81}\ (\pm 0.16)\sigma + \underset{(3.74)}{1.18}\ (\pm 0.31)\pi \quad \text{(Eq. 1)}$$
$$\underset{(-3.38)}{-1.19}\ (\pm 0.35)\pi^2 + 5.10$$

N	s	R	$F(3,3)$
7	0.12	0.98	34.67

In Eq. 1, the standard error for the estimate of a coefficient appears in parentheses after the coefficient; in parentheses below the coefficient is the t test. The statistics for the equation are the standard error of the estimate s, the multiple correlation coefficient R, and the F-ratio.

Based on Eq. 1, it may be concluded that electronic and lipophilic factors control the bacteriostatic activities for this series of sulfanilamides and that there is an optimum lipophilicity for the series. If the latter conclusion is true, a regression equation based on the activities for a larger number of N^1-benzoylsulfanilamides should retain the π and π^2 terms as found in Eq. 1, since the lipophilicities for the additional compounds should lie on the same parabola as is found for the smaller series. In this particular instance, an extension

Table I—*In Vitro* Activities against *Escherichia coli* of N^1-Benzoylsulfanilamides

Compound Number	Benzoyl Substituent	σ	π	Activities,[a] log (1/C)
1	4-OMe	−0.27	0.08	5.40
2	4-Me	0.17	0.42	5.40
3	H	0.0	0.0	5.25
4	4-Cl	0.23	0.87	5.10
5	3-CF_3	0.42	1.07	4.65
6	4-NO_2	0.78	0.02	4.50
7	4-CN	0.63	−0.31	4.05
8	3-Me	−0.07	0.52	5.40
9	4-iso-Pr	−0.15	1.40	5.40
10	4-Et	−0.15	0.92	5.62
11	4-*n*-Pr	−0.15	1.42	5.18
12	3,4-Me	−0.24	0.94	5.40
13	3-Me, 4-MeO	−0.44	0.60	5.25

[a] These activities were determined in the laboratories of J. K. S. and have been reported in previous discussions by J. K. Seydel and E. Wempe, *Arzeim.-Forsch.*, **14**, 705(1964), and by A. Cammarata, *J. Med. Chem.*, **11**, 1111(1968).

of the series (Compounds 1–13, Table I) eradicates the π, π^2 dependence:

$$\log(1/C) = \underset{(-4.51)}{-0.88\ (\pm 0.19)\sigma} \underset{(1.81)}{+0.59\ (\pm 0.32)\pi} \underset{(-1.62)}{-0.39\ (\pm 0.24)\pi^2} + 5.04 \qquad \text{(Eq. 2)}$$

N	s	R	$F(3,9)$
13	0.21	0.91	14.78

The equation correlating the total set of data is, therefore,

$$\log(1/C) = \underset{(-6.05)}{-1.06\ (\pm 0.17)\sigma} + 5.15 \qquad \text{(Eq. 3)}$$

N	s	R	$F(1,11)$
13	0.22	0.87	36.10

which indicates that only electronic factors influence the activities of these compounds.

With Eq. 3 as a guide, it is readily found that Compound 7 alone led to the π, π^2 dependence found in Eq. 1. This compound represents a terminal point displaced relative to the overall linear trend of activity with σ which holds for the other compounds (Compounds 1–6). When Compound 7 is deleted from the set used to derive Eq. 1, the resultant regression equation becomes

$$\log(1/C) = \underset{(-5.73)}{-0.84\ (\pm 0.14)\sigma} + \underset{(1.07)}{0.57\ (\pm 0.53)\pi} \underset{(-1.21)}{-0.63\ (\pm 0.52)\pi^2} + 5.17 \qquad \text{(Eq. 4)}$$

N	s	R	$F(3,2)$
6	0.11	0.98	20.11

The statistics for the coefficients in Eq. 4 clearly indicate that the π and π^2 terms should be deleted.

For the 13 N^1-benzoylsulfanilamides considered, a linear correlation between their bacteriostatic potency and their Hammett σ-values is suitable for directing later syntheses. An eventual π, π^2 dependence will most probably be found as a wider variety of multiple substitutions are made on the N^1-benzoyl moiety. In this event, it is likely that the optimum lipophilicity found will differ substantially from the optimum lipophilicity calculated on the basis of Eq. 1.

EXAMPLE 2

In testing alternative physical indexes in attempted correlations of structure–activity data, the rationale commonly used is to attribute physical significance only to those terms that appear as statistically significant in the derived regression equation. For example, if the use of the Hammett σ-constant does not lead to a regression equation in which this index is shown to be statistically significant, the inference that may be drawn is that an alternative electronic index, such as group polarizability P_E or group dipole moment μ, might be more suitable. These alternative indexes are often considered as measures of electronic properties not encompassed by the Hammett σ-value. Seldom, however, are these more specific electronic indexes investigated with respect to the general electronic index they are intended to displace. The result of this type of an investigation can considerably complicate the physical interpretation of a regression equation for a given body of structure–activity data.

Table II—Group Dipole Moments and Hammett σ-Values

Group	σ_p[a]	μ, obs.[b]	μ, estd.	$\lvert d \rvert$
SO_2[c]	0.728	−5.14	−4.00	1.14
SO[c]	0.567	−4.08	−3.33	0.75
CN	0.628	−4.05	−3.58	0.46
NO_2	0.778	−4.01	−4.20	0.19
COMe	0.516	−2.96	−3.12	0.16
CHO	0.216	−2.96	−1.86	1.09
CF_3	0.551	−2.60	−3.26	0.66
CCl_3	0.42[d]	−2.07	−2.72	0.65
$CHCl_2$	0.34[d]	−2.03	−2.38	0.35
CH_2Cl	0.184	−1.82	−1.73	0.08
Cl	0.226	−1.60	−1.91	0.31
OH	−0.357	−1.60	−0.51	2.11
Br	0.232	−1.57	−1.93	0.36
F	0.062	−1.48	−1.23	0.25
I	0.276	−1.42	−2.12	0.70
SMe	−0.047	−1.18	−0.77	0.40
Me	−0.170	0.35	−0.26	0.61
$SiMe_3$	−0.01	0.42	−0.93	1.35
OMe	−0.268	1.28	0.15	1.12
NH_2	−0.660	1.52	1.77	0.25
NMe_2	−0.600	1.61	1.53	0.08

[a] Hammett σ-values for *para*-substituents as given by K. B. Wiberg, "Physical Organic Chemistry," Wiley, New York, N. Y., 1964, p. 410. [b] L. E. Sutton, in "Determination of Organic Structures by Physical Methods," E. A. Braude and F. C. Nachod, Eds., Academic, New York, N. Y., 1955. [c] Assumed to have a Me group substituted on the S atom. The slight contribution made by the dipole moment of the Me group is neglected. [d] Calculated based on a quantum perturbation theory approach, F. L. J. Sixma, *Rec. Trav. Chim.*, **72**, 673(1953).

Certain pharmacological agents have been indicated as having their effect on a biological system more adequately interpreted in terms of the group dipole moments, μ, than of the Hammett σ-values for the substituents (4–6). An investigation of the relation between the group dipole moments and the Hammett σ-values for the substituents found in Table II, however, reveals a significant correlation between the two indexes:

$$\mu = \underset{(-9.16)}{-4.162\ (\pm 0.454)\sigma} - 0.969 \qquad \text{(Eq. 5)}$$

N	s	R	$F(1,19)$
21	0.83	0.90	83.88

Thus, a correlation of biological potency which involves μ or σ should be similarly interpreted. If, in two respective regression

Table III—Neuraminidase Inhibition by 1-Phenoxymethyl-3,4-dihydroisoquinolines

Phenoxy Substituent	σ	π	μ_v	log (1/C)
4-NO_2	0.78	0.50	−4.01	2.903
4-Br	0.27	1.13	−1.57	2.767
4-CN	0.66	0.14	−4.05	2.839
4-Cl	0.23	0.93	−1.60	2.807
4-F	0.06	0.31	−1.48	2.634
H	0.0	0.0	0.0	2.577
4-Me	−0.17	0.48	0.35	2.682
4-OMe	−0.27	−0.12	0.31	2.620
4-OH	−0.37	−0.87	0.00	2.244
4-OEt	−0.24	0.38	0.31	2.650
4-OPr	−0.25	0.88	0.31	2.790
4-OBu	−0.32	1.38	0.31	2.785
4-*t*-Bu	−0.20	1.68	0.35	3.149
3-Me	−0.07	0.56	0.18	2.782
3-F	0.34	0.47	−0.74	2.665
3-Cl	0.37	1.04	−0.80	2.818

Table IV—Acetyl Transferase-Catalyzed Acylation of Substituted Anilines by Acetylaminophenylazobenzenesulfonic Acid

Substituted Aniline	Charge Density on N of Aniline[a]	σ^-	π	Acylation Rate, log A[b]
4-Br	1.849	0.23	1.13	0.049
4-Cl	1.849	0.23	0.93	0.037
4-Me	1.853	−0.17	0.48	0.0
H	1.851	0.0	0.0	−0.155
4-NO_2	1.827	1.27	0.50	−0.468
4-SO_2NH_2	1.841	0.91	−1.16	−0.745

[a] From the results of Hückel molecular orbital calculations reported by A. Perault and B. Pullman, *Biochim. Biophys. Acta*, **66**, 86(1963). [b] From the data reported by K. B. Jacobson, *J. Biol. Chem.*, **236**, 343 (1961).

analyses, it is found that μ contributes significantly, whereas σ does not, or the converse, this finding alone is not sufficient to warrant an alternative interpretation. The compounds considered may have biological potencies that more closely parallel the order of one index than they do the other, but the indexes are not sufficiently independent, on the basis of Eq. 5, to make a distinction between the two possible physical interpretations.

Recently, viral neuraminidase inhibition potencies have been reported (6) for the compounds shown in Table III. The regression equation correlating these data was given (6) by

$$\log(1/C) = \underset{(8.78)}{0.271\ (\pm 0.031)\pi} + \underset{(1.68)}{0.061\ (\pm 0.036)\mu_v} + \underset{(2.95)}{0.029\ (\pm 0.010)\mu_v^2} + 2.551 \qquad \text{(Eq. 6)}$$

N	s	R	$F(3,12)$
16	0.079	0.927	28.96

A more appropriate representation of the correlation, based on the statistics for the coefficients of Eq. 6, is expressed as

$$\log(1/C) = \underset{(8.11)}{0.265\ (\pm 0.032)\pi} + \underset{(3.67)}{0.014\ (\pm 0.003)\mu_v^2} + 2.548 \qquad \text{(Eq. 7)}$$

N	s	R	$F(2,13)$
16	0.081	0.916	36.89

Equation 7 indicates that these compounds have their potencies determined by lipophilic factors and by the component of the group dipole moment, μ_v, which is directed along the 1,4-axis of the substituted moiety.

Upon comparing the electronic index μ_v^2 to its analog σ^2, as is suggested by Eq. 5, it is found that these indexes are related:

$$\mu_v^2 = 30.235\,\sigma^2 - 1.048 \qquad \text{(Eq. 8)}$$

N	s	R	$F(1,14)$
16	2.115	0.925	83.89

A much improved correlation can be obtained by using the unresolved group dipole moments:

$$\mu^2 = 29.659\,\sigma^2 - 0.205 \qquad \text{(Eq. 9)}$$

N	s	R	$F(1,14)$
16	1.483	0.960	164.16

Since the quantity σ^2 has been indicated as having a variety of possible origins (7) and has been shown to correlate with a free-radical index E_r (8) and a charge-transfer index E_{LEMO} (9), it is found that at least three differing physical interpretations can be presented for the electronic term found in Eq. 7. The resolution of the possible alternatives is, therefore, seen as a problem separate from the use of Eq. 7, or related forms, as a predictor of new agents. The most convenient equation to use for guiding the synthesis of new compounds may contain physical parameters which only obliquely reflect the actual factors controlling the observed response.

EXAMPLE 3

Another illustration of the potential complications attending the physical interpretation of derived regression equations is provided by the data found in Table IV. A more limited statistical analysis of the enzymatic acylation rates for this system has been reported (10). Here the authors would like to point out that at least five different statistically significant regression equations will correlate the data:

$$\log A = -0.028\ (\pm 0.004)\,\mu^2 + 0.012 \qquad \text{(Eq. 10)}$$

N	s	R	$F(1,4)$
6	0.111	0.952	38.58

$$\log A = \underset{(3.54)}{0.085\ (\pm 0.024)\,\mu} + \underset{(3.40)}{0.216\ (\pm 0.063)\pi} - 0.112 \qquad \text{(Eq. 11)}$$

N	s	R	$F(2,3)$
6	0.112	0.963	24.88

$$\log A = \underset{(-15.27)}{-0.335\ (\pm 0.022)\,\sigma^-} + \underset{(16.94)}{0.252\ (\pm 0.015)\pi} - 0.155 \qquad \text{(Eq. 12)}$$

N	s	R	$F(2,3)$
6	0.028	0.997	399.2

$$\log A = \underset{(8.99)}{18.169\ (\pm 2.021)\,q_N} + \underset{(12.16)}{0.290\ (\pm 0.024)\pi} - 33.82 \qquad \text{(Eq. 13)}$$

N	s	R	$F(2,3)$
6	0.048	0.993	140.7

$$\log A = \underset{(-6.96)}{-0.272\ (\pm 0.039)\sigma^2} + \underset{(8.41)}{0.264\ (\pm 0.031)\pi} - 0.179 \qquad \text{(Eq. 14)}$$

N	s	R	$F(2,3)$
6	0.061	0.989	85.96

There is little doubt that a charge-related property is affecting the enzymatic acylation rates, according to these correlations, and it is also likely that the lipophilicity of the compounds contributes an effect. An explicit description of the electronic interaction mechanism is difficult, however, because of the various physical significances that can be attributed to the different electronic indexes. In this particular case, Eq. 13 may be preferred as a basis for interpretation, since the charge density on the aniline nitrogen is the most fundamental of the electronic indexes that can be used.

CONCLUSIONS

From the examples presented, it is clear that a physical interpretation for a correlation of biological activities with some combination of physical parameters should be made with caution. It is also evident that the ability of a regression equation to act as a predictor for the biological activities of compounds within a series cannot be presented in full support of the interpretation lent to the terms appearing in the regression equation.

At present, a safe approach to follow if the intent is to gain insight into physical factors influencing the action of drug agents is: (*a*) investigate a correlation for internal consistency, as may be illustrated by *Example 1*; (*b*) establish whether the use of fundamental linear free energy (σ, π) and molecular orbital (q, S^E, S^N) indexes lead to essentially equivalent conclusions, *e.g.*, Eqs. 12 and 13 complement one another; and (*c*) when indexes having little precedent in correlating the rates and equilibria of simple chemical systems become involved in a correlation—*viz.*, μ, μ^2, and σ^2, investigate compounds designed specifically to distinguish between alternative physical interpretations. In the latter instance, since Hammett σ-values are simply additive, whereas group dipole moments are vectorially additive, it might be expected that multisubstituted compounds should provide the more appropriate test of dipole control of a biological response.

REFERENCES

(1) J. K. Seydel, *Mol. Pharmacol.*, **2**, 259(1966).

(2) A. Cammarata and R. C. Allen, *J. Pharm. Sci.*, **56**, 640(1967).

(3) T. Fujita and C. Hansch, *J. Med. Chem.*, **10**, 991(1967).

(4) J. W. Mc Farland, *Nat. Med. Chem. Symp. Amer. Chem. Soc. Proc., 11th*, **1968**.

(5) J. W. McFarland, L. H. Conover, H. L. Howes, Jr., J. E. Lynch, D. R. Chisholm, W. C. Austin, R. L. Cornwell, J. C. Danilewicz, W. Courtney, and D. H. Morgan, *J. Med. Chem.*, **12,** 1066(1969).
(6) M. S. Tute, *ibid.*, **13,** 48(1970).
(7) A. Cammarata and S. J. Yau, *ibid.*, **13,** 94(1970).
(8) A. Cammarata, S. J. Yau, J. H. Collett, and A. N. Martin, *Mol. Pharmacol.*, **6,** 65(1970).
(9) R. C. Allen and C. Cavallito, to be published.
(10) C. Hansch, E. W. Deutsch, and R. N. Smith, *J. Amer. Chem. Soc.*, **87,** 2738(1965).

ACKNOWLEDGMENTS AND ADDRESSES

Received February 11, 1970, from **Temple University, School of Pharmacy, Laboratory of Physical Medicinal Chemistry, Philadelphia, PA 19140,* and the *†Chemistry Department, Borstel Research Institute for Experimental Biology and Medicine, 2061 Borstel, Germany.*

Accepted for publication April 1, 1970.

Abstracted in part from a thesis presented by R. C. Allen to the Medical College of Virginia, Richmond, Va., in partial fulfillment of the Doctor of Philosophy degree requirements.

Solubilization of Some Steroid Hormones in Aqueous Solutions of Bile Salts

ARVIND L. THAKKAR

Abstract □ Solubilities of testosterone propionate, methyltestosterone, and 19-nortestosterone in aqueous sodium cholate and deoxycholate were determined. Solubilizing capacity values show that deoxycholate is a better solubilizer than cholate and that both bile salts solubilize more 19-nortestosterone than other testosterone derivatives. A possible mode of solubilization is discussed.

Keyphrases □ Steroid hormones—solubilization, aqueous solutions, bile salts □ Testosterone derivatives—solubilizing effect of sodium cholate, deoxycholate □ UV spectrophotometry—analysis

The ability of bile salts to enhance the water solubility of steroid hormones was noted as early as 1944 by Cantarow *et al.* (1). Since that time, micellar solubilization of steroids has been studied extensively by Ekwall (2) and Sjöblom (3). However, bile salt solubilization of hormonal steroids appears not to have been examined in detail. This study was undertaken to examine the solubilization of testosterone propionate, methyltestosterone, and 19-nortestosterone by the anionic surfactants, sodium cholate and sodium deoxycholate, and is part of a larger study of the solubilization of steroidal hormones by steroidal surfactants. A recent report from this laboratory (4) dealt with the solubilization of some androgenic steroids by ethoxylated cholesterol, a nonionic surfactant.

EXPERIMENTAL

Materials—Sodium cholate,[1] sodium deoxycholate,[1] methyltestosterone NF, testosterone propionate USP, and 19-nortestosterone[2] were used as received. Moisture contents of the bile salts, determined by drying overnight *in vacuo* at 110°, were taken into consideration when recording their weights. Bile salt solutions, prepared with distilled water, were not buffered; their pH ranged from 7.0 to 7.8 for cholate and from 7.2 to 8.6 for deoxycholate.

Solubility Determinations—Solubilities were determined by equilibration of several concentrations of bile salt solutions with the steroids, followed by spectrophotometric analyses of suitably diluted aliquots, as described previously (4). For solutions in which enhancement of steroid solubility was minimal, dilution with 50% (v/v) methanol was still necessary to lower the bile salt concentration to a point where it would not interfere with the UV spectrophotometric analytical procedure. In such cases, cells of 5-cm. pathlength were used. 19-Nortestosterone, which was not included in the previous study, has maximum absorbance in 50% (v/v) aqueous methanol at 244 mμ, with a molar absorption coefficient of 17.3×10^3.

[1] Special enzyme grade, Mann Research Laboratories, Inc., New York, N. Y.
[2] Purchased from Organon, Inc., West Orange, N. J.

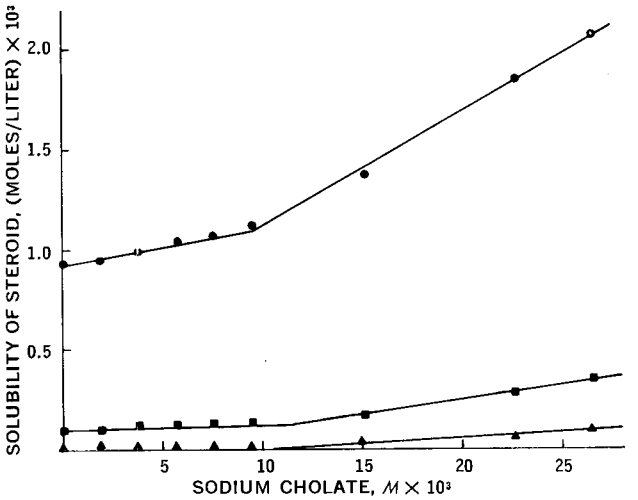

Figure 1—*Solubility of steroids in aqueous solutions of sodium cholate at 30°. Key:* ●, *19-nortestosterone;* ■, *methyltestosterone; and* ▲, *testosterone propionate.*

RESULTS AND DISCUSSION

Figures 1 and 2 show the relationship between solubility of the steroid solubilizates and the concentration of sodium cholate and sodium deoxycholate, respectively. At low concentrations of bile salts, only a marginal change in steroid solubility is observed. After these initial stages, up to $\sim 1.0 \times 10^{-2}$ *M* for sodium cholate and $\sim 6.0 \times 10^{-3}$ *M* for sodium deoxycholate, steroid solubility increases linearly with bile salt concentration. This behavior conforms well to the general features of micellar solubilization, but it is at variance with the report of Lach and Pauli (5) who found that the solubility of testosterone increased at a higher rate below the apparent critical micelle concentration (CMC) of deoxycholate than above it. In a comprehensive paper dealing with the solubiliza-