

# Development of quantitative structure–property relationship models for pseudoternary microemulsions formulated with nonionic surfactants and cosurfactants: application of data mining and molecular modeling

Mutasem O. Taha<sup>a,\*</sup>, Mutasim Al-Ghazawi<sup>b</sup>, Hend Abu-Amara<sup>c</sup>, Enam Khalil<sup>d</sup>

<sup>a</sup>Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan

<sup>b</sup>Department of Biopharmaceutics and Clinical Pharmacy, Faculty of Pharmacy, University of Jordan, Amman, Jordan

<sup>c</sup>Department of Pharmaceutics and Industrial Pharmacy, Al-Azhar University, Gaza, Palestinian National Authority, Palestine

<sup>d</sup>Department of Pharmaceutics and Pharmaceutical Technology, Faculty of Pharmacy, University of Jordan, Amman, Jordan

Received 5 September 2001; received in revised form 18 March 2002; accepted 25 March 2002

## Abstract

Data mining, computer-aided molecular modeling, descriptor calculation and multiple linear regression techniques were utilized to produce statistically significant and predictive models for O/W and W/O microemulsions. The literature was scanned over the last 20 years, subsequently, 68 phase diagrams from eight different references were collected. Molecular modeling techniques were then applied on the components of the microemulsion systems to generate plausible 3-D structures. Subsequently, various physicochemical descriptors were calculated based on the resulting 3-D structures. The generated descriptors were correlated with microemulsion existence areas utilizing multiple linear regression analysis (MLR). The generated models were statistically cross-validated and were found to be of significant predictive power. Furthermore, the resulting models allowed better understanding of the process of microemulsion formation. © 2002 Elsevier Science B.V. All rights reserved.

**Keywords:** QSPR; Microemulsions; Molecular modeling; Physicochemical descriptors; Multiple linear regression; Cross-validation

## 1. Introduction

Microemulsions are homogenous, transparent, isotropic, thermodynamically stable dispersions of water and oil (Warisnoicharoen et al., 2000). Such systems have received recent interest as potential vehicles for transdermal and oral drug delivery (Friman and Bäckman, 1996; Ho et al., 1996; Tenjarla, 1999; Watnasirichaikul et al., 2000).

Microemulsions are of three types: water-in-oil (W/O), oil-in-water (O/W) or bicontinuous (Tenjarla, 1999). They might be stabilized either by single surfactant (nonionic or anionic), mixture of surfactants, or cosurfactant/surfactant combination (Friberg, 1990; Constantinides and Scalart, 1997). Three-component microemulsions (i.e. stabilized by surfactant(s) only) are generally known as tertiary microemulsions, while those based on four components (i.e. with cosurfactant) are known as pseudoternary (or quaternary) microemulsions (Tenjarla, 1999).

Microemulsions are graphically represented as stability areas in triangular phase diagrams (Kreuter, 1994), where each triangular corner designates a certain component. An example phase diagram is illustrated in Fig. 1 (Attwood et al., 1992).

Few theories tried to explain microemulsion formation (Kreuter, 1994). However, the most famous is the geometrical packing theory (Israelachvili et al., 1976), which depicts microemulsions as tiny droplets of internal phase (ca. 200 nm) dispersed in the continuous phase and stabilized by efficient steric packing of surfactant molecules at the oil/water interface (Israelachvili et al., 1976; Friberg, 1990). However, this theory lacks explicit elucidation of the roles of oils and cosurfactants in the stabilization process (Tenjarla, 1999; Friberg, 1990). Oils were recently qualitatively implicated in the interfacial packing (Warisnoicharoen et al., 2000).

Nevertheless, despite the geometrical packing theory and the accumulation of experimental data concerning microemulsions, their formulation is still highly empirical and time consuming (Attwood et al., 1992; Aboofazeli et

\*Corresponding author. Tel.: +962-6-535-5000x2505.

E-mail address: mutasem@ju.edu.jo (M.O. Taha).

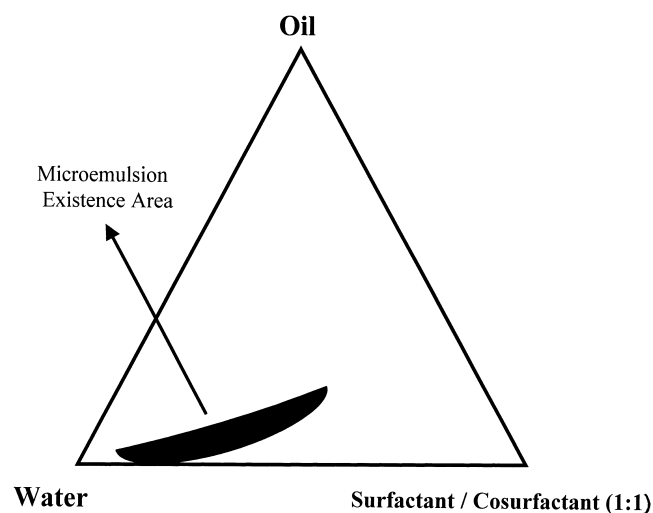


Fig. 1. Pseudoternary phase diagram illustrating a microemulsion existence area, containing: isopropyl myristate (oil), Tween 40 (surfactant), and sorbitol (cosurfactant) (Attwood et al., 1992).

al., 1994; Constantinides and Scalart, 1997). Accordingly, we were prompted to develop quantitative structure–property relationship (QSPR) models that can help cut down the trial time required in the preparation of microemulsions.

The geometrical packing theory allowed us to envisage microemulsion formation to be mediated via affinity binding processes between surfactant molecules at the interfacial film. Other microemulsion components (oils and cosurfactants) can be also considered to pack at the interface, influencing the integrity of the interfacial film.

Affinity processes are based on electrostatic interactions (e.g., H-bonding and dipole–dipole) and van der Waals dispersion forces (Martin, 1993). The steric bulkiness of interfacial film components is expected to increase interatomic distances, thus minimizing affinity interactions (Martin, 1993). Accordingly, it is anticipated that microemulsion stability should significantly correlate with electrostatic, dispersion and steric properties of the different components.

In the current study, phase diagrams corresponding to pseudoternary microemulsions, stabilized by nonionic surfactants were mined from published literature. Such systems have recently received great pharmaceutical interest due to their non-irritant and non-toxic properties (Aboofazeli et al., 1994; Thevenin et al., 1996).

Subsequently, computer-aided molecular modeling techniques were utilized to calculate various electrostatic, dispersion and steric descriptors for the different microemulsion components. Afterwards, multiple linear regression techniques were utilized to correlate the generated descriptors with microemulsion stability areas. The resulting models were utilized in probing the mechanism of microemulsion formation. The current study represents the

first attempt to develop QSPR models related to microemulsion formation and stability.

## 2. Experimental: data and calculation methods

### 2.1. Software

1. Alchemy 2000, 2.05, Tripos Inc. (<http://www.tripos.com/software/alchemy1.html>), USA (Tripos, 1998).
2. SciQSAR 3.0, Scivision (<http://www.scivision.com/SciQSAR.html>), USA (SciVision, 1999).
3. Pallas 2.0, CompuDrug Chemistry Ltd. (<http://www.compudrug.com>), USA (Compudrug, 1995).
4. SAS, Version 4.0 for Windows Release 6.12, SAS Institute Inc. (<http://www.sas.com>), USA.

### 2.2. Data-mining

The literature was surveyed over the last 20 years. Clearly defined and illustrated pseudoternary phase diagrams, corresponding to O/W and W/O microemulsions, produced at room temperature, using nonionic constituents, were collected for modeling. The percentage of microemulsion existing area (ME%) was determined by the cut and weight method (Kale and Allen, 1989). Table 1 illustrates the selected microemulsion phase diagrams, their components, mass ratios, percent areas (ME%) and cited references. The following definitions were adapted in this work for classifying the constituents of the microemulsion systems.

- Surfactants are relatively large molecules (350–1312 Da) possessing two large, structurally distinct parts illustrating opposite lipophilicity/hydrophilicity properties (Aulton, 1990; Wade and Weller, 1994).
- Cosurfactants are defined as small (46–186 Da) mono- or multi-hydroxy alcohols added to stabilize microemulsions (Kreuter, 1994). Literature articles defining large complex molecules as cosurfactants, e.g., Myverol 18-99 and Arlacel 186 (Constantinides and Scalart, 1997) were excluded.
- Oils are defined as small to moderately large (140–310 Da) alkyl or arylalkyl hydrocarbons that might contain one or few ester or ether linkages.

### 2.3. Molecular modeling

The two-dimensional (2-D) chemical structures of various cosurfactants and oils were generated from their corresponding chemical names or collected from references (Weast, 1989; Wade and Weller, 1994; Delgado and Remers, 1998). However, the chemical structures of different surfactants were collected from different sources. Table 2 gives the generic names and chemical compositions of the relevant surfactants. Subsequently, the 2-D

Table 1  
The selected microemulsion systems, their components, ratios, and corresponding references

No.	Surfactant	Cosurfactant	Oil	$K_m$	Other ratios	O/W area (%)	W/O area (%)	Reference	Reference no. <sup>a</sup>
1	Tween 80	Sorbitol	Isopropylmyristate (IPM)	1/1.5	–	5.47	–	Ktistis, 1990	1
2	Tween 80	Sorbitol	IPM	1/2	–	8.93	–	Ibid	1
3	Tween 80	Sorbitol	IPM	1/2.5	–	10.41	–	Ibid	1
4	Tween 80	Sorbitol	IPM	1/3	–	4.52	–	Ibid	1
5	Tween 80	Sorbitol	IPM	1/3.5	–	3.14	–	Ibid	1
6	Tween 80	Sorbitol	IPM	1/4	–	1.87	–	Ibid	1
7	Tween 60	Sorbitol	IPM	1/1.5	–	3.84	–	Attwood and Ktistis, 1989	2
8	Tween 60	Sorbitol	IPM	1/2	–	9.69	–	Ibid	2
9	Tween 60	Sorbitol	IPM	1/2.5	–	6.62	–	Ibid	2
10	Tween 60	Sorbitol	IPM	1/3	–	4.58	–	Ibid	2
11	Tween 60	Sorbitol	IPM	1/3.5	–	1.66	–	Ibid	2
12	Tween 60	Sorbitol	IPM	1/4	–	0.73	–	Ibid	2
13	Tween 40	Sorbitol	IPM	1/1	–	7.47	–	Attwood et al., 1992	3
14	Tween 40	Sorbitol	IPM	1/1.5	–	11.55	–	Ibid	3
15	Tween 40	Sorbitol	IPM	1/2	–	8.49	–	Ibid	3
16	Tween 40	Sorbitol	IPM	1/2.5	–	6.88	–	Ibid	3
17	Tween 40	Sorbitol	IPM	1/3	–	4.63	–	Ibid	3
18	Tween 40	Sorbitol	IPM	1/3.5	–	2.34	–	Ibid	3
19	Brij 76	<i>n</i> -butanol	<i>n</i> -decane	–	Alcohol/oil=1	3.60	22.1	Garti et al., 1995	4
20	Brij 76	<i>n</i> -butanol	<i>n</i> -tetradecane	–	Alcohol/oil=1	–	33.9	Ibid	4
21	Brij 76	<i>n</i> -butanol	<i>n</i> -hexadecane	–	Alcohol/oil=1	–	25.3	Ibid	4
22	Brij 76	Ethanol	<i>n</i> -tetradecane	–	Alcohol/oil=1	–	7.8	Ibid	4
23	Brij 76	1-propanol	<i>n</i> -tetradecane	–	Alcohol/oil=1	–	12.8	Ibid	4
24	Brij 76	1-pentanol	<i>n</i> -tetradecane	–	Alcohol/oil=1	0.80	13.9	Ibid	4
25	Brij 76	1-hexanol	<i>n</i> -tetradecane	–	Alcohol/oil=1	0.80	12.0	Ibid	4
26	Brij 97	Ethanol	<i>n</i> -tetradecane	–	Alcohol/oil=1	3.20	11.3	Ibid	4
27	Brij 97	1-propanol	<i>n</i> -tetradecane	–	Alcohol/oil=1	2.90	22.7	Ibid	4
28	Brij 97	1-pentanol	<i>n</i> -tetradecane	–	Alcohol/oil=1	3.90	33.3	Ibid	4
29	Brij 97	1-hexanol	<i>n</i> -tetradecane	–	Alcohol/oil=1	0.80	22.3	Ibid	4
30	Brij 76	Ethanol	<i>n</i> -dodecane	–	Alcohol/oil=1	–	8.3	Ibid	4
31	Brij 76	Ethanol	<i>n</i> -hexadecane	–	Alcohol/oil=1	–	3.2	Ibid	4
32	Brij 76	1-Propanol	<i>n</i> -dodecane	–	Alcohol/oil=1	0.50	21.5	Ibid	4
33	Brij 76	1-Butanol	<i>n</i> -dodecane	–	Alcohol/oil=1	–	28.9	Ibid	4
34	Brij 76	1-Pentanol	<i>n</i> -dodecane	–	Alcohol/oil=1	0.60	12.2	Ibid	4
35	Brij 76	1-Hexanol	<i>n</i> -dodecane	–	Alcohol/oil=1	0.20	13.5	Ibid	4
36	Brij 76	1-Heptanol	<i>n</i> -dodecane	–	Alcohol/oil=1	–	7.1	Ibid	4
37	Brij 76	1-Octanol	<i>n</i> -dodecane	–	Alcohol/oil=1	1.30	4.6	Ibid	4
38	Brij 76	1-Decanol	<i>n</i> -dodecane	–	Alcohol/oil=1	–	4.7	Ibid	4
39	Brij 76	1-Dodecanol	<i>n</i> -dodecane	–	Alcohol/oil=1	–	3.5	Ibid	4
40	BO-TX10	Ethanol	<i>n</i> -dodecane	–	Alcohol/oil=1	4.60	20.0	Ibid	4
41	Brij 97	1-Propanol	<i>n</i> -dodecane	–	Alcohol/oil=1	4.30	28.3	Ibid	4
42	BO-TX10	1-Pentanol	<i>n</i> -dodecane	–	Alcohol/oil=1	5.90	40.9	Ibid	4
43	Brij 97	1-Hexanol	<i>n</i> -dodecane	–	Alcohol/oil=1	0.30	39.8	Ibid	4
44	Brij 97	1-Heptanol	<i>n</i> -dodecane	–	Alcohol/oil=1	0.20	37.0	Ibid	4
45	BO-TX10	1-Octanol	<i>n</i> -dodecane	–	Alcohol/oil=1	1.20	30.7	Ibid	4
46	Brij 97	1-Decanol	<i>n</i> -dodecane	–	Alcohol/oil=1	0.50	22.8	Ibid	4
47	Brij 97	1-Dodecanol	<i>n</i> -dodecane	–	Alcohol/oil=1	0.60	18.8	Ibid	4
48	Span 20 and Tween 80 (2/3)	1-Propanol	Ethyl oleate	7/3	–	–	15.0	Alany et al., 2000	5
49	Span 20 and Tween 80 (2/3)	1-Butanol	Ethyl oleate	7/3	–	–	43.0	Ibid	5
50	Span 20 and Tween 80 (2/3)	1-Hexanol	Ethyl oleate	7/3	–	–	23.5	Ibid	5
51	Span 20 and Tween 80 (2/3)	1-Octanol	Ethyl oleate	7/3	–	–	1.5	Ibid	5
52	Span 20 and Tween 80 (2/3)	1,2-Propanediol	Ethyl oleate	7/3	–	–	11.5	Ibid	5
53	Span 20 and Tween 80 (2/3)	1,2-Pentandiol	Ethyl oleate	7/3	–	–	30.0	Ibid	5
54	Span 20 and Tween 80 (2/3)	1,2-Hexandiol	Ethyl oleate	7/3	–	–	43.0	Ibid	5
55	Span 20 and Tween 80 (2/3)	1,2-Octandiol	Ethyl oleate	7/3	–	–	48.0	Ibid	5

Table 1. Continued

No.	Surfactant	Cosurfactant	Oil	$K_m$	Other ratios	O/W area (%)	W/O area (%)	Reference	Reference no. <sup>a</sup>
56	Brij 96	Ethylene glycol	Mineral oil	3/1	–	6.18	7.79	Kale and Allen, 1989	6
57	Brij 96	Ethylene glycol	Mineral oil	9/1	–	6.84	9.72	Ibid	6
58	Brij 96	Propylene glycol	Mineral oil	3/1	–	5.74	3.37	Ibid	6
59	Brij 96	Propylene glycol	Mineral oil	9/1	–	5.95	8.12	Ibid	6
60	Brij 96	Glycerin	Mineral oil	1/1	–	5.59	3.79	Ibid	6
61	Brij 96	Ethylene glycol	Mineral oil	1/3	–	9.90	–	Ibid	6
62	Brij 96	Ethylene glycol	Mineral oil	1/1	–	7.66	–	Ibid	6
63	Brij 96	Propylene glycol	Mineral oil	1/1	–	1.59	–	Ibid	6
64	Brij 96	Glycerin	Mineral oil	1/3	–	12.27	–	Ibid	6
65	Brij 96	Glycerin	Mineral oil	3/1	–	6.17	–	Ibid	6
66	Brij 96	Glycerin	Mineral oil	9/1	–	6.30	–	Ibid	6
67	Tween 60	<i>n</i> -pentanol	Hexadecane	1.94/1	–	33.21	–	Hermansky and Mackay, 1980.	7
68	Tween 40	<i>n</i> -pentanol	Mineral oil	3.55/1	–	9.03	–	Mackay et al., 1980	8

<sup>a</sup> Reference numbers used in Table 3.

representations of different components were sketched using Alchemy 2000 2-D-sketcher (ChemPrint Pro, 07.18). Afterwards, reasonable three-dimensional (3-D) structures were generated using rule-based methods implemented in the 2-D to 3-D-Builder within Alchemy 2000. At this point each 3-D surfactant structure was cleaved into two parts: the hydrophilic segment (HS) and the lipophilic segment (LS). The cleavage was carried out in such a way as to include all oxygen functional groups into the HS, while the carbon tails were included in the LS. Hydrogen atoms were added at the division points. Fig. 2 shows an example cleavage carried out on Tween 80. The resulting 3-D structures were further optimized using the Alchemy 2000 force field with cut-off values of RMS=0.15 and  $\Delta E=0.0005$  (Tripos, 1998).

#### 2.4. Calculation of molecular descriptors

A group of physicochemical descriptors were calculated

for each component per system, i.e. oils, surfactants and cosurfactants, using Pallas and SciQSAR interfaced with Alchemy 2000. Pallas predictor was utilized to predict Log $P$  values based on structure fragments approach (Pallas reference manual. CompuDrug, 1995) for whole surfactant molecules, while the SciQSAR-based descriptors were calculated separately for the LS and HS parts of each surfactant molecule. SciQSAR generates 16 descriptors for each entry (SciVision, 1999):

1. Charge-related descriptors calculated utilizing charges based on the empirical model of partial equalization of orbital electronegativity developed by Gasteiger and Marsili (Gasteiger and Marsili, 1980). These include:
  - The sum of absolute values of the charges on each atom of the molecule in electrons (ABSQ)
  - The sum of absolute values of the charges on nitrogen and oxygen atoms in the molecule (ABSQon)
  - Dipole moment of the molecule (*Dipole*), based on

Table 2

The chemical names and compositions of different surfactants considered in this study

No.	Generic name	Chemical composition	Reference
1	Tween 80	Polyoxyethylene(20) Sorbitan monooleate	Aulton, 1990
2	Tween 60	Polyoxyethylene(20) Sorbitan monostearate	Aulton, 1990
3	Tween 40	Polyoxyethylene(20) Sorbitan monopalmitate	Aulton, 1990
4	Span 20	Sorbitan monolaurate	Aulton, 1990
5	Brij 96(97) or BO-TX 10	Polyoxyethylene(10) oleyl ether	Wade and Weller, 1994; Garti et al., 1995
6	Brij 76	Polyoxyethylene(10) stearyl ether	Wade and Weller, 1994

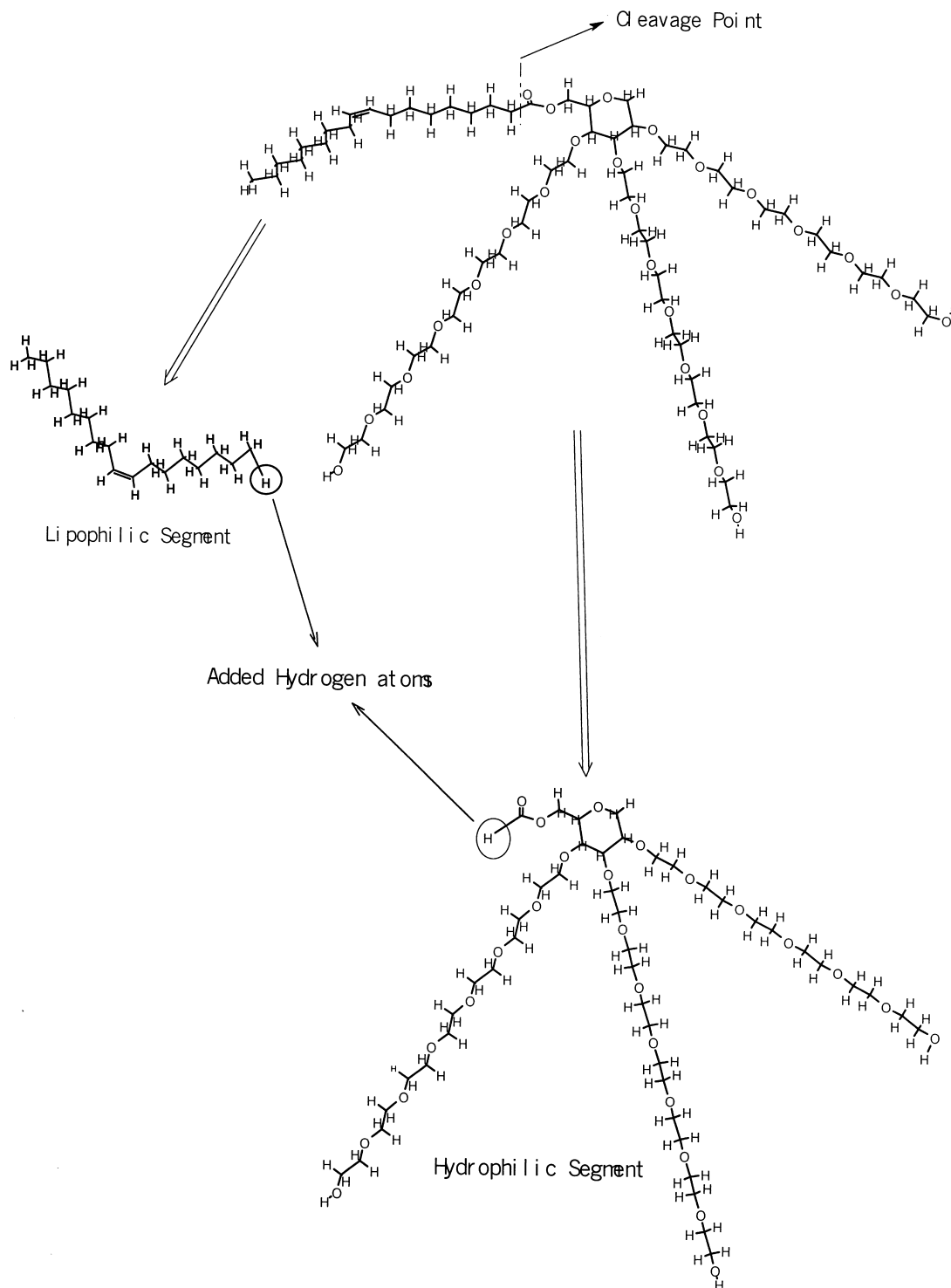


Fig. 2. A scheme illustrating the virtual cleavage carried out on Tween 80 into hydrophilic and lipophilic segments prior to descriptors' calculations. A similar procedure was carried out on other surfactants.

the molecular 3-D structure and component atomic charges. Units are in Debyes

- The largest negative charge over the atoms in a molecule (MaxQ<sup>-</sup>)
- The largest positive charge over the atoms in a molecule (MaxQ<sup>+</sup>)

2. Size and shape-related descriptors, which include.

- Molecular weight (MolWeight) in Daltons
- Molecular volume (*Volume*): a 3-D-dependent descriptor computed by the grid method of Bodor (Bodor et al., 1989)
- Kappa Alpha 3 ( $\kappa_{\alpha_3}$ ): third order shape index that

can discern isomers of the same molecule (Kier, 1985; SciVision, 1999)

- Wiener Index (WI): a dimensionless, 3-D-independent topological parameter formulated by Wiener (Wiener, 1947)
  - ${}^1\chi$ ,  ${}^3\chi$ ,  ${}^0\chi^V$  and  ${}^1\chi^V$ : are a group of connectivity descriptors that encode the 2-D structure of a molecule (Kier and Hall, 1986; Katritzky and Gordeeva, 1993)
3. Molecular polarizability descriptors:
- Molecular polarizability (Polar), also known as alpha ( $\alpha$ ), calculated based on the additive approach given by Miller (Miller, 1990). Units are  $\text{\AA}^3$ .
  - Specific polarizability (*Sp.Pol.*) defined as polarizability per unit volume
4. Calculated logarithm of the octanol/water partition coefficient (*LogP*), based on the neural network approach developed by Bodor (Bodor et al., 1989; Bodor and Huang, 1992).

The prefixes O-, Co-, S-, HS- and LS- were added to the different descriptor abbreviations to denote oil, cosurfactant and surfactant (HS or LS) descriptors, respectively, e.g., O-Volume denotes the volume of oil molecules.

### 2.5. Calculation of physicochemical descriptors related to surfactant mixtures

If a certain microemulsion is stabilized by a mixture of surfactants, e.g., systems 48–55 in Table 1, any particular mixture's descriptor was calculated as the average value of the respective components' descriptors according to their individual percentages, as illustrated in the following equation.

$$\text{Mixture descriptor} = R_1 \times D_1 + R_2 \times D_2$$

Where  $R_1$  is the ratio (w/w) of the first surfactant;  $D_1$  is the calculated descriptor of the first surfactant;  $R_2$  is the ratio (w/w) of the second surfactant and  $D_2$  is the calculated descriptor of the second surfactant. The HS-Volume for combined Span 20 and Tween 80 mixture (40:60 ratio) is shown here as an example:

$$\text{HS-Volume (Span 20/Tween 80)} = 0.4 \times 164.12 + 0.6 \times 1018.75 = 676.898 \text{ \AA}^3$$

The validity of this assumption is supported by the commonly accepted principle of calculating the hydrophile–lipophile balance (HLB) values of surfactant mixtures based on the average HLB values of their components (Attwood and Florence, 1983).

### 2.6. Mass ratio-based descriptors

The fixed ratio of two components in a pseudoternary

microemulsion system is an influential variable on microemulsion existence area (see Section 3.3). The fixed percentage of a particular surfactant in a mixture with a cosurfactant is expressed by the 'surfactant ratio' (SR) descriptor. Similarly, cosurfactant ratio (CoSR) and oil ratio (OR) were added to express the fixed ratios of cosurfactants or oils in given mixtures, respectively. The ratios were collected from triangle corners of different microemulsion phase diagrams.

### 2.7. Statistical modeling

The collected phase diagrams were classified into water-in-oil (W/O) and oil-in-water (O/W) microemulsion categories. It was assumed that phase diagrams illustrating W/O regions completely lack O/W areas (i.e. zero percent areas) and vice versa, unless the considered diagram shows two clearly defined W/O and O/W areas.

#### 2.7.1. Descriptor selection prior to multiple linear regression (MLR)

The generated molecular descriptors were too numerous (approximately 70) to allow significant MLR models. Accordingly, a two-stage manual descriptor selection process was carried out.

##### 2.7.1.1. Selection of molecular descriptors based on the graphical interface of SciQSAR

Many collinear SciQSAR-generated descriptors were filtered based on the graphical interface within the program. Collinear descriptors produced correlating line trends. Accordingly, a particular descriptor was chosen to represent a group of apparently highly collinear variables. Furthermore, it was decided to exclude *LogP* values calculated by SciQSAR in favor of the values calculated by the Pallas predictor, since it was not possible to calculate the *LogP* values of some surfactants (i.e. Tween 80 and Tween 60) using the *LogP* function of SciQSAR due to their excessive large sizes. Subsequently, the selected molecular descriptors and their associated mass ratio descriptors were transferred to an excel spreadsheet. The corresponding microemulsion percent areas (O/W-ME% and W/O-ME%) were recorded as the dependent variables. Table 3 shows the selected molecular descriptors (at this stage) and the associated percent areas.

##### 2.7.1.2. Descriptor selection based on correlation table (Table 4)

A subsequent, more rigorous descriptor-selection stage was carried out based on correlation Table 4. In this stage, any two descriptors within the same component class (i.e., surfactants, cosurfactants or oils) exhibiting cross-correlation  $r^2 > 0.65$ , were considered to be collinear (SciVision, 1999; Martin, 1978) and subsequently one of them was excluded. The selection was mainly based on the respective *F*-values of the rival descriptors. Statistically signifi-

Table 3

The different mined microemulsion systems with some of their corresponding molecular descriptors and microemulsion percent areas

No.	Ref. no.	System	SR (%)	HS- Volume	HS- $\kappa_{\alpha 3}$	HS-Dipole ABSQon	HS- ABSQon	HS-Sp.Pol. Volume	LS- Volume	LS- $\kappa_{\alpha 3}$	LS-Dipole	LS-Sp.Pol.	S-LogP	CoSR (%)	Co- Volume	Co- $\kappa_{\alpha 3}$	Co-Dipole AbsQon	Co- AbsQon	Co-Sp.Pol. LogP	Co- LogP	OR (%)	O- Volume	O- $\kappa_{\alpha 3}$	O-Dipole AbsQon	O- Sp.Pol. LogP	O- LogP	W/O- ME%	O/W- ME%	
1	1	Tween 80 + sorbitol + IPM	40	1018.75	48.76	3.616648	9.65	0.100374	293.61	15.74	0.134014	0.108229	5.79	60	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	5.47
2	1	Tween 80 + sorbitol + IPM	33.4	1018.75	48.76	3.616648	9.65	0.100374	293.61	15.74	0.134014	0.108229	5.79	66.67	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	8.93
3	1	Tween 80 + sorbitol + IPM	28.6	1018.75	48.76	3.616648	9.65	0.100374	293.61	15.74	0.134014	0.108229	5.79	71.43	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	10.41
4	1	Tween 80 + sorbitol + IPM	25	1018.75	48.76	3.616648	9.65	0.100374	293.61	15.74	0.134014	0.108229	5.79	75	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	4.52
5	1	Tween 80 + sorbitol + IPM	22.2	1018.75	48.76	3.616648	9.65	0.100374	293.61	15.74	0.134014	0.108229	5.79	77.78	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	3.14
6	1	Tween 80 + sorbitol + IPM	20	1018.75	48.76	3.616648	9.65	0.100374	293.61	15.74	0.134014	0.108229	5.79	80	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	1.87
7	2	Tween 60 + sorbitol + IPM	40	1018.69	48.76	3.626469	9.65	0.10038	300.12	16.00	0.010767	0.106521	6.31	60	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	3.84
8	2	Tween 60 + sorbitol + IPM	33.4	1018.69	48.76	3.626469	9.65	0.10038	300.12	16.00	0.010767	0.106521	6.31	66.67	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	9.69
9	2	Tween 60 + sorbitol + IPM	28.6	1018.69	48.76	3.626469	9.65	0.10038	300.12	16.00	0.010767	0.106521	6.31	71.43	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	6.62
10	2	Tween 60 + sorbitol + IPM	25	1018.69	48.76	3.626469	9.65	0.10038	300.12	16.00	0.010767	0.106521	6.31	75	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	4.58
11	2	Tween 60 + sorbitol + IPM	22.2	1018.69	48.76	3.626469	9.65	0.10038	300.12	16.00	0.010767	0.106521	6.31	77.78	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	1.66
12	2	Tween 60 + sorbitol + IPM	20	1018.69	48.76	3.626469	9.65	0.10038	300.12	16.00	0.010767	0.106521	6.31	80	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	0.73
13	3	Tween 40 + sorbitol + IPM	33.34	1018.80	48.76	3.546863	9.65	0.100369	266.00	14.00	0.01103	0.106387	5.25	66.66	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	8.49
14	3	Tween 40 + sorbitol + IPM	50	1018.80	48.76	3.546863	9.65	0.100369	266.00	14.00	0.01103	0.106387	5.25	50	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	7.47
15	3	Tween 40 + sorbitol + IPM	40	1018.80	48.76	3.546863	9.65	0.100369	266.00	14.00	0.01103	0.106387	5.25	60	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	11.55
16	3	Tween 40 + sorbitol + IPM	28.6	1018.80	48.76	3.546863	9.65	0.100369	266.00	14.00	0.01103	0.106387	5.25	71.43	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	6.88
17	3	Tween 40 + sorbitol + IPM	25	1018.80	48.76	3.546863	9.65	0.100369	266.00	14.00	0.01103	0.106387	5.25	75	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	4.63
18	3	Tween 40 + sorbitol + IPM	22.2	1018.80	48.76	3.546863	9.65	0.100369	266.00	14.00	0.01103	0.106387	5.25	77.78	162.98	3.36	2.36331	2.33	0.095752	-4.17	100	310.26	15.57	1.390134	0.61	0.10537	7.38	0	2.34
19	4	Brij 76 + butanol + dodecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	87.84	3.96	1.463911	0.40	0.099621	1.04	50	181.39	9.14	0.000015	0.00	0.10543	5.51	22.1	3.6
20	4	Brij 76 + pentanol + tetradecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	104.79	5.30	1.473282	0.40	0.101017	1.55	50	249.23	13.09	0.000013	0.00	0.10618	7.55	13.9	0.8
21	4	Brij 76 + hexanol + tetradecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	121.78	5.96	1.463851	0.40	0.101998	2.05	50	249.23	13.09	0.000013	0.00	0.10618	7.55	12	0.8
22	4	Brij 97 + ethanol + tetradecane	100	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	53.90	0.00	1.459151	0.40	0.09426	0.02	50	249.23	13.09	0.000013	0.00	0.10618	7.55	11.3	3.2
23	4	Brij 97 + propanol + tetradecane	100	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	70.91	4.00	1.47441	0.40	0.097532	0.53	50	249.23	13.09	0.000013	0.00	0.10618	7.55	22.7	2.9
24	4	Brij 97 + pentanol + tetradecane	100	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	104.79	5.30	1.473282	0.40	0.101017	1.55	50	249.23	13.09	0.000013	0.00	0.10618	7.55	33.3	3.9
25	4	Brij 97 + hexanol + tetradecane	100	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	121.78	5.96	1.463851	0.40	0.101998	2.05	50	249.23	13.09	0.000013	0.00	0.10618	7.55	22.3	0.8
26	4	Brij 76 + propanol + dodecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	70.91	4.00	1.47441	0.40	0.097532	0.53	50	215.30	11.11	0.00001	0.00	0.10587	6.53	21.5	0.5
27	4	Brij 76 + pentanol + dodecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	104.79	5.30	1.473282	0.40	0.101017	1.55	50	215.30	11.11	0.00001	0.00	0.10587	6.53	12.2	0.6
28	4	Brij 76 + hexanol + dodecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	121.78	5.96	1.463851	0.40	0.101998	2.05	50	215.30	11.11	0.00001	0.00	0.10587	6.53	13.5	0.2
29	4	Brij 76 + octanol + dodecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	155.75	7.96	1.46385	0.40	0.103315	3.07	50	215.30	11.11	0.00001	0.00	0.10587	6.53	4.6	1.3
30	4	BO-TX10 + ethanol + dodecane	100	447.48	29.56	0.988846	4.16	0.099403	310.47	16.81	0.124866	0.10826	6.05	50	53.90	0.00	1.459151	0.40	0.09426	0.02	50	215.30	11.11	0.00001	0.00	0.10587	6.53	20	4.6
31	4	Brij 97 + propanol + dodecane	100	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	70.91	4.00	1.47441	0.40	0.097532	0.53	50	215.30	11.11	0.00001	0.00	0.10587	6.53	28.3	4.3
32	4	BO-TX10 + pentanol + dodecane	100	447.48	29.56	0.988846	4.16	0.099403	310.47	16.81	0.124866	0.10826	6.05	50	104.79	5.30	1.473282	0.40	0.101017	1.55	50	215.30	11.11	0.00001	0.00	0.10587	6.53	40.9	5.9
33	4	Brij 97 + hexanol + dodecane	100	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	121.78	5.96	1.463851	0.40	0.101998	2.05	50	215.30	11.11	0.00001	0.00	0.10587	6.53	39.8	0.3
34	4	Brij 97 + heptanol + dodecane	100	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	138.78	7.16	1.473255	0.40	0.102726	2.56	50	215.30	11.11	0.00001	0.00	0.10587	6.53	37	0.2
35	4	BO-TX10 + octanol + dodecane	100	447.48	29.56	0.988846	4.16	0.099403	310.47	16.81	0.124866	0.10826	6.05	50	155.75	7.96	1.46385	0.40	0.103315	3.07	50	215.30	11.11	0.00001	0.00	0.10587	6.53	30.7	1.2
36	4	Brij 97 + decanol + dodecane	100	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	189.86	9.96	1.463825	0.40	0.104081	4.09	50	215.30	11.11	0.00001	0.00	0.10587	6.53	22.8	0.5
37	4	Brij 97 + dodecanol + dodecane	100	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	223.83	11.96	1.463798	0.40	0.10468	5.11	50	215.30	11.11	0.00001	0.00	0.10587	6.53	18.8	0.6
38	4	Brij 76 + ethanol + tetradecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	53.90	0.00	1.459151	0.40	0.09426	0.02	50	249.23	13.09	0.000013	0.00	0.10618	7.55	7.8	0
39	4	Brij 76 + propanol + tetradecane	100	447.69	29.56	0.9																							

Table 3. Continued

No.	Ref. no.	System	SR (%)	HS-Volume	HS- $\kappa_3$	HS-Dipole	HS-ABSQm	HS-Sp.Pol.	LS-Volume	LS- $\kappa_3$	LS-Dipole	LS-Sp.Pol.	S-LogP	CosR (%)	Co-Volume	Co- $\kappa_3$	Co-Dipole	Co-ABSQm	Co-Sp.Pol.	Co-LogP	OR (%)	O-Volume	O- $\kappa_3$	O-Dipole	O-ABSQm	O-Sp.Pol.	O-LogP	W/O-ME%	O/W-ME%
42	4	Brij 76 + butanol + dodecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	87.84	3.96	1.463911	0.40	0.099621	1.04	50	215.30	11.11	0.00001	0.00	0.10587	6.53	28.9	0
43	4	Brij 76 + heptanol + dodecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	138.78	7.16	1.473255	0.40	0.102726	2.56	50	215.30	11.11	0.00001	0.00	0.10587	6.53	7.1	0
44	4	Brij 76 + decanol + dodecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	189.86	9.96	1.463825	0.40	0.104081	4.09	50	215.30	11.11	0.00001	0.00	0.10587	6.53	4.7	0
45	4	Brij 76 + dodecanol + dodecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	223.83	11.96	1.463798	0.40	0.10468	5.11	50	215.30	11.11	0.00001	0.00	0.10587	6.53	3.5	0
46	4	Brij 76 + butanol + tetradecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	87.84	3.96	1.463911	0.40	0.099621	1.04	50	249.23	13.09	0.000013	0.00	0.10618	7.55	33.9	0
47	4	Brij 76 + butanol + hexadecane	100	447.69	29.56	0.988656	4.16	0.099356	317.09	17.07	0.00001	0.106606	6.5	50	87.84	3.96	1.463911	0.40	0.099621	1.04	50	283.16	15.08	0.000012	0.00	0.10642	8.57	25.3	0
48	5	Span 20/Tween 80 + propanol + ethyloleate	70	676.90	30.24	4.0621504	6.65	0.099504	255.47	13.44	0.084748	0.1072206	4.562	30	70.91	4.00	1.47441	0.40	0.097532	0.53	100	355.12	18.37	1.528065	0.62	0.10702	8.43	15.5	0
49	5	Span 20/Tween 80 + butanol + ethyloleate	70	676.90	30.24	4.0621504	6.65	0.099504	255.47	13.44	0.084748	0.1072206	4.562	30	87.84	3.96	1.463911	0.40	0.099621	1.04	100	355.12	18.37	1.528065	0.62	0.10702	8.43	43.81	0
50	5	Span 20/Tween 80 + hexanol + ethyloleate	70	676.90	30.24	4.0621504	6.65	0.099504	255.47	13.44	0.084748	0.1072206	4.562	30	121.78	5.96	1.463851	0.40	0.101998	2.05	100	355.12	18.37	1.528065	0.62	0.10702	8.43	24.56	0
51	5	Span 20/Tween 80 + octanol + ethyloleate	70	676.90	30.24	4.0621504	6.65	0.099504	255.47	13.44	0.084748	0.1072206	4.562	30	155.75	7.96	1.46385	0.40	0.103315	3.07	100	355.12	18.37	1.528065	0.62	0.10702	8.43	1.74	0
52	5	Span 20/Tween 80 + 1,2-propanediol + ethyloleate	70	676.90	30.24	4.0621504	6.65	0.099504	255.47	13.44	0.084748	0.1072206	4.562	30	79.15	3.92	2.125495	0.78	0.095426	-0.71	100	355.12	18.37	1.528065	0.62	0.10702	8.43	12.44	0
53	5	Span 20/Tween 80 + 1,2-pentandiol + ethyloleate	70	676.90	30.24	4.0621504	6.65	0.099504	255.47	13.44	0.084748	0.1072206	4.562	30	113.16	3.76	2.173882	0.78	0.099182	0.31	100	355.12	18.37	1.528065	0.62	0.10702	8.43	31.05	0
54	5	Span 20/Tween 80 + 1,2-hexandiol + ethyloleate	70	676.90	30.24	4.0621504	6.65	0.099504	255.47	13.44	0.084748	0.1072206	4.562	30	130.12	4.92	2.165131	0.78	0.100352	0.82	100	355.12	18.37	1.528065	0.62	0.10702	8.43	43.97	0
55	5	Span 20/Tween 80 + 1,2-octandiol + ethyloleate	70	676.90	30.24	4.0621504	6.65	0.099504	255.47	13.44	0.084748	0.1072206	4.562	30	164.12	6.92	2.16501	0.78	0.101928	1.84	100	355.12	18.37	1.528065	0.62	0.10702	8.43	49.11	0
56	6	Brij 96 + ethylenglycol + mineral oil	90	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	10	62.28	4.01	2.127774	0.79	0.09181	-1.25	100	367.81	20.00	0.010923	0.00	0.10687	11.12	9.72	6.84
57	6	Brij 96 + propylenglycol + mineral oil	90	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	10	79.15	3.92	2.125495	0.78	0.095426	-0.71	100	367.81	20.00	0.010923	0.00	0.10687	11.12	8.12	5.95
58	6	Brij 96 + glycerin + mineral oil	90	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	10	87.46	2.88	1.599463	1.17	0.09364	-1.98	100	367.81	20.00	0.010923	0.00	0.10687	11.12	0	6.3
59	6	Brij 96 + ethylenglycol + mineral oil	75	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	25	62.28	4.01	2.127774	0.79	0.09181	-1.25	100	367.81	20.00	0.010923	0.00	0.10687	11.12	7.79	6.18
60	6	Brij 96 + propylenglycol + mineral oil	75	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	25	79.15	3.92	2.125495	0.78	0.095426	-0.71	100	367.81	20.00	0.010923	0.00	0.10687	11.12	3.37	5.74
61	6	Brij 96 + glycerin + mineral oil	75	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	25	87.46	2.88	1.599463	1.17	0.09364	-1.98	100	367.81	20.00	0.010923	0.00	0.10687	11.12	0	6.17
62	6	Brij 96 + ethylenglycol + mineral oil	50	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	62.28	4.01	2.127774	0.79	0.09181	-1.25	100	367.81	20.00	0.010923	0.00	0.10687	11.12	0	7.66
63	6	Brij 96 + propylenglycol + mineral oil	50	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	79.15	3.92	2.125495	0.78	0.095426	-0.71	100	367.81	20.00	0.010923	0.00	0.10687	11.12	0	1.59
64	6	Brij 96 + glycerin + mineral oil	50	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	50	87.46	2.88	1.599463	1.17	0.09364	-1.98	100	367.81	20.00	0.010923	0.00	0.10687	11.12	3.79	5.59
65	6	Brij 96 + ethylenglycol + mineral oil	25	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	75	62.28	4.01	2.127774	0.79	0.09181	-1.25	100	367.81	20.00	0.010923	0.00	0.10687	11.12	0	9.9
66	6	Brij 96 + glycerin + mineral oil	25	447.30	29.56	0.989551	4.16	0.099442	310.56	16.81	0.124181	0.108229	6.05	75	87.46	2.88	1.599463	1.17	0.09364	-1.98	100	367.81	20.00	0.010923	0.00	0.10687	11.12	0	12.27
67	7	Tween 60 + pentanol + hexadecane	66	1018.69	48.76	3.626469	9.65	0.10038	300.12	16.00	0.010767	0.106521	6.31	34	104.79	5.30	1.473282	0.40	0.101017	1.55	100	283.16	15.08	0.000012	0.00	0.10642	8.57	0	33.21
68	8	Tween 40 + pentanol + mineral oil	78	1018.80	48.76	3.546863	9.65	0.100369	266.00	14.00	0.01103	0.106387	5.25	21.98	104.79	5.30	1.473282	0.40	0.101017	1.55	100	367.81	20.00	0.010923	0.00	0.10687	11.12	0	9.03



Table 4

Correlation matrix (cross-correlation  $r^2$ ) of the different selected molecular descriptors for W/O and O/W microemulsion systems

	SR	HS- Volume	HS- SP.Pol.	HS- Dipole	HS- $\kappa_{\alpha 3}$	HS- ABSQ	LS- Volume	LS- Sp.Pol.	LS- Dipole	LS- $\kappa_{\alpha 3}$	S-logP	CoSR	Co- Volume	Co- Sp.Pol.	Co- Dipole	Co- $\kappa_{\alpha 3}$	Co- ABSQon	Co-LogP	OR	O- Volume	O- Sp.Pol.	O- Dipole	O- $\kappa_{\alpha 3}$	O-ABSQ on	O-LogP
SR	1	0.64	0.64	0.48	0.60	0.63	0.20	0.00	0.01	0.20	0.10	0.32	0.14	0.26	0.70	0.09	0.75	0.68	0.66	0.36	0.08	0.51	0.26	0.54	0.05
HS-Volume		1	0.95	0.77	0.93	0.99	0.35	0.10	0.03	0.38	0.14	0.19	0.30	0.05	0.56	0.05	0.73	0.52	0.44	0.12	0.26	0.67	0.06	0.71	0.02
HS-Sp.Pol.			1	0.57	0.99	0.93	0.19	0.06	0.02	0.21	0.05	0.26	0.29	0.09	0.57	0.06	0.80	0.59	0.38	0.09	0.35	0.49	0.04	0.53	0.01
HS-Dipole				1	0.52	0.80	0.73	0.08	0.01	0.75	0.49	0.03	0.20	0.01	0.40	0.02	0.41	0.27	0.52	0.24	0.03	0.90	0.14	0.89	0.00
HS- $\kappa_{\alpha 3}$					1	0.91	0.15	0.08	0.04	0.17	0.02	0.28	0.30	0.08	0.54	0.06	0.79	0.58	0.32	0.06	0.41	0.45	0.02	0.49	0.02
HS-ABSQ on						1	0.38	0.10	0.03	0.41	0.16	0.17	0.30	0.05	0.55	0.05	0.72	0.51	0.45	0.13	0.24	0.70	0.06	0.73	0.02
LS-Volume							1	0.04	0.00	0.99	0.90	0.02	0.07	0.00	0.18	0.00	0.12	0.07	0.40	0.29	0.02	0.65	0.21	0.62	0.01
LS-Sp.Pol.								1	0.95	0.06	0.00	0.02	0.04	0.04	0.00	0.00	0.01	0.00	0.00	0.03	0.08	0.05	0.05	0.06	0.12
LS-Dipole									1	0.00	0.07	0.03	0.02	0.04	0.01	0.00	0.00	0.00	0.05	0.10	0.11	0.00	0.12	0.00	0.14
LS- $\kappa_{\alpha 3}$										1	0.87	0.01	0.08	0.00	0.19	0.00	0.13	0.08	0.40	0.28	0.02	0.67	0.20	0.64	0.00
S-logP											1	0.08	0.02	0.00	0.09	0.00	0.03	0.02	0.33	0.32	0.12	0.46	0.26	0.42	0.03
CoSR												1	0.17	0.02	0.13	0.02	0.32	0.20	0.00	0.06	0.53	0.07	0.10	0.09	0.19
Co-Volume													1	0.18	0.14	0.32	0.27	0.01	0.03	0.02	0.33	0.24	0.04	0.26	0.20
Co-Sp.Pol.														1	0.29	0.67	0.24	0.60	0.28	0.30	0.00	0.03	0.30	0.03	0.26
Co-Dipole															1	0.09	0.77	0.69	0.58	0.30	0.11	0.50	0.22	0.53	0.03
Co- $\kappa_{\alpha 3}$																1	0.11	0.44	0.08	0.09	0.00	0.03	0.08	0.03	0.06
Co-ABSQon																	1	0.82	0.43	0.14	0.35	0.51	0.07	0.56	0.00
Co-LogP																		1	0.47	0.25	0.16	0.35	0.18	0.38	0.04
OR																			1	0.81	0.03	0.46	0.72	0.46	0.32
O-Volume																				1	0.27	0.21	0.98	0.20	0.67
O-Sp.Pol.																					1	0.06	0.36	0.09	0.53
O-Dipole																						1	0.11	0.99	0.02
O- $\kappa_{\alpha 3}$																							1	0.10	0.79
O-ABSQ on																								1	0.02
O-LogP																									1

cant and/or superior *F*-value warranted the selection of the corresponding descriptor.

Descriptor variables were selected to represent three properties per microemulsion component, namely: molecular size, electrostatic and dispersion properties.

Eventually, only one descriptor remained to represent a particular property. For example, either *O-Volume* or  $O-\kappa_{\alpha 3}$  survived to express the size-related (shape-related) properties of oil molecules (more details in Section 3.4.1).

### 2.7.2. Multiple linear regression modeling

Stepwise backward regression analysis was performed to achieve the most statistically significant models. It was decided to utilize this routine after extensive exploration of other modeling strategies. It was concluded that this method allows the emergence of descriptors that are rendered significant when combined with others. However, such variables might be deemed redundant if considered singly.

Initially, the selected descriptors were utilized to generate two tentative models for O/W and W/O systems. Each model was then optimized by transforming the dependent variable (ME%) and expanding one or more of the independent descriptors as necessary. Subsequently, statistically problematic outliers were removed (Ramsey and Schafer, 1997). Finally, redundant explanatory variables were removed based on their respective *P*-values. Statistical modeling was carried out using SAS software for Windows, Release 6.12.

### 2.7.3. Statistical validation

The most significant QSAR models were cross-validated by splitting each of their respective training sets (microemulsion systems) into two subsets, one to fit a sub-model while the other to test the generated sub-model (Ramsey and Schafer, 1997; Maran et al., 1999). To this end, the collected microemulsion systems were classified into statistical strata where every unique reference is considered as a separate stratum (see Table 3). References 7 and 8 were arbitrarily added to the reference 6 cluster, since each provided one phase diagram. Subsequently, test data-points were selected from each stratum using random tables. Each stratum furnished nearly 20% of its data points to the test subset (Ramsey and Schafer, 1997; SciVision, 1999). The remaining data-points were kept as the training subsets.

This procedure was repeated three times. Accordingly, three unique test subsets with their corresponding training sets were selected to cross-validate each model (O/W or W/O). The resulting fitted sub-models were utilized to predict ME% of the corresponding test sets. Finally, the predicted ME% values were correlated with their experimental counterparts for each test subset.

## 3. Results and discussion

### 3.1. Data-mining

Phase diagrams corresponding to pseudoternary microemulsions produced from nonionic surfactants were collected from published literature. W/O and O/W microemulsion areas were cut and weighed to determine ME%. It is assumed that phase-diagrams illustrating W/O regions completely lack O/W areas and vice versa, unless otherwise indicated in the considered article (e.g., reference 4 in Table 1). The validity of this assumption is based on the anticipation that the respective authors probably scanned the whole phase diagrams of their systems searching for different microemulsion types. However, bicontinuous microemulsions (Thevenin et al., 1996) were excluded, as it is hard to define exactly their O/W and W/O regions. Furthermore, articles defining large complex molecules as cosurfactants, e.g., Myverol 18-99 (Constantinides and Scalart, 1997) were also excluded. Such molecules are large and amphiphilic, thus they were considered to be closer to surfactants in their properties. Table 1 summarizes the mined microemulsion systems.

### 3.2. Molecular modeling and descriptor calculations

In principle, the structural formula of an organic compound encodes all the information that determines the chemical, biological and physical properties of that compound (Grover et al., 2000). However, quantitative structure–activity (property) relationships (QSARs or QSPRs) define mathematically the relationship between a given type of activity (biological, chemical and physical) within a set of (usually) congeneric ‘single’ compounds and one or more physicochemical or structural parameters (Dear-den, 1994). However, in the current study, different physicochemical and structural parameters were generated for each ‘mixture’ of component compounds representing an overall microemulsion system. Subsequently, a statistical correlation was defined between the mixtures’ physicochemical descriptors and the corresponding ME%.

Accordingly, a variety of different physicochemical descriptors were calculated for each microemulsion component. To this end, the 2-D chemical structures of the different components (oils, surfactants and cosurfactants) were converted to reasonable 3-D structures using the rule-based methods of Alchemy 2000. However, such methods fail (generally) in accessing global energy minima, particularly for flexible molecules (Goodman, 1998). Nevertheless, the fact that the 3-D models were generated for all components utilizing the same approach is expected to minimize the overall energy-related error in the regression analysis. Furthermore, the resulting 3-D structures were energy-optimized using Alchemy 2000 force field to further reduce energy-related problems. Fig.

3 illustrates the generated 3-D structures of a representative oil (ethyl oleate), cosurfactant (sorbitol) and surfactant (Tween 80).

Subsequently, the resulting 3-D structures were utilized in calculating 3-D dependent descriptors via SciQSAR: molecular volume (Bodor et al., 1989) and dipole moment (Gasteiger and Marsili, 1980). However, other 3-D independent variables were also calculated, i.e. charge-related, polarization-related, connectivity indices and partition coefficients ( $\text{Log}P$ ) as described in Section 2.4.

Surfactant structures were cleaved into hydrophilic segments (HS) and lipophilic segments (LS) before energy minimization. The cleavage aimed at: (i) To generate separate descriptors covering the hydrophilic and lipophilic

parts of the surfactant molecules, which should enable better understanding of their roles in microemulsion formation. (ii) Some surfactant molecules, such as Tween 40, 60 and 80, were too large to have their descriptors calculated by SciQSAR (i.e. as single entity). However, surfactant  $\text{Log}P$  values were calculated for whole molecules.

Despite the fact that many earlier researchers utilized HLB values to explain microemulsion stability (Issa et al., 1977), it was decided to exclude this property from the current modeling study. HLB is an empirically defined quantity that takes only a limited number of surfactants' physicochemical properties into consideration (Becher, 1966). Furthermore, we noticed from our experience in microemulsion formulation the complete inadequacy of this property in predicting microemulsion stability.

### 3.3. Mass-ratio based descriptors

Mass ratio-based descriptors are introduced herein to express the fixed cosurfactant-to-oil or surfactant-to-cosurfactant ( $K_m$ ) ratios. Such descriptors are necessary in pseudoternary microemulsion systems since two of the microemulsion components must be incorporated in the system as a fixed-ratio mixture and subsequently expressed in the phase diagram at one of the triangle heads (Aboofazeli and Lawrence, 1993; Garti et al., 1995). Surfactant-to-cosurfactant and cosurfactant-to-oil ratios were reported to exert significant effects on ME% (Atwood et al., 1992; Aboofazeli et al., 1994; Garti et al., 1995).

Accordingly, the descriptor 'surfactant ratio' (SR) is defined herein as the fixed percentage of surfactant in a mixture with the associated cosurfactant. While the ratio of a cosurfactant in a mixture (with a surfactant or oil) is expressed through the descriptor 'cosurfactant ratio' (CoSR). Similarly, the mass ratio of an oil in a mixture (with a cosurfactant) is covered through an 'oil ratio' (OR) descriptor. If one of the components was represented solely at a triangular corner of the phase diagram it was assigned a ratio-descriptor value of 100%.

### 3.4. Statistical modeling

#### 3.4.1. Descriptor-selection prior to multiple linear regression (MLR)

Despite the diversity of the calculated molecular descriptors, they can be bundled into size- (or shape-), dispersion forces-, charge- and partition-related properties. Accordingly, it is not unexpected to detect extensive collinearity within each category of descriptors, which warranted manual selection prior to multiple linear regression modeling. Collinear explanatory variables tend to reduce the predictive power of statistical models (Martin, 1978). The selection process was carried out over two subsequent stages.

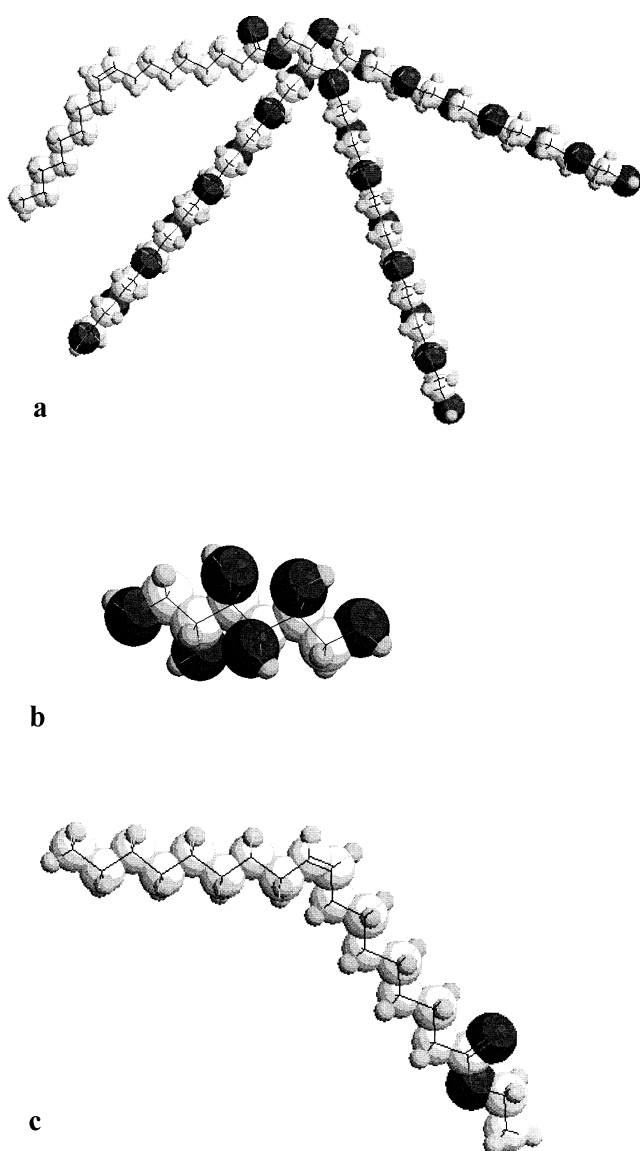


Fig. 3. CPK (Corey-Pauling-Koltun) views of the energy minimized 3-D structures of: (a) Tween 80, (b) sorbitol and (c) ethyl oleate. (White: carbon atoms; dark gray: oxygen atoms; light gray: hydrogen atoms).

The first stage was performed via the graphical interface of SciQSAR. Apparently collinear variables were defined and qualitatively filtered. The selection of shape and size cosurfactants' descriptors is illustrated here as an example. In this case, the descriptors Co-Volume, Co- $\kappa_{\alpha_3}$ , Co- $\chi^V$ , Co- $\chi^V$ , Co- $\chi$ , Co- $\chi$  and Co-WI were all strongly correlated. Fig. 4 shows the line trends of some of these descriptors. It is clearly evident from the figure that Co- $\kappa_{\alpha_3}$  is somehow less collinear with the other size variables. Accordingly, it was decided to drop Co- $\chi^V$ , Co- $\chi^V$ , Co- $\chi$ , Co- $\chi$ , and Co-WI in favor of Co-volume, as it infers superior information compared to connectivity indices. Correspondingly, Co-Volume and Co- $\kappa_{\alpha_3}$  were selected as the cosurfactants' size- and shape-related descriptors. Similar procedures were adapted to select size- (or shape-), electrostatic- and van der Waals-related descriptors of other microemulsion components. Eventually, at the end of this level, two descriptors expressed each property category: 3-D-dependent and 3-D-independent. Table 3 shows the mined microemulsion systems and the corresponding selected descriptors at this stage.

A second rigorous descriptor-filtration stage was carried out based on Table 4. Accordingly, only orthogonal descriptors were selected to express molecular size, shape, electrostatic and dispersion properties characterizing each microemulsion component. The selection was carried out such that if any two descriptors were collinear ( $r^2 > 0.65$ ), the one correlating with the least number of other descriptors (Table 4) within the same component class was selected. However, if two collinear descriptors were found to correlate with the same number of variables, then the selection was based on their corresponding  $F$ -values. The following examples illustrate the selection process at this stage.

- LS-*Sp.Pol.* was found to be strongly collinear with LS-*Dipole* ( $r^2 = 0.95$ ), while both were orthogonal with the remaining descriptors. Accordingly, the selection was based on their respective  $F$ -values. The statistically superior  $F$ -value of LS-*Sp.Pol.* ( $F = 6.75$ ) over that of LS-*Dipole* ( $F = 5.90$ ) justified the selection of LS-*Sp.Pol.* to express dispersion and electrostatic properties of the surfactants' lipophilic segments in O/W microemulsions. Furthermore, in LS-*Sp.Pol.* ( $F = 0.64$ ) was also favored over LS-*Dipole* ( $F = 0.18$ ) for W/O microemulsion models. It must be mentioned that such variables of poor individual  $F$ -values were enrolled in MLR in case that they might become statistically significant upon combination with other descriptors.
- Regarding the surfactant partition descriptor (S-Log $P$ ), it was found to be strongly collinear with LS- $\kappa_{\alpha_3}$  and LS-Volume. However, LS- $\kappa_{\alpha_3}$  and LS-Volume were preselected to represent size- and shape-related surfactant properties in the O/W and W/O models, respectively. Therefore, a decision was taken to drop S-Log $P$

and keep the more informative size- and shape-related variables.

Based on the previous arguments, the following orthogonal descriptors were selected for the subsequent MLR of W/O systems: SR, CoSR, OR, HS- $\kappa_{\alpha_3}$ , HS-*Dipole*, LS-Volume, LS-*Sp.Pol.*, Co-Volume, Co-*Sp.Pol.*, Co-ABSQon, O- $\kappa_{\alpha_3}$ , O-*Sp.Pol.* and O-*Dipole*.

While the following descriptors were selected for modeling O/W systems: SR, CoSR, OR, HS-Volume, LS- $\kappa_{\alpha_3}$ , LS-*Sp.Pol.*, Co-Volume, Co-*Sp.Pol.*, Co-ABSQon, O- $\kappa_{\alpha_3}$ , O-*Sp.Pol.* and O-*Dipole*.

### 3.4.2. Statistical modeling of O/W microemulsion data

The final set of selected descriptors (12 variables) was utilized in the QSPR modeling of O/W microemulsions. Stepwise backward regression analysis was carried out to reach the statistically most significant model. It was concluded, from different modeling trials, that this method allows the emergence of descriptors that are rendered significant upon combination with others. However, such variables might be deemed redundant if considered singly. Accordingly, an initial tentative model that included all the selected descriptors was generated ( $n = 68$ ,  $R^2 = 0.74$ ,  $F = 14.80$ ).

Subsequently, different dependent variable (O/W-ME%) transformations were explored to improve  $R^2$  (Ramsey and Schafer, 1997). Empirically, it was found that root transformations provided the optimum improvements in  $R^2$ , as well as they yielded homogenous distribution of the residuals plot. However, the fourth root of O/W-ME% was found to give the highest  $R^2$  (0.80),  $F$ -value (20.23) and optimal mean square of error (M.S.E. = 0.11) and thus was chosen for further optimization.

Subsequently, through inspecting the scatter plots of the dependent variable (O/W-ME%)<sup>1/4</sup> versus each descriptor, it was concluded that LS- $\kappa_{\alpha_3}$  shows parabolic relationship with (O/W-ME%)<sup>1/4</sup>, prompting the incorporation of (LS- $\kappa_{\alpha_3}$ )<sup>2</sup> in the model. Still, the statistical criteria of the resulting model were only moderately improved ( $R^2 = 0.83$ ,  $F = 22.34$  and M.S.E. = 0.10). However, the fitness of the model was significantly enhanced by removing microemulsion systems (20, 21 and 29, Table 3), which were acting as a group of satellite outliers ( $R^2 = 0.88$ ,  $F = 32.82$ , M.S.E. = 0.07). The outliers were also removed in the cross-validation procedure (Section 3.5).

Finally, redundant descriptors were discarded based on their  $P$ -values. It was noticed that the removal of the redundant descriptor of the highest  $P$ -value influenced the significance of other descriptors, such that some insignificant variables became significant. Accordingly, redundant variables were removed in a one-per-step manner rather than all at once. The highest  $P$ -value descriptor was

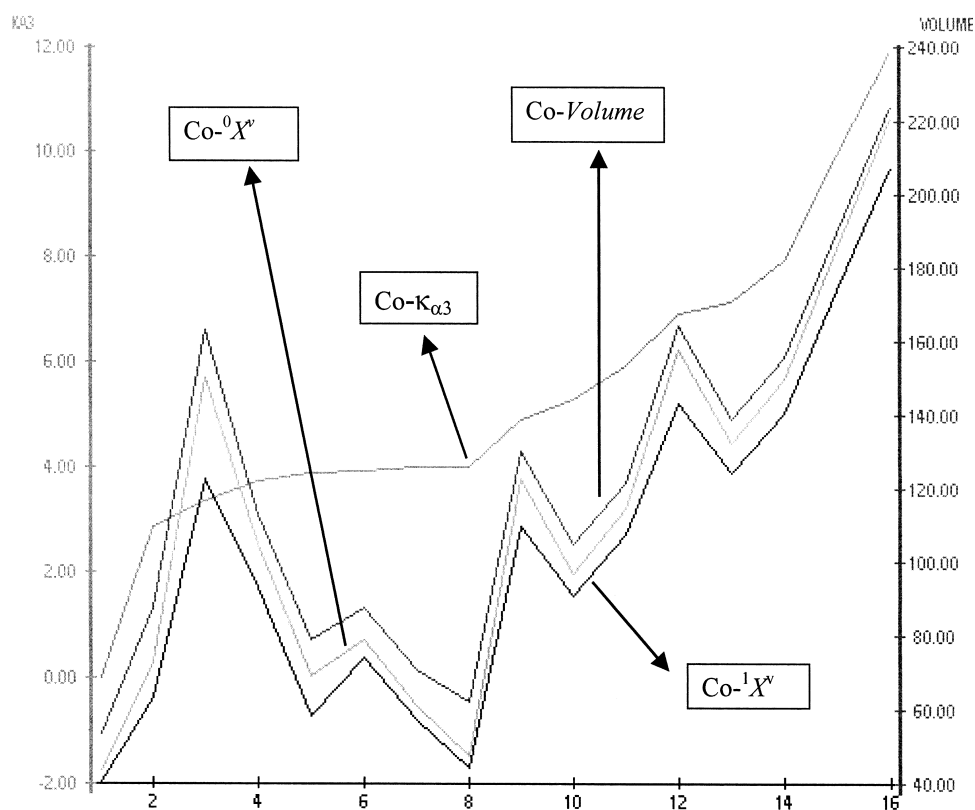


Fig. 4. SciQSAR graphical presentation of the different cosurfactants considered in this study (x-axis) versus their corresponding  $\text{Co-}\kappa_{\alpha_3}$ ,  $\text{Co-Volume}$ , the connectivity descriptors  $\text{Co}^0\chi^v$  and  $\text{Co}^1\chi^v$ .

removed first. Afterwards, the descriptor with the highest  $P$ -value in the subsequent equation was removed, and so on. The elimination process was terminated when the  $P$ -values of all descriptors were below 0.05. The following descriptors were removed in a stepwise manner:  $\text{O-}\kappa_{\alpha_3}$  ( $P$ -value = 0.43),  $\text{Co-ABSQon}$  ( $P$ -value = 0.62),  $\text{Co-Sp.Pol.}$  ( $P$ -value = 0.42). Eq. (1) represents the final most significant model. The 95% confidence limits (CL) of different regression coefficients are shown in brackets ( $[\pm\text{CL}]$ ).

$$\begin{aligned}
 (\text{O/W-ME}\%)^{1/4} = & -167.00[\pm 114.04] \\
 & -2.32[\pm 0.70] \text{ SR} \\
 & -0.0063[\pm 0.0032] \text{ HS-Volume} \\
 & -322.89[\pm 308.45] \text{ LS-Sp.Pol.} \\
 & -2.51[\pm 0.94] \text{ CoSR} \\
 & -0.0035[\pm 0.002] \text{ Co-Volume} \\
 & -1.08[\pm 0.27] \text{ O-Dipole} \\
 & -963.87[\pm 308.48] \text{ O-Sp.Pol.} \\
 & +42.30[\pm 18.45] \text{ LS-}\kappa_{\alpha_3} \\
 & -1.41[\pm 0.61] (\text{LS-}\kappa_{\alpha_3})^2 \\
 n = & 65, R^2 = 0.88, F = 44.80, \\
 \text{M.S.E.} = & 0.07
 \end{aligned}
 \tag{1}$$

The validity of this model was established through cross-validation (see Section 3.5). Fig. 5 illustrates the scatter plot of fitted values produced by Eq. (1) versus the corresponding experimental  $(\text{O/W-ME}\%)^{1/4}$  values.

### 3.4.3. Statistical modeling of W/O microemulsion data

A similar strategy was followed to reach the most significant QSPR model of W/O microemulsions. An initial tentative model that included all the selected descriptors (13 variables) was generated ( $n=68$ ,  $R^2=0.71$ ,  $F=11.30$ ,  $\text{M.S.E.}=69.23$ ).

Different transformations were attempted on the response variable (W/O-ME%) to improve  $R^2$ . Again, it was noticed that root transformations were yielding the best correlation coefficients ( $R^2$ ). Furthermore, they allowed homogenous distribution of the residuals plot. However, the sixth root of W/O-ME% was found to be the best transformation for further optimization ( $R^2=0.88$ ,  $F=34.80$ ,  $\text{M.S.E.}=0.09$ ).

Removing microemulsions 60 and 61 (Table 3), which were detected as outliers in the residuals plot, significantly enhanced the model ( $R^2=0.92$ ,  $F=50.40$ ,  $\text{M.S.E.}=0.06$ ). Subsequently, redundant descriptors were removed in a step-by-step manner, as guided by their  $P$ -values. The following descriptors were removed:  $\text{O-Sp.Pol.}$  ( $P$ -value = 0.96),  $\text{Co-ABSQon}$  ( $P$ -value = 0.87),  $\text{CoSR}$  ( $P$ -value =

0.76), *HS-Dipole* ( $P$ -value=0.59), *LS-Volume* ( $P$ -value=0.40). Eq. (2) is the most significant final QSPR model. The 95% confidence limits of different regression coefficients are shown in brackets.

$$\begin{aligned} (\text{W/O-ME}\%)^{1/6} = & -13.10[\pm 11.45] + 1.54[\pm 0.59] \text{SR} \\ & - 0.042[\pm 0.016] \text{HS-}\kappa_{\alpha 3} \\ & + 105.93[\pm 94.48] \text{LS-}Sp.Pol. \\ & - 0.0037[\pm 0.0030] \text{Co-Volume} \\ & + 39.82 [36.90] \text{Co-}Sp.Pol. \\ & - 0.045 [\pm 0.031] \text{O-}\kappa_{\alpha 3} \\ & + 0.59[\pm 0.15] \text{O-Dipole} \\ n = 66, R^2 = 0.92, F = 92.51, \\ \text{M.S.E.} = 0.058 \end{aligned} \quad (2)$$

The validity of model Eq. (2) was established through cross-validation (see Section 3.5). Fig. 6 illustrates the scatter plot of fitted values produced by Eq. (2) versus the corresponding experimental  $(\text{W/O-ME}\%)^{1/6}$  values.

### 3.5. Cross-validation of models (1) and (2)

Statistical cross-validation is required to rule out the possibility of generating successful MLR models by chance, and to check the predicting power of a particular statistical model (Ramsey and Schafer, 1997). To cross-validate each of the developed models, certain test subsets were selected randomly from Table 3, however, after removing the outliers (details in Section 2.7.3). Subsequently, the remaining data points (i.e. fit subset) were utilized to generate sub-models based on either Eqs. (1) or (2). The resulting sub-models were used to predict microemulsion areas of the corresponding test sub-sets. Finally, predicted O/W-ME% and W/O-ME% were correlated with the respective experimental values of the test sets. This procedure was repeated three times such that three unique test subsets covered 60% of the total data points for each microemulsion type. Table 5 shows the three test subsets utilized in the cross-validation of model

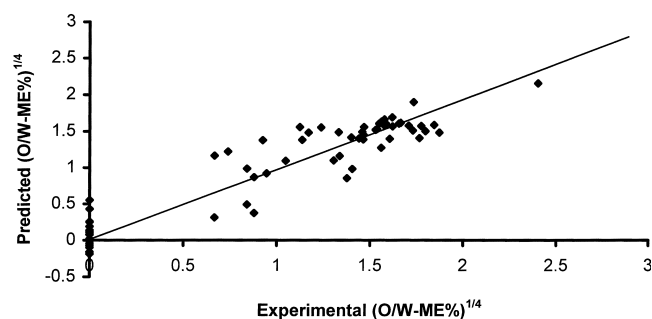


Fig. 5. Scatter plot of fitted versus experimental values of  $(\text{O/W-ME}\%)^{1/4}$  produced by Eq. (1).

(1) and their corresponding training (fit) subsets. However, Table 6 summarizes the respective results of the three rounds of cross-validation. The average values of fit and test correlation coefficients were 0.88 and 0.86, respectively, which clearly illustrate the consistency of the total O/W microemulsion data points, as well as the statistical significance of Eq. (1).

Table 7 illustrates the data subsets utilized in the cross-validation of model Eq. (2). Table 8 shows the corresponding cross-validation results. The average values of fit and test correlation coefficients were 0.92 and 0.91, respectively, which illustrates the harmony of the W/O microemulsion data and the statistical significance of model Eq. (2).

### 3.6. Interpretation of the statistical models

#### 3.6.1. The W/O model

From model Eq. (2), it seems that SR,  $\text{HS-}\kappa_{\alpha 3}$ , *LS-Sp.Pol.*, *Co-Volume*, *Co-Sp.Pol.*,  $\text{O-}\kappa_{\alpha 3}$  and *O-Dipole*, combined together, were significantly influencing the formation and stability of W/O microemulsions. The negative sign of the  $\text{HS-}\kappa_{\alpha 3}$  term suggests an inverse relationship between the shape (or size) of the surfactant's hydrophilic segment and the microemulsion existing area. Accordingly, one can conclude that the steric bulkiness of hydrophilic head groups interferes in the efficient packing of surfactant molecules at the interface and thus reduce the possibility of microemulsion formation. This is consistent with previous findings, which related the inability of Brij 96 in forming an isotropic oily phase to the large hydrophilic head of the surfactant (Issa et al., 1977). On the other hand, the statistical redundancy of *HS-Dipole* and its omission from model Eq. (2) suggests the negligible influence of the hydrophilic heads' electrostatic properties on W/O microemulsion formation, which is probably related to their hydration. Efficient hydration is expected to shield the electrostatic forces (attraction or repulsion) generated by the hydrophilic groups. Furthermore, hydration is expected further to amplify the destabilizing steric effects of the hydrophilic heads.

The emergence of *LS-Sp.Pol.*, which is a van der Waals

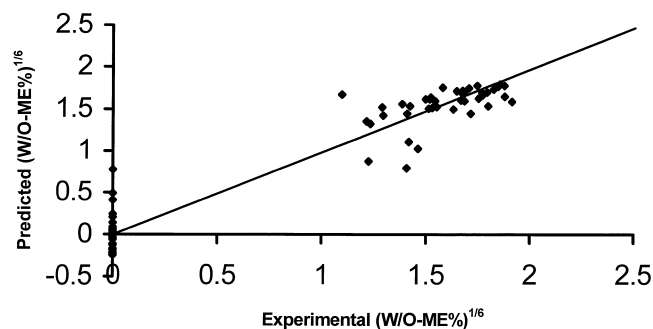


Fig. 6. Scatter plot of fitted versus experimental values of  $(\text{W/O-ME}\%)^{1/6}$  produced by Eq. (2).

Table 5  
Fit and test subsets utilized in the cross-validation of the O/W QSPR model Eq. (1)

Set to fit (training)		Set to test	
Subset no.	Data-points (numbers as in Table 3)	Subset no.	Data-points (numbers as in Table 3)
1	1,2,3,5,6,7,9,10,11,12,13,14, 15,16,18,19,22,23,25,26,27, 28,30,31,32,33,35,36,38,40, 41,42,43,44,45,46,48,49,51, 53,54,55,56,58,59,61,62,63, 64,65,66 and 68	2	4,8,17,24,34,37,39,47,50,52, 57,60 and 67
3	2,3,4,5,6,7,8,9,10,12,13,14, 15,16,17,19,22,24,25,26,28, 31,32,33,34,35,36,37,38,39, 40,41,42,43,47,48,50,52,53, 54,55,57,58,59,60,61,62,63, 64,66,67 and 68	4	1,11,18,23,27,30,44,45,46,49, 51,56 and 65
5	1,2,3,4,5,7,8,9,11,12,14,15, 16,17,18,23,24,25,26,27,28, 30,32,33,34,36,37,38,39,40, 43,44,45,46,47,49,50,51,52, 53,54,56,57,58,59,60,61,64, 65,66,67 and 68	6	6,10,13,19,22,31,35,41,42,48, 55,62 and 63

related descriptor (Martin, 1993), in Eq. (2) suggests the important role played by the surfactants' lipophilic chains on the binding phenomenon at the interface, albeit through van der Waals attraction. So far as we are aware this factor was not studied before in relation to the formation of microemulsions.

The role of the cosurfactant size is clearly evident through the emergence of *Co-Volume* in Eq. (2). The associated negative coefficient suggests that cosurfactants with large volumes might be incapable of forming microemulsions. Large-sized cosurfactants are expected to increase distances between different components at the interface and thus destabilize the different attractive forces responsible for the integrity of the interfacial film. Shorter-chain alcohols were reported to produce extended microemulsion zones (Thevenin et al., 1996).

The emergence of *Co-Sp.Pol.* in Eq. (2) suggests that the polarizabilities of cosurfactant molecules exert an influential direct effect on the formation of W/O microemulsions. Accordingly, it seems logical to conclude that cosurfactants influence W/O microemulsion forma-

tion, at least partially, through stabilizing the interfacial film via van der Waals attraction. The fact that the charge-related descriptor *Co-ABSQon* was found to be statistically redundant implies the negligible effect of the cosurfactant's electrostatic properties on the formation and stability of the interfacial film. The emergence of *Co-Sp.Pol.*, combined with the omission of *Co-ABSQon* suggest that interfacial cosurfactant molecules seem to position themselves towards the aqueous phase where the hydrophobic van der Waals interactions supersede the hydration-shielded electrostatic forces.

The presence of *O-Dipole* in Eq. (2) suggests the involvement of oil electrostatic properties in the formation and stabilization of W/O microemulsions. The positive coefficient associated with *O-Dipole* indicates a direct relationship with W/O microemulsion stability. Accordingly, oil molecules seem to contribute to the stability of the interfacial film through electrostatic attraction at the interface. This is consistent with results reported earlier, which concluded that as the oil dielectric constant (polarity) increased, the total amount of water solubilized by the oily

Table 6  
Cross-validation results of the O/W microemulsion model Eq. (1)

Set to fit	<i>n</i>	$R^2$ (fit)	<i>F</i> (fit)	M.S.E. (fit)	Set to predict	<i>n</i>	$R^2$ (test)	<i>F</i> (test)	M.S.E. (test)
1	52	0.87	32.59	0.071	2	13	0.87	75.32	0.071
3	52	0.87	31.81	0.076	4	13	0.90	103.87	0.060
5	52	0.89	37.76	0.068	6	13	0.81	47.24	0.089

Table 7  
Fit and test subsets utilized in the cross-validation of W/O QSPR model Eq. (2)

Set to fit (training)		Set to test	
Subset no.	Data-points (numbers as in Table 3)	Subset no.	Data-points (numbers as in Table 3)
7	1,2,3,4,6,8,9,10,11,12,13,15, 16,17,18,19,20,21,22,24,26, 28,29,31,32,33,34,35,37,38, 39,41,42,43,44,45,46,47,48, 50,52,53,54,55,56,57,58,59, 62,63,64,65 and 67	8	5,7,14,23,25,27,30,36,40,49, 51,66 and 68
9	1,2,3,4,5,7,8,9,10,12,13,14, 16,17,18,19,21,23,24,25,26, 27,28,29,30,31,34,35,36,37, 38,39,40,41,43,44,45,47,48, 49,50,51,52,55,56,58,62,63, 64,65,66,67 and 68	10	6,11,15,20,22,32,33,42,46,53, 54,57 and 59
11	1,2,4,5,6,7,8,9,10,11,13,14, 15,17,18,19,20,21,23,24,25, 26,27,28,30,31,32,33,35,36, 37,38,39,40,42,44,45,46,48, 49,51,53,54,55,56,57,58,59, 63,64,65,66,67 and 68.	12	3,12,16,22,29,34,41,43,47,50, 52 and 62

phase increased (Issa et al., 1977). Furthermore, polar oils such as alkanolic acids and their corresponding ethyl esters were reported to produce large oily ( $L_2$ ) isotropic regions, while lower polarity oils like triglycerides generated fairly narrow  $L_2$  regions (Aboofazeli et al., 1995).

The negative contribution of  $O-\kappa_{\alpha_3}$  in Eq. (2) suggests that large-sized oils destabilize W/O microemulsions probably via increasing the distances between attracting components at the interfacial film. Oil molecular size was reported to influence the formation of stable isotropic regions along the surfactant/oil axis (Aboofazeli et al., 1995).

Finally, the surfactant-to-cosurfactant fixed ratio (SR, often referred to as  $K_m$ ) is apparently statistically significant, as evident in Eq. (2). High surfactant concentrations are necessary to enhance the extent of water solubilization in oils (W/O). Consistent with this result,  $K_m$  was previously reported to play an important role in the extent of microemulsion existing area (Thevenin et al., 1996).

### 3.6.2. The O/W model

Model Eq. (1) suggests the statistical significance of the

combination of SR, HS-Volume,  $LS-\kappa_{\alpha_3}$ ,  $LS-Sp.Pol.$ , CoSR, Co-Volume, O-Dipole and O-Sp.Pol. in the formation and stability of O/W microemulsions. The negative coefficient of HS-Volume implies that large-sized surfactant head groups interfere in microemulsion formation. Large hydrophilic heads are expected to result in steric repulsion and inefficient packing at the interface with the consequent reduction in the chance of interfacial film formation. On the other hand, the role played by the sizes (shapes) of the surfactants' lipophilic segments is more complex, as indicated by the emergence of  $LS-\kappa_{\alpha_3}$  and  $(LS-\kappa_{\alpha_3})^2$  in Eq. (1). A positive coefficient associated with  $LS-\kappa_{\alpha_3}$  combined with a negative one with  $(LS-\kappa_{\alpha_3})^2$  indicate parabolic relationship with microemulsion stability, such that within small to moderate ranges the sizes of the surfactants' lipophilic segments are directly proportional to the stability of O/W microemulsions. However, within larger LS size ranges, the  $(LS-\kappa_{\alpha_3})^2$  term becomes significant thus leading to an overall negative effect on W/O microemulsion formation. This behavior can be explained based on contradicting factors of efficient packing and steric repulsion at the interfacial film. It seems that within

Table 8  
Cross-validation results of the W/O microemulsion model Eq. (2)

Set to fit	$n$	$R^2$ (fit)	$F$ (fit)	M.S.E. (fit)	Set to predict	$n$	$R^2$ (test)	$F$ (test)	M.S.E. (test)
7	53	0.91	67.66	0.063	8	13	0.93	153.56	0.047
9	53	0.92	70.61	0.063	10	13	0.93	149.03	0.037
11	54	0.93	84.55	0.052	12	12	0.86	59.55	0.099



small to medium ranges, the sizes of the LS chains contribute towards O/W microemulsion stability via forming an uninterrupted interfacial film. However, within larger size ranges the destabilizing steric effects of the LS chains become more significant due to the associated increase in distances separating different attracting entities at the interface. Furthermore, the emergence of a parabolic relationship suggests the existence of an optimal size for the LS groups allowing optimal O/W microemulsion stability.

Eq. (1) shows the significance of *LS-Sp.Pol.* in the formation of O/W microemulsions, which suggests a significant role played by van der Waals forces exhibited by the LS parts. However, in contrast to model Eq. (2) (i.e. W/O), *LS-Sp.Pol.* is associated with a negative coefficient in model Eq. (1) (i.e. O/W) suggesting an inverse relationship with microemulsion stability. A possible explanation could be that lipophilic tails expressing strong van der Waals attractive forces tend to be withdrawn deeper towards the oily internal phase, which seems to interfere in the continuity and packing of the interfacial film.

The role of the cosurfactant size is evident through the emergence of *Co-Volume* in Eq. (1). However, the associated negative coefficient suggests that smaller-sized cosurfactants improve the chances of O/W microemulsion formation, probably through allowing closer proximities between attracting moieties at the interface.

However, oils seem to influence O/W microemulsion formation through their dipole moments (*O-Dipole*) and specific polarizabilities (*O-Sp.Pol.*). The large coefficient of *O-Sp.Pol.* mirrors the significance of the oil dispersion forces in O/W microemulsion formation, despite its negligible value in the W/O case. The negative effect of *O-Sp.Pol.* on O/W microemulsion stability suggests that oils possessing stronger van der Waals attraction forces drag the lipophilic tails of the surfactant molecules deeper towards the oily phase with the concomitant reduction in interfacial film packing and stability.

Finally, the emergence of surfactant and cosurfactant fixed ratios (SR and CoSR, respectively) in Eq. (1) suggests their significant value upon O/W microemulsion formation. The negative coefficients of both variables suggest that if surfactants or cosurfactants were used solely or in extremely high ratios, they would tend to destabilize O/W microemulsions. Accordingly, it is preferable to use surfactant/cosurfactant mixtures rather than pure materials.

However, it is important to remember that surfactants and cosurfactants are not always incorporated together in fixed-ratio mixtures. Sometimes, a particular cosurfactant is mixed with the selected oil in a starting fixed-ratio mixture, while the surfactant is added as a pure material to search for the microemulsion area. Microemulsion systems 19–47 in Table 1 represent examples of such cases. Accordingly, CoSR and SR are actually independent of each other, i.e. CoSR might indicate cosurfactant to

surfactant ratio or cosurfactant to oil ratio, while SR always indicates the surfactant to cosurfactant ratio only. The orthogonality of CoSR and SR is further established by their cross-correlation coefficient  $r^2=0.32$  (Table 4), which allows for their combined emergence in model Eq. (1).

In conclusion, this study has shown that data mining, molecular modeling, descriptor calculation followed by multiple linear regression analysis were successful in producing statistically significant and predictive QSPR models for O/W and W/O microemulsions.

The emergence of complex ME% root functions in Eqs. (1) and (2) indicates complex, nonlinear relationships connecting the physicochemical properties of the different microemulsion components and the corresponding microemulsion areas.

However, the generated models shed some light upon the factors influencing microemulsion formation and stability.

## Acknowledgements

The authors would like to thank the Deanship of Scientific Research at the Jordan University for providing funds towards acquiring Alchemy 2000 and SciQSAR 3.0. We also would like to thank Dr Bassam Amro for his helpful discussions.

## References

- Aboofazeli, R., Lawrence, M.J., 1993. Investigation into the formation and characterization of phospholipid microemulsions. I. Pseudo-ternary phase diagrams of systems containing water–lecithin–alcohol–isopropyl myristate. *Int. J. Pharm.* 93, 161–175.
- Aboofazeli, R., Lawrence, B.C., Wicks, R.S., Lawrence, M.J., 1994. Investigation into the formation and characterization of phospholipid microemulsions. II. Pseudo-ternary phase diagrams of systems containing water–lecithin–alcohol–isopropyl myristate and either an alkanolic acid, amine, alkanediol, polyethylene glycol alkyl ether or alcohol as cosurfactant. *Int. J. Pharm.* 111, 63–72.
- Aboofazeli, R., Patel, N., Thomas, M., Lawrence, M.J., 1995. Investigations into the formation and characterization of phospholipid microemulsions. IV. Pseudo-ternary phase diagrams of systems containing water–lecithin–alcohol and oil: the influence of oil. *Int. J. Pharm.* 125, 107–116.
- Alany, R.G., Rades, T., Agatonovic-Kustrin, S., Davies, M.N., Tucker, G.I., 2000. Effect of alcohols and diols on the phase behavior of quaternary systems. *Int. J. Pharm.* 196, 141–145.
- Attwood, D., Florence, A.T., 1983. *Surfactant Systems, Their Chemistry, Pharmacy and Biology*. Chapman and Hall, New York.
- Attwood, D., Mallon, C., Ktistis, G., Taylor, J.C., 1992. A study on factors influencing the droplet size in non-ionic oil in water microemulsions. *Int. J. Pharm.* 88, 417–422.
- Attwood, D., Ktistis, G., 1989. A light scattering study on oil in water microemulsions. *Int. J. Pharm.* 52, 165–171.
- Aulton, M.E., 1990. *Pharmaceutics. The Science of Dosage Form Design*. ELBS, London.

- Beacher, P., 1966. Emulsions: Theory and Practice. Reinhold Publishing Corporation, New York.
- Bodor, N., Huang, M.J., 1992. An extended version of a novel method for the estimation of partition-coefficients. *J. Pharm. Sci.* 81, 272–281.
- Bodor, N., Gabanyi, Z., Wong, C.K., 1989. A new method for estimation of partition-coefficient. *J. Am. Chem. Soc.* 111, 3783–3786.
- CompuDrug Ltd, 1995. Pallas Reference Manual. ChempuDrug Chemistry Ltd, USA.
- Constantinides, P.P., Scalart, P.J., 1997. Formulation and physical characterization of water in oil microemulsions containing long-versus medium-chain glycerides. *Int. J. Pharm.* 158, 57–68.
- Dearden, J.C., 1994. Application of quantitative structure property relationships to pharmaceuticals. *Chemometr. Intell. Lab. Syst.* 24, 77–87.
- Delgado, N.J., Remers, A.W., 1998. Wilson and Gisvold Textbook of Organic Medicinal and Pharmaceutical Chemistry. Lippincott Company, Philadelphia.
- Friberg, S.E., 1990. Micelles, microemulsions, liquid-crystals, and the structure of stratum corneum lipids. *J. Soc. Cosmet. Chem.* 41, 155–171.
- Friman, S., Bäckman, L., 1996. A new microemulsion formulation of cyclosporin. Pharmacokinetics and clinical features. *Clin. Pharmacokinet.* 30, 181–193.
- Garti, N., Aserin, A., Ezrahi, S., Wachtel, E., 1995. Water solubilization and chain length compatibility in non-ionic microemulsions. *J. Colloid Interf. Sci.* 169, 428–436.
- Gasteiger, J., Marsili, M., 1980. Iterative partial equalization of orbital electronegativity—A rapid access to atomic charges. *Tetrahedron* 36, 3219–3228.
- Goodman, M.J., 1998. Chemical Application of Molecular Modeling. Royal Society of Chemistry, Cambridge, UK.
- Grover, M., Singh, B., Bakshi, M., Singh, S., 2000. Quantitative structure–property relationships in pharmaceutical research—part 1. *Pharm. Sci. Technol. Today* 3, 8–34.
- Hermansky, C., Mackay, A.R., 1980. Light scattering measurements in nonionic microemulsions. *J. Colloid Interf. Sci.* 73, 324–332.
- Ho, H.-O., Chih-Chuan, H., Ming-Thau, S., 1996. Preparation of microemulsions using polyglycerol protein drugs. *J. Pharm. Sci.* 85, 138–143.
- Israelachvili, L., Mitchell, J.D., Niham, W.B., 1976. Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers. *J. Chem. Soc., Faraday Trans. 2* 72, 1525–1567.
- Issa, L., Florence, A.T., Treguier, J.P., Seiller, M., Puisieux, F., 1977. The influence of surfactant HLB and the nature of the oil phase on the phase diagrams of nonionic structure–oil–water systems. *J. Colloid Interf. Sci.* 59, 319–327.
- Kale, J.N., Allen, V.L., 1989. Studies on microemulsions using Brij 96 as surfactant and glycerin, ethylene glycol and propylene glycol as cosurfactants. *Int. J. Pharm.* 57, 87–93.
- Katritzky, A.R., Gordeeva, E.V., 1993. Traditional topological indices versus electronic, geometric, and combined molecular descriptors in QSAR and QSPR Research. *J. Chem. Inf. Comput. Sci.* 33, 835–857.
- Kier, L.B., 1985. A shape index from molecular graphs. *Quant. Struct.-Act. Relatsh.* 4, 109–116.
- Kier, L.B., Hall, L.H., 1986. Molecular Connectivity in Structure–Activity Analysis. Research Studies Press, Letchworth, UK.
- Kreuter, J., 1994. Colloidal Drug Delivery Systems. Marcel Dekker, New York.
- Ktistis, G., 1990. A viscosity study on oil in water microemulsions. *Int. J. Pharm.* 61, 213–218.
- Mackay, A.R., Jacobson, K., Tourian, J., 1980. Measurement of pH and  $pK_a$  in O/W microemulsions. *J. Colloid Interf. Sci.* 76, 515–523.
- Maran, U., Karelson, M., Katritzky, A.R., 1999. A comprehensive QSAR treatment of genotoxicity of heteroaromatics and aromatic amines. *Quant. Struct.-Act. Relatsh.* 18, 3–10.
- Martin, A., 1993. Physical Chemistry. Lea and Febrieger, Philadelphia.
- Martin, Y.C., 1978. Quantitative Drug Design. Marcel Dekker, New York.
- Miller, K.J., 1990. Additivity methods in molecular polarizability. *J. Am. Chem. Soc.* 112, 8533–8542.
- Ramsey, L.F., Schafer, W.D., 1997. The Statistical Sleuth. Duxbury Press, California.
- SciVision, 1999. SciQSAR™ 3.0 User Guide. Academic Press, Massachusetts.
- Tenjarla, S., 1999. Microemulsions: an overview and pharmaceutical applications. *Crit. Rev. Ther. Drug. Carrier Syst.* 16, 461–521.
- Thevenin, A.M., Grossiord, L.J., Poelman, C.M., 1996. Sucrose esters cosurfactant microemulsion systems for transdermal delivery: assessment of biocontinuous structures. *Int. J. Pharm.* 137, 177–186.
- Tripos Inc, 1998. Alchemy 2000 Reference Manual. Tripos Inc, Missouri.
- Wade, A., Weller, J.P., 1994. Handbook of Pharmaceutical Excipients. The Pharmaceutical Press, London.
- Warisnoicharoen, W., Lansley, A.B., Lawrence, M.J., 2000. Nonionic oil in water microemulsions: the effect of oil type on phase behavior. *Int. J. Pharm.* 198, 7–27.
- Watanasirichaikul, S., Davies, M.N., Rades, T., Tucker, G.I., 2000. Preparation of biodegradable insulin nanocapsules from biocompatible microemulsions. *Pharm. Res.* 17, 684–689.
- Weast, C.R., 1989. CRC—Handbook of Chemistry and Physics. CRC Press, Florida.
- Wiener, H., 1947. Correlation of heats of isomerization and differences in heats of vaporization of isomers among the paraffin hydrocarbons. *J. Am. Chem. Soc.* 69, 2636–2644.