Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Pharmaceutics 295 (2005) 135–155
international journal of pharmaceutics
www.elsevier.com/locate/ijpharm

QSPR modeling of pseudoternary microemulsions formulated employing lecithin surfactants: Application of data mining, molecular and statistical modeling

Mutasem O. Taha [a,*], Heba Abdel-Halim [b], Mutasim Al-Ghazawi [c], Enam Khalil [b]

[a] *Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman 11942, Jordan*
[b] *Department of Pharmaceutics and Pharmaceutical Technology, Faculty of Pharmacy, University of Jordan, Amman, Jordan*
[c] *Department of Biopharmaceutics and Clinical Pharmacy, Faculty of Pharmacy, University of Jordan, Amman, Jordan*

Received 31 August 2004; received in revised form 26 January 2005; accepted 9 February 2005
Available online 1 April 2005

Abstract

Data mining, computer aided molecular modeling, descriptor calculation, genetic algorithm and multiple linear regression analysis techniques were combined together to generate predictive quantitative structure property relationship (QSPR) models explaining the formation of lecithin-based W/O microemulsions. Ninety-four microemulsion phase diagrams were collected from five different references published over the past few years. Computer-based molecular modeling techniques were then applied on the components of the collected microemulsion systems to generate corresponding plausible three-dimensional (3D) structures. The resulting 3D models were utilized to calculate a group of molecular physicochemical descriptors. Thereafter, genetic algorithm and backward stepwise regression analysis were separately assessed as means for selecting optimal descriptor sets for statistical modeling. The selected descriptors were correlated with microemulsion existence areas employing multiple linear regression analysis. The resulting W/O models were statistically validated and found to be of significant predictive power. The models allowed better understanding of the process of microemulsion formation. Unfortunately, all QSPR modeling efforts directed towards O/W microemulsions failed completely.

Keywords: Microemulsions; Lecithin; Physicochemical descriptors; QSPR; Molecular modeling; Genetic algorithm; Multiple linear regression

1. Introduction

Microemulsions are homogenous, transparent, optically isotropic, thermodynamically stable dispersions of water and oil stabilized by relatively large amounts of surfactant(s) frequently in combination with cosurfactant(s) (Friberg, 1990; Aboofazeli et al., 1995; Tenjarla, 1999). Microemulsions can be classified into three categories: water-in-oil (W/O), oil-in-water (O/W) or bicontinuous (Tenjarla, 1999). Three-component microemulsions (i.e., stabilized by surfactant(s) only) are generally known as tertiary microemulsions, while

* Corresponding author. Tel.: +962 6 5355000x2505; fax: +962 6 5355533.
E-mail address: mutasem@ju.edu.jo (M.O. Taha).

0378-5173/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2005.02.012

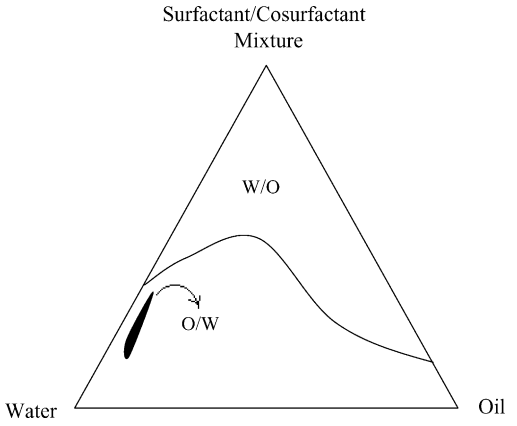

Fig. 1. Pseudoternary phase diagram showing W/O and O/W microemulsion areas of a system comprised of surfactant/cosurfactant mixture, oil and water (Aboofazeli et al., 1994a).

those based on four components (i.e., with cosurfactant) are known as pseudoternary or quaternary (Tenjarla, 1999). Microemulsions are graphically represented as stability areas in triangular phase diagrams (Kreuter, 1994), where each triangular corner designates certain component. Fig. 1 illustrates an example phase diagram of a microemulsion system comprised of surfactant/cosurfactant mixture, oil and water, and illustrating W/O and O/W areas (Aboofazeli et al., 1994a).

Microemulsions have wide industrial applications. For example, microemulsion-based polymerization processes represent an effective route towards novel interesting polymeric materials (Schmuhl et al., 1998; Xu et al., 1999). Moreover, microemulsion dispersions are promising candidates as means for controlled drug delivery (Tenjarla, 1999), and as drug carriers for oral, topical and parenteral administration (Constantinides, 1995; Thevenin et al., 1996; Tenjarla, 1999; Rosano et al., 1979). Furthermore, microemulsions have been shown to possess promising potential in the fields of cosmetics and various consumer products (Ho et al., 1996; Friman and Bäckman, 1996; Tenjarla, 1999; Watnasirichaikul et al., 2000).

Microemulsion formulations based on nonionic surfactants (e.g., Tweens and Spans) proved to be appropriate for topical pharmaceutical applications. However, their applicability for oral and parenteral routes is limited by the toxicity profiles of nonionic surfactants (Tenjarla, 1999). On the other hand, the superior safety profiles of phospholipids (e.g., lecithins) rendered their corresponding microemulsions appropriate for parenteral and oral routes. Lecithins are virtually nontoxic in acute and short-term oral studies (Tenjarla, 1999). Furthermore, lecithins are generally nonirritating and nonsensitizing for animal and human skin (Fiume, 2001).

Despite the increased interest in microemulsions and the abundance of relevant experimental and theoretical data, their formulation is still highly empirical and time-consuming (Attwood et al., 1992; Aboofazeli et al., 1994a,b). Few theories tried to explain microemulsion formation (Kreuter, 1994). However, the most famous is the geometrical packing theory, which depicts microemulsions as tiny droplets of internal phase (ca. 200 nm) dispersed in the continuous phase and stabilized by efficient packing of surfactant molecules at the interface (Israelachvili et al., 1976; Friberg, 1990; Tenjarla, 1999). However, this theory is of limited practical value in the preparation of microemulsions. Furthermore, the geometrical packing theory lacks explicit elucidation of the rules of oils and cosurfactants in the stabilization of the interfacial film. It is noteworthy to mention that oil and cosurfactant molecules were recently implicated in the interfacial packing of some microemulsion systems (Aboofazeli et al., 1995; Aboofazeli and Lawrence, 1993; Tenjarla, 1999; Kreuter, 1994).

Nevertheless, the geometrical packing theory implies that microemulsion stability is a function of the interfacial affinity interactions (e.g., electrostatic and van der Waals forces) that promote the integrity and continuity of the interfacial film. On the other hand, the steric bulkiness of interfacial molecules is expected to disturb the continuity of the interfacial film by allowing interfacial inter-molecular gaps. Accordingly, it can be assumed that the stability of a particular microemulsion system is related to the molecular physicochemical characteristics of its interfacial components.

The ready accessibility to various calculated molecular physicochemical descriptors utilizing computer-based molecular models (Gasteiger and Marsili, 1980; Kier and Hall, 1986; Bodor et al., 1989; Miller, 1990; Bodor and Huang, 1992; Hall et al., 2001) prompted us to investigate the possibility of statistically correlating molecular descriptors calculated for microemulsion components with corresponding

microemulsion stabilities, i.e., developing statistical quantitative structure–property relationships (QSPR). Such statistical models should cut down the trial time required for preparing microemulsions, furthermore, they should provide better understanding of microemulsion formation and stability.

The current study commenced by collecting phase diagrams of pseudoternary microemulsions from published literature. However, we restricted our data-mining to lecithin-based microemulsions due to the recent pharmaceutical interest in lecithin surfactants. Subsequently, computer-aided molecular modeling techniques were utilized to generate 3D representations of respective microemulsion components. The resulting in silico molecular models were subsequently utilized to calculate a variety of corresponding molecular descriptors. Afterwards, genetic algorithm (GA) and backward stepwise regression analysis techniques were separately evaluated as means of selecting optimal descriptors combinations for statistical correlation. Thereafter, multiple linear regression (MLR) analysis was utilized to correlate the selected descriptors with microemulsion stability areas. The resulting statistical relationships were thoroughly validated and utilized to probe the mechanism of microemulsion formation.

The validity of QSPR modeling in probing the formation of microemulsions stabilized by non-ionic surfactants has been thoroughly investigated (Taha et al., 2002).

2. Methods

2.1. Software

1. CS ChemDraw Ultra 6.0, Cambridge Soft Corp. (http://www.cambridgesoft.com/), USA.
2. Alchemy 2000, 2.05, Tripos Inc. (http://www.tripos.com), USA.
3. SciQSAR 3.0, Scivision (http://www.scivision.com/SciQSAR.html), USA.
4. QsarIS, Scivision Inc. (www.scivision.com), USA.
5. 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 past few years. Phase diagrams corresponding to lecithin-based pseudoternary microemulsions were collected. Both O/W and W/O microemulsion categories were included. Only published systems of clearly defined and illustrated phase diagrams were considered for modeling. The percentages of microemulsion existence areas were determined by cut and weight method (Taha et al., 2002). Table 1 illustrates the selected microemulsion phase diagrams; their components, mass ratios and the corresponding references. We collected 94 microemulsion systems, which should allow the investigation of the effects of a maximum of 19 descriptors on microemulsion stability without jeopardizing the statistical significance of the final QSPR models (i.e., *F*-statistic). The optimal ratio of explanatory descriptors-to-observations is 1:5 (Ramsey and Schafer, 1997). Furthermore, we concentrated on microemulsions reported by a particular research group (Aboofazeli et al., 1995, 1994a,b; Aboofazeli and Lawrence, 1993) to minimize any operator-related variability in microemulsion stability areas. Variation in the sources of data might lead to significant error or "noise" in the corresponding QSPR models due to variations in microemulsion stabilities unexplainable by the physicochemical properties of the microemulsion components.

Microemulsion components were classified into surfactants, cosurfactants and oils according to the following definitions and assumptions:

- Surfactants (lecithins) are complex mixtures of phospholipids characterized with molecular weight range of 500–700 Da, and two structurally distinct parts of opposite lipophilicity/hydrophilicity properties (see Table 2 for detailed composition) (Aboofazeli et al., 1994a). Fig. 2 shows the general chemical structures of some phospholipid components of lecithins (Mathew and Holde, 1991). Carboxylic acids ($pK_{a1} \approx 3.0$–4.5), amines ($pK_{a1} \approx 8.0$–9.0) and phosphates ($pK_{a1} \approx 2.13$) moieties were assumed to exist entirely in their ionized forms since microemulsions are generally formulated using distilled water (pH ≈ 6.0).
- Cosurfactants are defined as small (60–190 Da) mono or multi-hydroxy alcohols or carboxylic acids that might contain ether linkages. Cosurfactants are added to stabilize microemulsions (Kreuter, 1994).
- Oils are defined as moderate to large alkyl hydrocarbons (ca. 140–900 Da) that might contain ester, ether or carboxylic acid moieties.

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

No.	Microemulsion constituents			Microemulsion area			
	Surfactant	Cosurfactant	Oil	Km	%O/W area	%W/O area	Reference
1	O 200[a]	*n*-Butanol	Oleic acid	1:1		28.2	(Aboofazeli et al., 1995)
2	O 200	*n*-Butanol	Octanoic acid	1:1.94	0.17	29.7	(Aboofazeli et al., 1995)
3	O 200	*n*-Butanol	Oleic acid	1:1.94	0.42	31	(Aboofazeli et al., 1995)
4	O 200	*n*-Butanol	Ethyl octanoate	1:1.94	0.42	37.78	(Aboofazeli et al., 1995)
5	O 200	*n*-Butanol	Ethyl oleate	1:1.94	0.68	42.23	(Aboofazeli et al., 1995)
6	O 200	*n*-Propranol	Soya bean	1:1	–	41.46	(Aboofazeli et al., 1995)
7	O 200	*n*-Propranol	Miglyol 812	1:1	–	21.2	(Aboofazeli et al., 1995)
8	O 200	*n*-Propranol	Octanoic acid	1:1	0.29	38.56	(Aboofazeli et al., 1995)
9	O 200	*n*-Propranol	Oleic acid	1:1	0.24	40.34	(Aboofazeli et al., 1995)
10	O 200	*n*-Propranol	Ethyl octanoate	1:1	0.075	44.5	(Aboofazeli et al., 1995)
11	O 200	*n*-Propranol	Ethyl oleate	1:1	0.1	47.43	(Aboofazeli et al., 1995)
12	O 200	*n*-Propranol	Soya been	1:1.94	–	16.72	(Aboofazeli et al., 1995)
13	O 200	*n*-Propranol	Miglyol 812	1:1.94	–	20.5	(Aboofazeli et al., 1995)
14	O 200	*n*-Propranol	Oleic acid	1:1.94	0.44	27.75	(Aboofazeli et al., 1995)
15	O 200	*n*-Propranol	Octanoic acid	1:1.94	–	29.17	(Aboofazeli et al., 1995)
16	O 200	*n*-Propranol	Ethyl oleate	1:1.94	1.28	34.75	(Aboofazeli et al., 1995)
17	O 200	*n*-Propranol	Ethyl octanoate	1:1.94	0.31	34	(Aboofazeli et al., 1995)
18	O 200	*n*-Butanol	Ethyl oleate	1:1	–	35.02	(Aboofazeli et al., 1995)
19	O 200	*n*-Butanol	Soya been	1:1.94	–	17.55	(Aboofazeli et al., 1995)
20	O 200	*n*-Butanol	Miglyol 812	1:1.94	–	20.26	(Aboofazeli et al., 1995)
21	O 200	*n*-Butanol	IPM[b]	1:0.6	3.45	42.32	(Attwood et al., 1992)
22	O 200	*n*-Butanol	IPM	1:0.45	4.11	30.09	(Attwood et al., 1992)
23	O 200	*n*-Butanol	IPM	1:0.33	3.71	20.7	(Attwood et al., 1992)
24	E 200[c]	*n*-Butanol	IPM	1:0.6	1.02	54.28	(Attwood et al., 1992)
25	E 200	*n*-Butanol	IPM	1:0.45	3.45	50.36	(Attwood et al., 1992)
26	O 200	*sec*-Butanol	IPM	1:1	–	43.03	(Aboofazeli and Lawrence, 1993)
27	O 200	*sec*-Butanol	IPM	1.5:1	0.78	45.31	(Aboofazeli and Lawrence, 1993)
28	O 200	*sec*-Butanol	IPM	1.77:1	1.24	42.87	(Aboofazeli and Lawrence, 1993)
29	O 200	*sec*-Butanol	IPM	1.94:1	0.6	37.23	(Aboofazeli and Lawrence, 1993)
30	O 200	*n*-Butanol	IPM	1:1	–	37.2	(Aboofazeli and Lawrence, 1993)
31	O 200	*n*-Butanol	IPM	1.5:1	1.23	40.12	(Aboofazeli and Lawrence, 1993)
32	O 200	*n*-Butanol	IPM	1.77:1	1.03	42.6	(Aboofazeli and Lawrence, 1993)
33	O 200	*n*-Butanol	IPM	1.94:1	1.37	38.66	(Aboofazeli and Lawrence, 1993)
34	E 200	*n*-Pentanol	IPM	1:1	–	30.7	(Aboofazeli and Lawrence, 1993)
35	E 200	*n*-Pentanol	IPM	1.5:1	–	41.14	(Aboofazeli and Lawrence, 1993)
36	E 200	*n*-Pentanol	IPM	1.77:1	0.86	39.3	(Aboofazeli and Lawrence, 1993)
37	E 200	*n*-Pentanol	IPM	1.94:1	1.51	39.51	(Aboofazeli and Lawrence, 1993)
38	E 200	*tert*-Butanol	IPM	1:1	0.36	48.44	(Aboofazeli and Lawrence, 1993)
39	E 200	*tert*-Butanol	IPM	1.5:1	0.62	47.62	(Aboofazeli and Lawrence, 1993)
40	E 200	*tert*-Butanol	IPM	1.77:1	0.86	56.01	(Aboofazeli and Lawrence, 1993)
41	E 200	*tert*-Butanol	IPM	1.94:1	0.7	54.81	(Aboofazeli and Lawrence, 1993)
42	E 200	Isobutanol	IPM	1:1	–	35.09	(Aboofazeli and Lawrence, 1993)
43	E 200	Isobutanol	IPM	1.5:1	0.57	38.28	(Aboofazeli and Lawrence, 1993)
44	E 200	Isobutanol	IPM	1.77:1	0.77	36.88	(Aboofazeli and Lawrence, 1993)
45	E 200	Isobutanol	IPM	1.94:1	1.11	39.86	(Aboofazeli and Lawrence, 1993)
46	E 200	*sec*-Butanol	IPM	1:1	0.61	44	(Aboofazeli and Lawrence, 1993)
47	E 200	*sec*-Butanol	IPM	1.5:1	0.41	45.36	(Aboofazeli and Lawrence, 1993)
48	E 200	*sec*-Butanol	IPM	1.77:1	1.14	48.95	(Aboofazeli and Lawrence, 1993)
49	E 200	*sec*-butanol	IPM	1.94:1	1.21	50.39	(Aboofazeli and Lawrence, 1993)
50	E 200	*n*-Butanol	IPM	1:1	–	36.05	(Aboofazeli and Lawrence, 1993)
51	E 200	*n*-Butanol	IPM	1.5:1	0.99	42.81	(Aboofazeli and Lawrence, 1993)
52	E 200	*n*-Butanol	IPM	1.77:1	0.47	40.9	(Aboofazeli and Lawrence, 1993)

Table 1 (*Continued*)

No.	Microemulsion constituents			Microemulsion area			
	Surfactant	Cosurfactant	Oil	Km	%O/W area	%W/O area	Reference
53	E 200	*n*-Butanol	IPM	1.94:1	0.85	46.22	(Aboofazeli and Lawrence, 1993)
54	E 200	Isopropanol	IPM	1:1	0.53	47.5	(Aboofazeli and Lawrence, 1993)
55	E 200	Isopropanol	IPM	1.5:1	0.86	45.68	(Aboofazeli and Lawrence, 1993)
56	E 200	Isopropanol	IPM	1.77:1	0.544	44.96	(Aboofazeli and Lawrence, 1993)
57	E 200	Isopropanol	IPM	1.94:1	0.59	43.36	(Aboofazeli and Lawrence, 1993)
58	E 200	*n*-Propanol	IPM	1:1	0.73	46.39	(Aboofazeli and Lawrence, 1993)
59	E 200	*n*-Propanol	IPM	1.5:1	0.99	45.37	(Aboofazeli and Lawrence, 1993)
60	E 200	*n*-Propanol	IPM	1.77:1	1.18	39.17	(Aboofazeli and Lawrence, 1993)
61	E 200	*n*-Propanol	IPM	1.94:1	1.04	47.06	(Aboofazeli and Lawrence, 1993)
62	E 200	*n*-Pentanoic acid	IPM	1:1	–	17.85	(Aboofazeli et al., 1994b)
63	E 200	*n*-Hexanoic acid	IPM	1:1	–	15.89	(Aboofazeli et al., 1994b)
64	E 200	1,2-Butanediol	IPM	1:1	–	23.44	(Aboofazeli et al., 1994b)
65	E 200	1,2-Hexanediol	IPM	1:1	–	56.23	(Aboofazeli et al., 1994b)
66	E 200	1,2-Pentanediol	IPM	1:1	–	40.77	(Aboofazeli et al., 1994b)
67	E 200	Diethylene glycol monobutyl ether	IPM	1:1	0.95	27.92	(Aboofazeli et al., 1994b)
68	E 200	Diethylene glycol monopentyl ether	IPM	1:1	–	59.4	(Aboofazeli et al., 1994b)
69	E 200	Diethylene glycol monohexyl ether	IPM	1:1	–	42.95	(Aboofazeli et al., 1994b)
70	E 200	*n*-Hexanol	IPM	1:1	–	25.66	(Aboofazeli et al., 1994b)
71	E 170[d]	Isopropanol	IPM	1:1	0.46	51.2	(Aboofazeli et al., 1994a)
72	E 170	Isopropanol	IPM	1.5:1	0.49	56.37	(Aboofazeli et al., 1994a)
73	E 170	Isopropanol	IPM	1.77:1	0.19	47.61	(Aboofazeli et al., 1994a)
74	E 170	Isopropanol	IPM	1.94:1	0.25	45.55	(Aboofazeli et al., 1994a)
75	E 170	*n*-Butanol	IPM	1:1	–	35.72	(Aboofazeli et al., 1994a)
76	E 170	*n*-Butanol	IPM	1.5:1	–	42.24	(Aboofazeli et al., 1994a)
77	E 170	*n*-Butanol	IPM	1.77:1	–	42.93	(Aboofazeli et al., 1994a)
78	E 170	*n*-Butanol	IPM	1.94:1	1.11	42.95	(Aboofazeli et al., 1994a)
79	E 170	*sec*-Butanol	IPM	1:1	–	41.41	(Aboofazeli et al., 1994a)
80	E 170	*sec*-Butanol	IPM	1.5:1	–	46.58	(Aboofazeli et al., 1994a)
81	E 170	*sec*-Butanol	IPM	1.77:1	1.06	44.58	(Aboofazeli et al., 1994a)
82	E 170	*sec*-Butanol	IPM	1.94/1	0.6	46.85	(Aboofazeli et al., 1994a)
83	E 170	Isobutanol	IPM	1:1	–	33.94	(Aboofazeli et al., 1994a)
84	E 170	Isobutanol	IPM	1.5:1	–	46.97	(Aboofazeli et al., 1994a)
85	E 170	Isobutanol	IPM	1.77:1	1.15	43.51	(Aboofazeli et al., 1994a)
86	E 170	Isobutanol	IPM	1.94:1	1.45	45.22	(Aboofazeli et al., 1994a)
87	E 170	*tert*-Butanol	IPM	1:1	–	48.77	(Aboofazeli et al., 1994a)
88	E 170	*tert*-Butanol	IPM	1.5:1	0.44	50.99	(Aboofazeli et al., 1994a)
89	E 170	*tert*-Butanol	IPM	1.77:1	0.5	47.15	(Aboofazeli et al., 1994a)
90	E 170	*tert*-Butanol	IPM	1.94:1	0.28	48.15	(Aboofazeli et al., 1994a)
91	E 170	*n*-Pentanol	IPM	1:1	–	26.66	(Aboofazeli et al., 1994a)
92	E 170	*n*-Pentanol	IPM	1.5:1	–	40.13	(Aboofazeli et al., 1994a)
93	E 170	*n*-Pentanol	IPM	1.77:1	–	44.44	(Aboofazeli et al., 1994a)
94	E 170	*n*-Pentanol	IPM	1.94:1	–	45.38	(Aboofazeli et al., 1994a)

[a] Ovithin lecithin (egg lecithin).
[b] Isopropyl myristate.
[c] Epikuron lecithin (soya bean lecithin).
[d] Epikuron lecithin (soya bean lecithin).

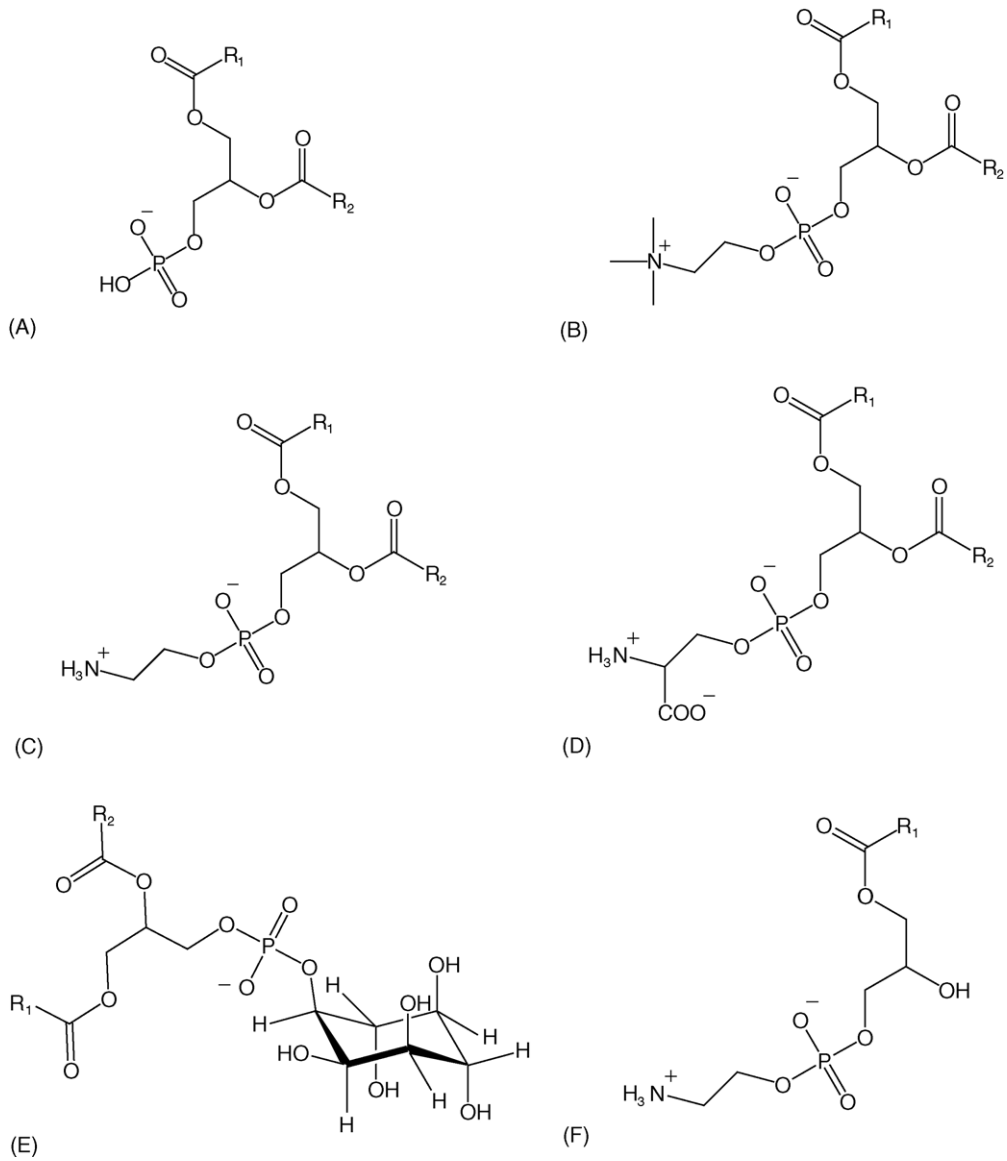

Fig. 2. Examples of some phospholipid components of different lecithin-type surfactants: (A) phosphatidic acid, (B) phosphatidylcholine, (C) phosphatidyl-ethanolamine, (D) phosphatidylserine, (E) phosphatidylinositol, and (F) lysophosphatidylcholine. R' and R'' are lipophilic tails originating from different fatty acids (Mathew and Holde, 1991).

Carboxylic acids (in oils or cosurfactants) were modeled in their ionized forms ($pK_a \approx 3.0$–4.5).

2.3. Molecular modeling

The chemical structures of all cosurfactants and most oils were generated from their corresponding chemical names. However, the structures of lecithin surfactants, and remaining oils were collected from a variety of resources. Tables 2 and 3 list the chemical compositions of the surfactant and oil fractions within the collected microemulsion systems.

The two-dimensional (2D) chemical structures of different components were sketched using Chem-

Table 2
Average compositions of soybean and egg lecithins (Aboofazeli et al., 1994a)

Lecithin	% w/w of polar head groups			% w/w of total fatty acid			
	Pc[a]	Lyso-Pc[b]	Other PL[c]	Palmitic and stearic	Oleic	Linoleic	Linolenic
Epikuron 200 (soybean)	94	3	1	18	10	64	7
Epikuron 170 (soybean)	70	4	23[d]	18	10	64	7
Ovithin 200 (egg)	92	3	2	47	32	17	–

[a] Phosphatidylcholine.
[b] Lysophosphatidylcholine.
[c] Phospholipid.
[d] 12% Phosphatidylethanolamine and 11% other phospholipids phosphatidyl serine 5.5% and phosphatidylinositol 5.5%.

Draw Ultra 6.0. Subsequently, they were imported into Alchemy 2000® and converted to corresponding three-dimensional (3D) representations using the 2D–3D rule based methods employed in Alchemy 2000® (Tripos, 1998).

Subsequently, the 3D structures were further optimized using Alchemy 2000® molecular mechanics force field and energy optimization. The minimization process was performed using the conjugate gradient algorithm employed in Alchemy 2000® (Tripos, 1998). The minimization cutoff values were RMS = 0.05 and $\Delta E = 0$ for cosurfactants, while they were 0.35 and 0, respectively, for surfactants and oil molecules. Fig. 3 shows the optimized 3D structures of representative surfactant, oil and cosurfactant.

2.4. Calculated descriptors

The 3D structures were utilized to calculate a number of physicochemical descriptors. Sixteen descriptors were calculated for each microemulsion component utilizing SciQSAR®:

- ABSQ: the sum of absolute values of charges on each atom of the molecule in electrons. It is calculated employing the empirical atomic charges model based on partial equalization of orbital electronegativity (Gasteiger and Marsili, 1980).
- ABSQon: the sum of absolute values of charges on nitrogen and oxygen atoms in the particular molecule.
- $MaxQ^+$: is the largest positive charge over any atom within a particular molecule.
- $MaxQ^-$: is the largest negative charge over any atom within a particular molecule.
- Dipole: is the dipole moment of the molecule calculated based on the 3D structure and charges generated by the Gasteiger–Marsili method.
- Polar: molecular polarizability is calculated based on the 3D-independent additive approach described byMiller (1990).
- Sp.pol: specific polarizability determined by dividing polarizability by volume.
- log *P*: the logarithm of octanol/water partition coefficient estimated by SciQSAR employing a neural

Table 3
Chemical names and compositions of different oils incorporated in the collected microemulsion systems

No.	Generic name	Chemical composition	Reference
1	Ethyl oleate	Ethyl 9: octadecenoate	Wade and Weller, 1994
2	Isopropyl myristate (IPM)	Methylethyl tetradecanoate	Wade and Weller, 1994
3	Miglyol 812	Medium chain triglycerids of fatty acids of which not less than 95% are the saturated fatty acids; octanoic (caprylic) acid and decanoic (capric) in 50:50 ratio	Butter, 1993
4	Oleic acid	9: Octadecenoic acid	Wade and Weller, 1994
5	Soybean oil	A mixture of triglycerides with the following percentages: ca. 53% linoleic triglycerides, ca. 8% linolenic triglycerides, ca. 22%oleic triglycerides, ca. 11% palmitic triglycerides and ca. 4% stearic triglycerides	Wade and Weller, 1994

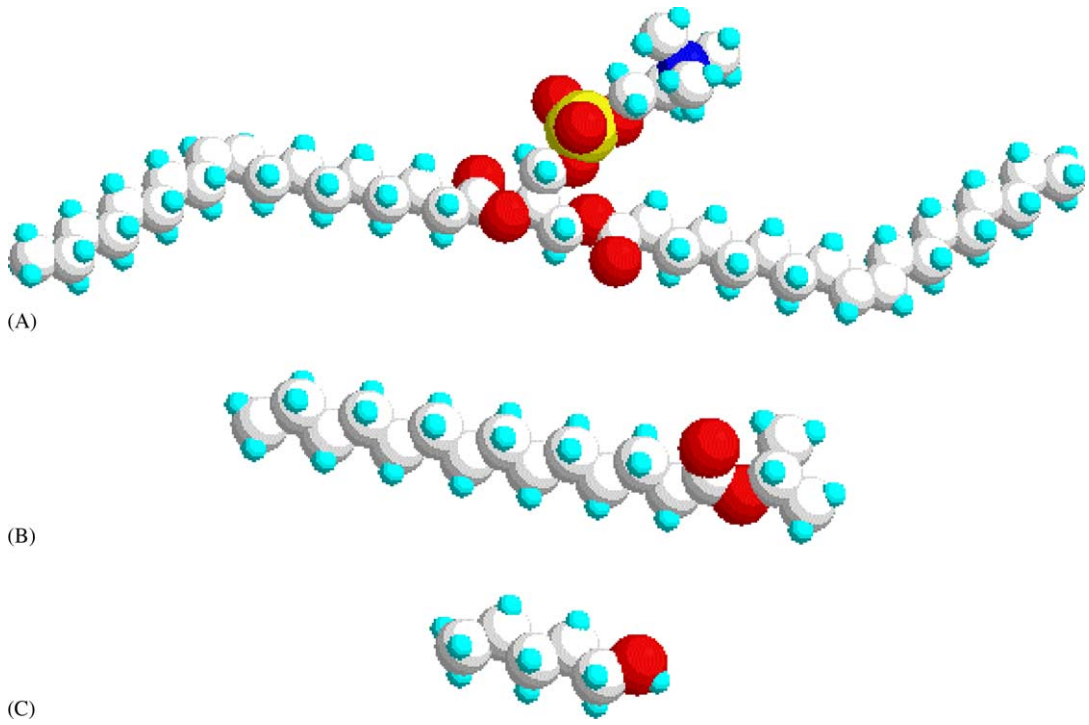

Fig. 3. Optimized 3D structures of representative: (A) surfactant (phosphatidyl-choline found in lecithins), (B) oil (isopropyl myristate) and (C) cosurfactant (*n*-pentanol). Red: oxygen atoms, white: carbon atoms, light blue: hydrogen atoms, dark blue: nitrogen atoms, yellow: phosphorus atoms.

network approach that utilizes a variety of 2D and 3D descriptors (Bodor and Huang, 1992).

- MW: molecular weight in Daltons.
- Volume: the molecular volume of a molecule. It is a 3D-dependent descriptor computed by the grid method of Bodor et al. (1989).
- WI (Wiener index): is a dimensionless, 3D-independent topological parameter based on the hydrogen-suppressed graph of the molecule. It encodes the number of bonds between all pairs of atoms. The larger and more branched a molecule is, the higher is the Wiener index (Wiener, 1947).
- $\kappa_{\alpha 3}$ (Kappa Alpha 3): a third order shape index for molecules. It is a 3D-independent descriptor that encodes information on the degree of cyclicity and the degree of centralization/separation in the branching of a molecule (Kier, 1985; SciVision, 1999).
- ${}^{1}\chi$, ${}^{3}\chi$, ${}^{0}\chi^{V}$ and ${}^{1}\chi^{V}$: are a group of dimensionless third-order molecular connectivity descriptors that encode the 2D structure of a molecule including molecular branching and length (Kier and Hall, 1986; Katritzky and Gordeeva, 1993).

2.4.1. Surfactant descriptors

The collected microemulsion systems are based on three lecithin surfactants, namely, Epikuron 200, Epikuron 170 and Ovithin 200. However, each lecithin surfactant is characterized with a particular phospholipid composition, in which a variety of hydrophilic phosphate moieties are attached to a variety of hydrophobic alkyl chains (fatty acid esters). Table 2 illustrates the components of each lecithin surfactant and their corresponding approximate percentages (Aboofazeli et al., 2000), while Fig. 2 shows the chemical structures of different phosphate hydrophilic heads found in lecithins.

For simplification purposes, it was decided to calculate separately the molecular descriptors that correspond to the hydrophilic and hydrophobic fragments of each particular lecithin. To this end, the hydrophilic head groups, i.e., phosphatidylcholine, phosphatidyl

ethanolamine, phosphatidylserine, phosphatidylinositol, and lysophosphatidylcholine (Fig. 2, Table 2), were modeled into 3D structures (see Section 2.3). Subsequently, corresponding descriptors were calculated for each phosphate head group. The hydrophilic descriptors of the overall lecithin (e.g., Epikuron 200) were calculated as the average of the respective descriptors of the individual hydrophilic heads according to their respective percentages in the surfactant mixture (as in Table 2). Similarly, the lipophilic descriptors of a particular lecithin were calculated as the average values of the descriptors of the lipophilic fragments within the particular lecithin. The following equation illustrates the calculation:

$$\text{lecithin descriptor} = \sum R_i \times \text{Di}$$

where R_i is the ratio (w/w) of the particular fragment (phosphate head moiety or alkyl chain, Table 2) and Di the molecular descriptor of that particular fragment.

Despite that the exact palmitic-to-stearic acids ratios are not reported in Table 2, soybean lecithin was reported to contain ca. 12% and 4% (i.e., a ratio of 3:1) palmitic and stearic acids, respectively (Windhole, 1983). Accordingly, this ratio was utilized to calculate the lipophilic descriptors of soybean lecithins. Furthermore, it was also assumed that the same ratio (i.e., palmitic: stearic) exists in egg lecithin.

Table 4 shows the calculated descriptors of individual hydrophilic phospholipid head groups and lipophilic tails, as well as the overall hydrophilic and lipophilic descriptors of each lecithin surfactant.

2.4.2. Cosurfactants' descriptors

Cosurfactants were incorporated as pure materials within the collected microemulsion systems. Accordingly, their descriptors were directly utilized in statistical modeling without further processing (Table 4).

2.4.3. Oils' descriptors

Generally, oils were incorporated as pure materials in most of the collected microemulsions. Accordingly, their descriptors were directly utilized in statistical modeling without further processing (Table 5). However, few microemulsion systems contained complex oil mixtures, i.e., miglyol 812 and soybean oil (systems 7, 12, 13, 19 and 20 in Table 1). The overall descriptors of complex oils were calculated as the average descriptors of their individual components:

$$\text{descriptor of complex oil} = \sum R_i \times \text{Di}$$

where R_i is the ratio (w/w) of a particular oil component and Di is the molecular descriptor of that component. Table 5 illustrates the descriptors of individual components within soybean oil and miglyol 812, respectively. The table also show the overall calculated descriptors for both oils.

2.5. Mass ratio descriptor

The surfactant-to-cosurfactant fixed ratio (w/w), also known as Km, is an important factor affecting microemulsion existence area (Aboofazeli et al., 1994b). This ratio is expressed herein by the "surfactant ratio" (SR) descriptor (w/w):

$$\text{SR} = \frac{\text{surfactant}}{\text{surfactant} + \text{cosurfactant}}$$

2.6. Statistical modeling employing stepwise backward regression analysis

In this technique, the modeler starts by constructing the largest possible statistical model by including all possible descriptors as explanatory (X) variables and the monitored response as the Y variable. Subsequently, the modeler eliminates, in a stepwise manner, problematic descriptors, i.e., redundant and collinear variables, until achieving the simplest regression equation capable of explaining significant percentage (>85%) of the variation in the response Y variable. The modeler usually stops removing explanatory descriptors upon reaching to a point were any further deletion of descriptors causes dramatic fall in R^2 (Ramsey and Schafer, 1997; SciVision, 1999).

Backward stepwise regression modeling was performed within SAS® software environment. The collected phase diagrams were classified into two categories, water in oil (W/O) and oil in water (O/W) systems. Subsequently, microemulsion percent areas (O/W-ME% and W/O-ME%) together with the corresponding molecular descriptors were transferred into an excel spreadsheet, which was imported into SAS®. The stability areas (O/W-ME% and W/O-ME%) were enlisted as independent (response) variables, while the

Table 4
The physicochemical descriptors calculated for the surfactants and cosurfactants incorporated in the microemulsions enlisted in Table 1

Microemulsion component	Compound or fragment	$^{1}\chi$	$^{V}\chi^{0}$	$^{V}\chi^{1}$	log *P*	Volume	WI	$^{3}\chi$	MW	Sp Pol	Dipole	Polar	MaxQ^{+}	$\kappa_{\alpha 3}$	ABSQ	ABSQon	MaxQ^{-}
Hydrophilic Segments of lecithins	Lysophosphatidylcholin	8.16	11.67	6.76	−2.09	242.10	751	4.15	285	0.0940	24.63	22.80	0.240	11.38	5.19	2.71	−1.000
	Phosphatidylcholine	9.20	12.62	7.21	2.09	261.24	980	4.67	313	0.0950	24.29	24.73	0.240	11.72	5.65	2.93	−1.000
	Phosphatidylethanolamine	8.05	9.88	6.05	−2.19	208.91	614	4.10	271	0.0920	25.09	19.22	0.230	8.93	5.41	2.92	−1.000
	Phosphatidylinositol	11.64	13.45	8.29	0.61	289.92	1606	8.35	389	0.0970	9.60	28.20	0.230	6.04	8.25	4.62	−1.000
	Phosphatidylserine	9.32	11.06	6.57	3.81	230.86	967	5.31	314	0.0900	14.47	20.75	0.260	9.25	7.07	4.27	−1.000
Lipophilic Segments of lecithins	Arachidonic acid	9.41	12.98	8.01	5.70	308.35	1140	4.21	260	0.1131	0.27	34.87	0.057	16.96	2.28	0.00	−0.088
	Linoleic acid	8.41	12.09	7.71	5.81	287.36	816	3.71	236	0.1099	0.26	31.59	0.057	15.48	1.97	0.00	−0.088
	Linolenic acid	8.41	11.83	7.36	5.80	280.85	816	3.71	234	0.1118	0.29	31.39	0.057	15.22	2.02	0.00	−0.088
	Oleic acid	8.41	12.35	8.06	5.81	293.65	816	3.71	238	0.1082	0.13	31.78	0.057	15.74	1.92	0.00	−0.088
	Palmitic acid	7.41	11.19	7.41	5.72	266.08	560	3.21	212	0.1064	0.01	28.30	0.027	14.00	1.65	0.00	−0.065
	Palmitoleic acid	7.41	10.93	7.06	5.63	259.64	560	3.21	210	0.1083	0.14	28.11	0.057	13.74	1.70	0.00	−0.088
	Stearic acid	8.41	12.61	8.41	5.81	300.08	816	3.71	240	0.1065	0.01	31.97	0.027	16.00	1.87	0.00	−0.065
Overall lecithins[a]	Epikuron 200																
	Hydrophilic part	9.16	12.56	7.18	1.92	260.15	969	4.65	312	0.0946	24.31	24.61	0.24	11.68	5.63	2.93	−1.000
	Lipophilic part	8.04	11.65	7.49	5.62	276.87	761	3.54	227	0.1060	0.20	30.26	0.051	14.87	1.86	0.00	−0.081
	Epikuron 170																
	Hydrophilic part	8.98	11.93	6.92	1.34	248.35	940	4.74	305	0.0923	22.45	23.41	0.230	10.62	5.73	3.05	−1.000
	Lipophilic part	8.04	11.65	7.49	5.62	276.87	761	3.54	227	0.1060	0.20	30.26	0.051	14.87	1.86	0.00	−0.081
	Ovothin 200																
	Hydrophilic part	8.96	12.28	7.03	1.83	254.40	946	4.55	305	0.0927	23.83	24.06	0.230	11.42	5.52	2.87	−1.000
	Lipophilic part	8.21	12.06	7.85	5.83	286.70	752	3.60	233	0.1092	0.12	31.02	0.045	15.31	1.87	0.00	−0.079
Cosurfactants	*n*-Butanol	2.41	3.57	2.02	1.39	87.84	20	0.71	74	0.0996	1.46	8.75	0.210	3.96	1.09	0.40	−0.395
	sec-Butanol	2.27	3.73	1.95	1.38	87.82	18	0.82	74	0.0997	1.51	8.75	0.210	3.96	1.05	0.39	−0.392
	Isobutanol	2.27	3.73	1.88	1.38	87.75	18	0.82	74	0.0997	1.47	8.75	0.210	3.96	1.08	0.39	−0.395
	Tert-Butanol	2.00	3.95	1.72	1.37	87.79	16	0.00	74	0.0997	1.57	8.75	0.210	0.00	1.00	0.39	−0.390
	1,2-Butanediol	2.81	3.89	2.10	0.85	96.19	31	1.39	90	0.0976	2.17	9.39	0.210	2.92	1.75	0.78	−0.392
	Diethylene glycol monobutyl ether	5.41	7.21	4.18	2.32	173.46	220	2.21	162	0.1001	1.46	17.37	0.210	9.88	2.59	1.15	−0.393
	n-Propranol	1.91	2.86	1.52	0.50	70.91	10	0.50	60	0.0975	1.47	6.92	0.210	4.00	0.98	0.40	−0.395
	Isopropanol	1.73	3.03	1.41	0.45	70.82	9	0.00	60	0.0977	1.59	6.92	0.210	0.00	0.94	0.39	−0.392
	n-Pentanol	2.91	4.28	2.52	1.51	104.73	35	0.96	88	0.1011	1.47	10.59	0.210	5.30	1.19	0.40	−0.395
	n-Pentanoic acid	3.27	4.44	2.47	1.21	104.17	52	1.13	101	0.0926	12.98	9.65	0.204	5.47	2.01	1.35	−1.000
	1,2-Pentanediol	3.31	4.59	2.60	1.37	113.15	50	1.48	104	0.0992	2.17	11.22	0.210	3.76	1.85	0.78	−0.392
	Diethylene glycol monopentyl ether	5.91	7.92	4.68	2.53	190.49	286	2.46	176	0.1008	1.47	19.20	0.209	11.00	2.70	1.15	−0.393
	n-Hexanol	3.41	4.98	3.023	1.676	121.77	56	1.21	102	0.1020	1.46	12.42	0.209	5.96	1.30	0.40	−0.395
	n-Hexanoic acid	3.77	5.15	2.968	1.762	121.15	79	1.39	115	0.0948	15.81	11.48	0.204	6.69	2.12	1.35	−1.000
	1,2-Hexanediol	3.81	5.30	3.098	1.778	130.12	76	1.75	118	0.1004	2.165	13.06	0.210	4.92	1.96	0.78	−0.392
	Diethylene glycol monohexyl ether	6.41	8.63	5.178	2.569	207.37	364	2.71	190	0.1014	1.463	21.04	0.209	11.90	2.80	1.15	−0.393

[a] The corresponding descriptors were calculated utilizing the percentages of individual component phospholipids as in Table 3 and associated reference.

Table 5
The physicochemical descriptors calculated for the oil components incorporated in the microemulsions enlisted in Table 1

Oil	Component molecules	$^{1}\chi$	$^{v}\chi^{0}$	$^{v}\chi^{1}$	log *P*	Volume	WI	$^{3}\chi$	MW	Sp Pol	Dipole	Polar	MaxQ^{+}	$\kappa_{\alpha 3}$	ABSQ	ABSQon	MaxQ^{-}
Ethyl octanoate	–	5.81	8.27	4.97	2.731	191.84	265	2.58	172	0.1035	1.42	19.85	0.247	8.63	1.95	0.62	−0.327
Ethyl oleate	–	10.81	15.08	9.62	3.395	355.16	1720	5.08	311	0.1070	1.51	38.01	0.247	18.37	3.06	0.62	−0.327
Octanoic acid (octanoate ion)	–	4.77	6.56	3.97	1.517	155.13	158	1.89	143	0.0977	21.56	15.15	0.204	8.62	2.33	1.35	−1.000
Oleic acid (oleate ion)	–	9.77	13.37	8.62	4.363	318.47	1313	4.39	281	0.1046	47.34	33.31	0.204	18.27	3.44	1.35	−1.000
Soybean oil[a]	Linoleic acid triglycerides	31.03	41.33	26.44	2.530	970.29	30600	15.80	879	0.1100	1.43	106.40	0.247	50.19	9.06	1.83	−0.327
	Linolenic acid triglycerides	31.03	40.55	25.38	2.530	950.71	30600	15.80	873	0.1110	1.29	105.80	0.247	49.42	9.21	1.83	−0.327
	Oleic acid triglycerides	31.03	42.10	27.49	2.530	990.43	30600	15.80	886	0.1080	1.73	107.00	0.247	50.96	8.91	1.83	−0.327
	Palmitic acid triglycerides	28.03	38.64	25.54	2.530	908.18	22558	14.30	807	0.1060	1.36	96.53	0.247	45.82	8.12	1.83	−0.327
	Stearic acid triglycerides	31.03	42.88	28.54	2.530	1009.10	30600	15.80	892	0.1070	1.34	107.50	0.247	51.73	8.75	1.83	−0.327
	Overall virtual soybean oil	30.08	40.38	26.05	2.479	948.66	29103	15.30	855	0.1070	1.44	103.30	0.242	48.89	8.74	1.80	−0.327
Miglyol 812[a]	Octanoic acid triglycerides	16.03	21.67	13.54	2.530	500.59	4270	8.27	471	0.1050	1.41	52.49	0.247	22.45	5.57	1.83	−0.327
	Decanoic acid triglycerides	19.03	25.92	16.54	2.530	602.48	7096	9.77	555	0.1050	1.40	63.50	0.247	28.22	6.21	1.83	−0.327
	Overall virtual Miglyol 812	17.53	23.79	15.04	2.530	551.53	5683	9.02	513	0.1050	1.40	57.99	0.247	25.34	5.89	1.83	−0.327

[a] The corresponding descriptors were calculated utilizing the percentages of individual component oils as in Table 2 and associated reference (Wade and Weller, 1994).

molecular descriptors were considered as explanatory variables.

Modeling commenced by constructing initial tentative models that included all molecular descriptors as explanatory variables and the respective microemulsion areas (O/W-ME% and W/O-ME%) as response variables. Thereafter, a variety of mathematical transformations (e.g., root, power, reciprocal and logarithmic transformations) were applied on the response variables (i.e., O/W-ME% and W/O-ME%) to find the optimal transformations that yield the best statistical criteria, i.e., R^2, *F*-statistic, mean square error (MSE) and the homogeneity of the residuals plots. The logarithmic transformations of microemulsion stability areas provided the best results. Afterwards, a descriptor-assessment process was performed, such that each descriptor was removed and reincorporated to evaluate its significance on the statistical parameters of the tentative model. The process was repeated till all descriptors were assessed. Eventually, descriptors that illustrated any effect on the statistical criteria of the model (R^2 and *F*-value) were collected and considered for subsequent modeling steps.

Subsequently, a cross-correlation matrix was constructed to assess the colinearity patterns among surviving descriptors. The least significant descriptor within any set of collinear variables ($r^2 > 0.90$) was removed from the tentative model. Successful statistical regression models should not include any collinear variables, i.e., explanatory variables of similar meanings. The emergence of collinear variables in a particular QSPR equation leads to significant prediction errors resulting from combining their parallel errors (SciVision, 2000, 1999).

Subsequently, redundant explanatory variables were removed in a stepwise manner according to their respective probability-of-significance values (*p*-values). The variable with the highest *p*-value was 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 (95% significance level).

Backward stepwise regression analysis succeeded in modeling the formation of W/O microemulsions, however, it failed completely with O/W microemulsions. The optimal W/O-QSPR model was further enhanced by removing statistical outliers, i.e., systems number

6, 22, 23, 64 and 67 in Table 1. The final W/O model was free from collinear descriptors (cross-correlation threshold of $R^2 \geq 0.65$).

2.7. Statistical modeling utilizing genetic algorithm (GA)

The genetic algorithm (GA) embedded in QSARIS® was employed in the current study. Microemulsion percent areas (O/W-ME% and W/O-ME%) and corresponding molecular descriptors were imported into QSARIS®. The logarithmic transformations of O/W-ME% and W/O-ME% were enlisted as independent response variables, while the corresponding calculated molecular descriptors were enrolled as explanatory variables.

GA techniques rely on the evolutionary operations of "crossover and mutation" to select an optimal combination of descriptors capable of explaining microemulsion stability across diverse training systems. GA operates through a cycle of the following stages: (i) encoding mechanism; (ii) definition of a fitness function; (iii) creating a population of chromosomes; (iv) genetic manipulation of chromosomes (Hall et al., 2001; SciVision, 2000).

The coding scheme used in QSARIS® is gene-based. In this scheme, the possible regression models (chromosomes) differ from one another by the set of independent variables (descriptors) that comprise each model. If the general number of independent variables (descriptors) is equal to P (in this particular case $P = 49$ variables, 16 descriptors for each of the three microemulsion components plus the SR descriptor), then any chromosome corresponding to any model consists of a string of P binary digits (bits) called "genes". Each value in the string represents an independent variable (0 = absent, 1 = present). Each chromosome is associated with a fitness value that reflects how good it is compared to other solutions. From diagnostic experiments, it was decided to employ the adjusted correlation coefficient (R^2) of each chromosome-model as the fitness function, as it seems to allow optimal microemulsion QSPR models to emerge.

The following points describe subsequent GA selection steps and related control parameters (Hall et al., 2001; SciVision, 2000):

- Creating an initial population: The user must specify a number of initial random chromosomes. In the current research, we decided to start with 100 initial random chromosomes.
- Choosing a parent: parent selection in GA aims at providing more reproductive chances (mating) for the fittest chromosomes. Our diagnostic trials indicated that the "Tournament Selection" option yielded optimal models. In this scheme the individual chromosome must win a competition with a randomly selected set of chromosomes. The winner of the tournament is the chromosome with the highest fitness of the tournament competitors. The winner is then incorporated in a mating pool composed of tournament winners, which drives the genetic algorithm to improve the fitness of each succeeding generation (Angeline, 1995; Hall et al., 2001; SciVision, 2000).
- Mating process: Mating is an operation during which two parents' chromosomes are combined to generate new solutions (offspring). For a couple of parents two parameters are to be configured. (i) The probability of mating, which can take values between 0.0 and 1.0 (set to 0.90 in the current project). (ii) The number of offspring chromosomes from the same parents (set to 2 in the present work). QSARIS® offers three possible crossover operators for mating: (a) One-point crossover; (b) two-point crossover; (c) uniform crossover. Diagnostic trials performed on the current data indicated that uniform crossover yielded superior QSPR models. In uniform crossover, each gene, for a given offspring, can be independently chosen from one parent or the other. The other offspring simply receives the complementary value (Angeline, 1995; Hall et al., 2001; SciVision, 2000).
- Mutation operator: this operator modifies any single chromosome with a given probability, which can take values between 0.0 and 1.0 (set to 0.70 in the current project). A mutation operator changes one or more bits in the chromosome to its complement. It is possible to define single or two-point randomly chosen mutations. In addition, uniform mutation is possible, where at least one gene is changed (Angeline, 1995; Hall et al., 2001; SciVision, 2000). Diagnostic trials on the current data suggested uniform mutation as the optimal choice.
- The offspring process, which aims at displacing an existing member with better offspring. The following variants are possible in QSARIS®: (i) Replace weakest members of the population with the best

Table 6
GA control parameters employed in the QSPR modeling of W/O microemulsions

	Parameter	Description
1	Initial population	100
2	Mating	Uniform crossover
3	Probability of mating	0.9
4	Mutation	Uniform mutation
5	Mutation parameter	0.7
6	Choosing parents	Tournament selection
9	Number of offspring from the same parents	2
7	Number of the all generated offspring for population update	60
8	Number of replaced worst parent solutions for best offspring solutions	6
10	Probability for a variable to be included	0.05
11	Total number of generations	4000
12	Fitness function	Adjusted R^2

offspring, and (ii) derive the next population from the best solutions only. In the present case we achieved optimal QSAR models through replacing the weakest parent chromosomes with best offspring.

- Maximum number of generations: this is needed to exit from GA basic cycle and to complete the algorithm.

Optimal GA parameters were configured experimentally as it is practically impossible to foretell their corresponding effects. Consequently, we conducted few diagnostic modeling trials to arrive to the best possible GA configuration.

GA-MLR modeling succeeded in producing a QSPR model that describes W/O microemulsion formation, however, it failed completely in developing O/W-QSPR model. Table 6 summarizes the optimal GA parameters employed in the development of optimal W/O microemulsion QSPR model.

The final W/O-QSPR model was further optimized by removing statistical outliers: systems 6, 22, 23, 64, 68, and 70 in Table 1.

2.8. Validation of the optimal models

Optimal QSPR models were cross-validated as follows. The total training set was divided into two subsets (after removing the outliers): fit and test subsets. The test subset was randomly selected to represent ca. 20% of the total mined microemulsion systems. This procedure was repeated three times; accordingly, three test subsets with their corresponding training fit sets were selected for cross-validation. The three test sets covered ca. 60% of the total data points. The procedure avoided selecting the same data point in more than one test subset (Ramsey and Schafer, 1997; Maran et al., 1999; Taha et al., 2002; SciVision, 1999).

The fit sets were utilized to generate three sub-models employing the same group of descriptors that emerged in the original QSPR model undergoing validation. The resulting sub-models were utilized to predict percent microemulsion areas of the corresponding test sets. Finally, the predicted values were correlated with their experimental counterparts for each test subset to determine the corresponding test correlation coefficients.

Tables 7 and 8 illustrate the training and testing subsets employed in the validation of the optimal QSPR equations. However, Tables 9 and 10 summarize the cross-validation results of Eqs. (1) and (2).

3. Results

3.1. QSPR models

Upon exploring various statistical modeling strategies, two methods were found to yield the best results, namely, backward regression analysis and genetic algorithm-based QSPR modeling.

Eq. (1) illustrates the final QSPR model achieved for W/O microemulsions employing backward regression analysis. The 95% confidence limits (CL) of different regression coefficients are shown in brackets ([±CL]).

$$\begin{aligned}\log(\text{W/O-ME\%}) = {} & -11.116[\pm 18.988] + 0.313[\pm 0.125]\text{SR} - 0.103[\pm 0.069]\text{Co-}^{1}\chi \\ & -0.121[\pm 0.106]\text{Co-}^{3}\chi + 0.021[\pm 0.022]\text{Co-}\kappa_{\alpha 3} + 61.687[\pm 88.880]\text{Co-MaxQ}^{+} \\ & +0.685[\pm 0.285]\text{Co-ABSQon} + 0.900[\pm 1.000]\text{Co-MaxQ}^{-} + 0.047[\pm 0.055]\text{Co-log}P \\ & -0.236[\pm 0.043]\text{O-ABSQon} - 0.144[\pm 0.066]\text{O-MaxQ}^{-}, \\ n = 89,\ R^2 = 0.87,\ F = 50.40,\ \text{MSE} = 0.0022 & \qquad (1)\end{aligned}$$

Table 7
Fit and test subsets utilized in the cross validation of the W/O QSPR model Eq. (1) developed employing backward stepwise regression analysis

Set to fit (training)		Set to test	
Subset no.	Data points[a]	Subset no.	Data points[a]
1	1, 2, 4, 5, 7, 8, 10–13, 15–18, 20, 21, 24–27, 29, 30, 32–35, 37, 38, 40, 42–45, 47–50, 52–55, 57–60, 62, 63, 65, 68–71, 73, 74–76, 78–81, 83–86, 88–91, 93 and 94	2	3, 9, 14, 19, 28, 31, 36, 39, 41, 46, 51, 56, 61, 66, 72, 77, 82, 87 and 92
3	1, 2, 3, 5, 8–16, 18–21, 25, 26, 28, 29, 31–34, 36–42, 44–46, 48, 50–54, 56–58, 60–62, 65, 66, 69—83, 85–87, 89–92 and 94	4	4, 7, 17, 24, 27, 30, 35, 43, 47, 49, 55, 59, 63, 68, 76, 80, 84, 88 and 93
5	1, 3, 4, 5, 7, 9–12, 14–19, 24, 25, 27, 28, 30–33, 35–41, 43–47, 49–56, 58, 59, 61, 63, 65, 66, 68, 70, 72—77, 79–88, 90, 92, 93 and 94	6	2, 8, 13, 20, 21, 26, 29, 34, 42, 48, 57, 60, 62, 69, 71, 78, 83, 89 and 91

[a] Numbers as in Table 1.

Table 8
Fit and test subsets utilized in the cross validation of the W/O QSPR model Eq. (2) developed employing genetic algorithm and multiple linear regression analysis (GA-MLR)

Set to fit (training)		Set to test	
Subset no.	Data points[a]	Subset no.	Data points[a]
1	1, 2, 4, 5, 7, 8, 10–13, 15–18, 20, 21, 24–27, 29, 30, 32–35, 37, 38, 40, 42–45, 47–50, 52–55, 57–60, 62, 63, 65, 67, 69, 71, 73–76, 78–81, 83–86, 88–91, 93 and 94	2	3, 9, 14, 19, 28, 31, 36, 39, 41, 46, 51, 56, 61, 66, 72, 77, 82, 87 and 92
3	1–3, 5, 7–12, 14–16, 18–21, 25, 26, 28, 29, 31–34, 36–42, 44–46, 48, 50–54, 56–58, 60–62, 65, 66, 69, 71–75, 77–79, 81–83, 85–87, 89–92 and 94	4	4, 13, 17, 24, 27, 30, 35, 43, 47, 49, 55, 59, 63, 67, 76, 80, 84, 88 and 93
5	1, 3, 4, 5, 7, 9–12, 14–19, 24, 25, 27, 28, 30–33, 35–41, 43–47, 49–56, 58–61, 63, 65–67, 72–77, 79, 80, 81, 82, 84–88, 90, 92, 93 and 94	6	2, 8, 13, 20, 21, 26, 29, 34, 42, 48, 57, 60, 62, 69, 71, 78, 83, 89 and 91

[a] Numbers as in Table 1.

Table 9
Cross validation of W/O microemulsion model Eq. (1)

Set to fit[a]	n	R^2 (fit)	F (fit)	MSE (fit)	Set to predict[a]	n	R^2 (test)	F (test)	MSE (test)
1	70	0.87	37.84	0.0024	2	19	0.84	86.60	0.0027
3	70	0.85	34.54	0.0024	4	19	0.91	182.26	0.0017
5	70	0.88	42.89	0.0019	6	19	0.85	95.42	0.0029

[a] The subsets' numbers are as in Table 7.

Table 10
Cross validation of W/O microemulsion model Eq. (2)

Set to fit	n	R^2 (fit)	F (fit)	MSE (fit)	Set to predict[a]	n	R^2 (test)	F (test)	MSE (test)
1	69	0.89	33.44	0.0021	2	19	0.89	137.75	0.0015
3	69	0.88	31.38	0.0020	4	19	0.84	88.03	0.0030
5	69	0.91	40.82	0.0015	6	19	0.81	73.96	0.0036

[a] The subsets' numbers are as in Table 8.

where n is the number of training microemulsion systems, R^2 the correlation coefficient, F the Fisher statistic, MSE the means square error. The definitions and abbreviations of the different descriptors are shown in Section 2.4. The prefixes HS-, LS-, Co- and O-, were added to the descriptors' abbreviations to denote hydrophilic surfactant segment, lipophilic surfactant segment, cosurfactant and oil descriptors, respectively, e.g., O-MaxQ$^-$ denotes the maximum negative charge within oil molecules.

Fig. 4 illustrates the scatter plot of calculated log (W/O-ME%) values produced by Eq. (1) versus the corresponding experimental log (W/O-ME%) values.

On the other hand, the combination of genetic algorithm and multiple linear regression analysis (GA-MLR) yielded Eq. (2) as the most optimal QSPR equation after 4000 iterations.

$$\begin{aligned}\log(\text{W/O-ME\%}) = {} & -41.662[\pm 17.540] + 0.340[\pm 1.480]\text{SR} - 4.175[\pm 3.940]\text{LS-ABSQ} \\ & + 1.487 \times 10^{-4}[\pm 5.800 \times 10^{-4}]\text{Co-WI} + 315.200[\pm 93.570]\text{Co-MaxQ}^+ \\ & + 1.661[\pm 0.730]\text{Co-ABSQ} - 3.694[\pm 1.586]\text{Co-ABSQon} - 3.649[\pm 1.500]\text{Co-MaxQ}^- \\ & + 0.172[\pm 0.102]\text{Co-log}P - 169.000[\pm 71.320]\text{Co-SpPol} - 0.079[\pm 0.082]\text{O-}{}^0\chi^{\text{V}} \\ & + 0.127[\pm 0.170]\text{O-}{}^3\chi + 0.02032[\pm 0.017]O\text{-}\kappa_{a3} - 0.1819[\pm 0.071]\text{O-ABSQon}, \end{aligned}$$

$$n = 88,\ R^2 = 0.89,\ F = 46.11,\ \text{MSE} = 0.0019 \quad (2)$$

Fig. 5 illustrates the scatter plot of calculated log (W/O-ME%) values produced by Eq. (2) versus the corresponding experimental log (W/O-ME%) values.

The correlation coefficients (R^2) of both QSPR models indicate they can explain the variation in the stability domains of 87% and 89% of the collected W/O microemulsion systems, which correspond to significant explanatory capacities. However, the remaining unexplained variation in W/O microemulsion stabilities (13% and 11%, respectively) is probably due to impurities in different microemulsion components, unreported variations in the preparation conditions (e.g., temperature, stirring, etc.) or certain inter-operator variations.

However, both QSPR models exhibited high calculated F-values (Fisher statistic) suggesting that the collected W/O microemulsion systems represent the overall population of lecithin-based pseudoternary W/O microemulsions, i.e., the modeled observations are good samples of the overall population, and that the models are predictive and represent real relationships and not just chance correlations. Furthermore, subsequent extensive cross-validation illustrated the statistical significance and predictive powers of both Eqs. (1) and (2) as shown in Tables 9 and 10 (see Section 3.2).

It remains to be mentioned that despite exploring different statistical modeling strategies, all attempts to develop significant QSPR model(s) for O/W microemulsions proved futile. The best achieved O/W statistical model was of very poor criteria ($R^2 = 0.5$, $F = 2.9$, MSE = 0.43).

3.2. Cross-validation of the successful models

We implemented the leave-20%-out crossvalidation protocol often utilized to assess the predictive potential of statistical regression models (Ramsey and Schafer, 1997). Tables 9 and 10 summarize the results of three rounds of cross-validation performed in Eqs. (1) and (2). The average values of fit and test correlation coefficients (R^2) for Eq. (1) were found to be identical (i.e., 0.87), while they were 0.89 and 0.84, respectively, for Eq. (2). The fact that the test R^2 values over three randomized trials ranged from 0.81 to 0.91 indicate that the models can explain the variation in microemulsion stability regions of 81–91% of the microemulsions within test set, which correspond to good statistical significance and predictive powers. Furthermore, the results rule-out the possibility of chance correlation between the selected physicochemical descriptors and W/O microemulsion stabilities.

4. Discussion

4.1. Molecular and statistical modeling and descriptor calculations

It is possible, in principle; to collect all the information that predetermines the chemical, biological,

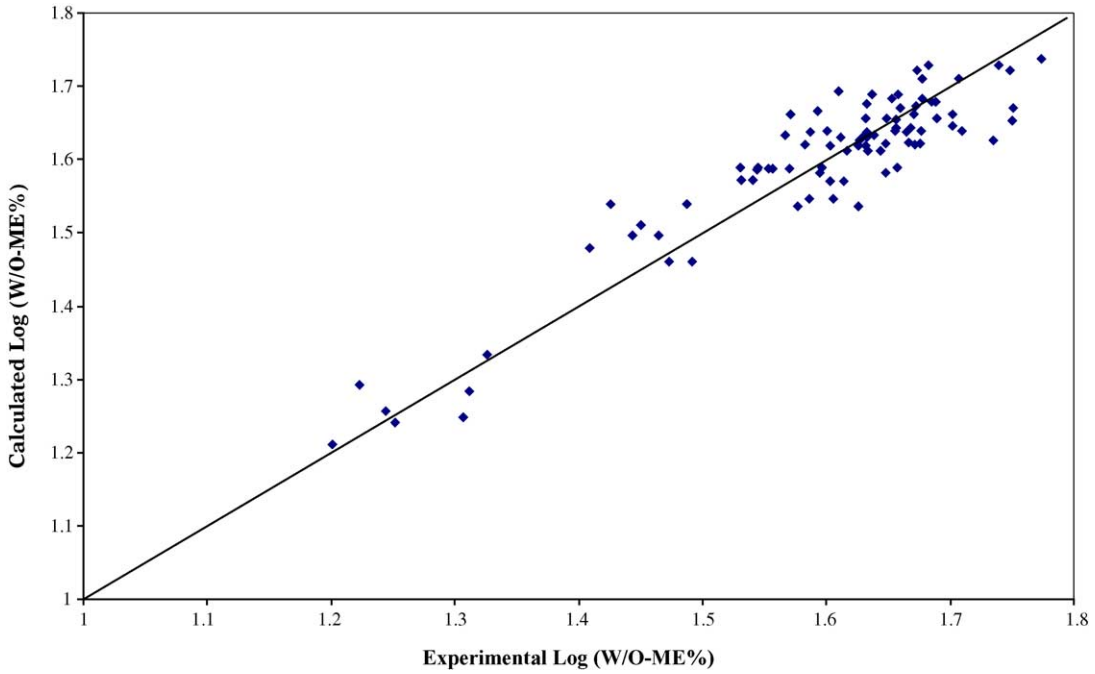

Fig. 4. Scatter plot of calculated log (W/O-ME%) values produced by Eq. (1) vs. the corresponding experimental log (W/O-ME%) values.

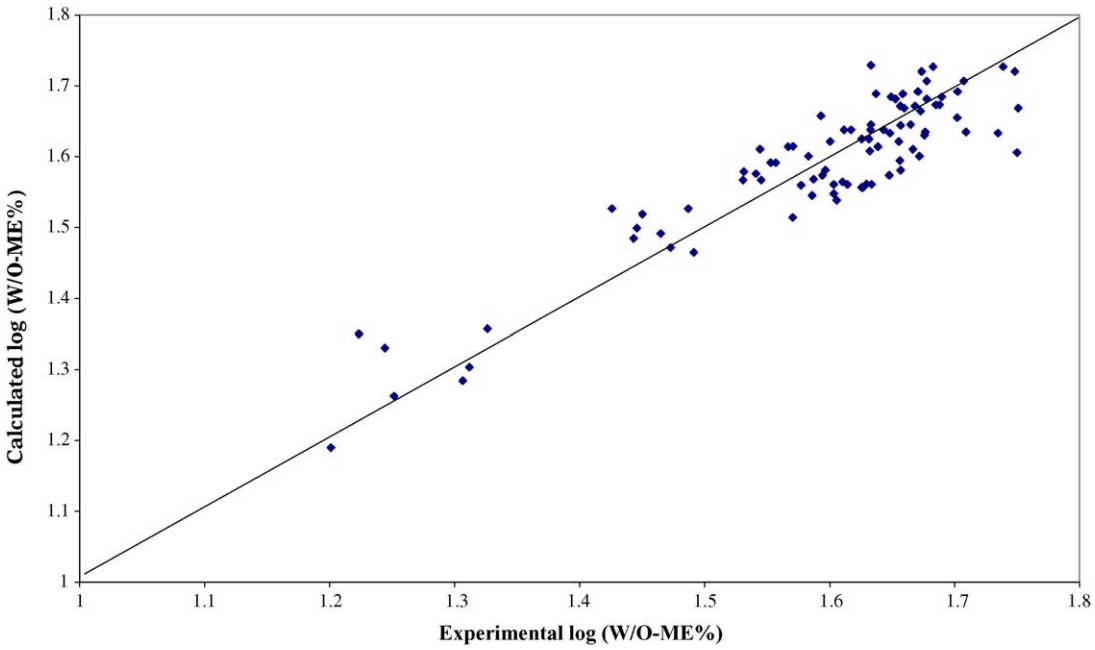

Fig. 5. Scatter plot of calculated log (W/O-ME%) values produced by Eq. (2) vs. the corresponding experimental log (W/O-ME%) values.

and physical properties of a particular compound from its chemical formula (Grover et al., 2000). On the other hand, quantitative structure activity relationships (QSARs) or structure property relationships (QSPRs) define mathematically the statistical relationship between a given type of activity (chemical, biological or physical) within a set of individual compounds and one or more physicochemical or structural parameters (Dearden, 1994).

Actually, QSARs and QSPRs are well-established in the fields of drug discovery and materials (polymer) research. These techniques convert extensive experimental data into mathematical patterns more appropriate for decision-making. For example, QSAR equations are routinely developed during drug discovery efforts by correlating the physicochemical properties of known bioactive compounds with their bioactivities. The resulting models are usually utilized to predict the bioactivities of chemical entities before preparation, which cuts time, effort and money spent in preparing less-than-optimal drug candidates (Selassie et al., 2002). In contrast to the process of microemulsion formation, which is mediated by simple molecular packing at the oil/water interface, the process of drug-receptor binding is highly specific, as it depends on the exact match between the drug molecule and the corresponding receptor. Still, QSAR analysis succeeded in constructing countless number of successful statistical models that explain variation in affinity across diverse ligands against many receptor targets. This fact prompted us to extend this interesting technique (QSAR and QSPR analyses) to analyze microemulsion formation and stability.

The stability of any microemulsion system is undoubtedly directly proportional to the degree of molecular packing within the interfacial films that separate oil and aqueous phases. Molecular packing is a function of the attractive and repulsive forces among interacting molecules. Molecular affinity is mediated by three major types of interactions: (i) electrostatic interactions (attraction of oppositely charged ions and repulsion of similarly charged ions), (ii) dipole–dipole interactions associated with partial atomic charges (e.g., hydrogen bonding), and (iii) van der Waals' attraction and repulsion forces. (Martin, 1993).

SciQSAR utilizes the three-dimensional (3D) molecular structure to calculate various descriptors to cover the electrostatic, steric, topological and hydrophobic properties of the corresponding molecules (SciVision, 1999). In the current work, the electrostatic properties of microemulsion components are encoded within a set of charge descriptors (i.e., ABSQ, ABSQon, $MaxQ^{+}$, $MaxQ^{-}$ and dipole, see Section 2.4). Furthermore, the fact that dipole-dipole interactions are mediated by partial atomic charges implies the usefulness of the same charge descriptors as means to encode dipole–dipole interactions. On the other hand, van der Waals' interactions are encoded in two types of descriptors, namely, (i) molecular volume, and (ii) molecular polarizability, i.e., Polar and Sp Pol. (see Section 2.4). Molecular volume is defined as the space occupied by the electronic clouds of atoms comprising a particular molecule (also known as van der Waals' volume). Any particular atom will exert very powerful repulsion, i.e., van der Waals' repulsion, on any entity that violates its van der Waals' space (Bodor et al., 1989). On the other hand, molecular polarizability (also known as α) and specific polarizability (polarizability per unit volume) indicate the relative ease by which the electronic cloud of certain atom is distorted (i.e., polarized) when the atom is placed in the path of an electromagnetic radiation (Miller, 1990). Accordingly, atomic polarizability is tightly related to the ability of atoms to undergo spontaneous momentary polarization and subsequent attraction towards other atoms upon contact at distances exceeding their van der Waals' radii (Miller, 1990).

However, it must be remembered that the molecular interaction forces (electrostatic, dipole, and van der Waals') are heavily dependent on the intermolecular distances between interacting molecules (Martin, 1993). Accordingly, it is necessary to encode the topological properties of interfacial molecules that control their intermolecular spatial relationships. Consequently, we employed a set of connectivity indices (WI, $\kappa_{\alpha 3}$, $^{1}\chi$, $^{3}\chi$, $^{0}\chi^{V}$ and $^{1}\chi^{V}$) to assess the influence of molecular topological factors on microemulsion stability.

Therefore, 16 different descriptors were calculated for each microemulsion component, i.e., an overall of 48 molecular descriptors that cover the three components (lecithins, cosurfactants, and oils). Subsequently, we utilized two statistical techniques, i.e., genetic algorithm and stepwise multiple linear regression analysis (see Sections 2.6 and 2.7), to select two optimal combinations of physicochemical descriptors capable

of explaining microemulsion formation and stabilities (Eqs. (1) and (2)).

However, in order to calculate the molecular descriptors of certain molecule it must be represented in three-dimensional form (3D). Accordingly, 3D models were generated for each microemulsion component employing rule-based and energy optimization methods as implemented in Alchemy 2000®. Generally, rule based methods followed by energy optimization yield molecular models of reasonable local energy minima. However, this combination usually fails in achieving molecular global minima (Goodam, 1998). Nevertheless, since the components of all collected microemulsions were treated employing the same molecular modeling sequence, i.e., rule based methods followed by energy optimization, it is expected that the overall energy-related error will be of minimal impact on the final regression model. Fig. 3 shows the optimized 3D structures of representative surfactant, oil and cosurfactant. The generated 3D structures were subsequently utilized to calculate different relevant physicochemical properties.

Being complex mixtures, the calculation of surfactants' descriptors presented a special challenge (Table 2). Consequently, it was assumed that each lecithin mixture (i.e., Epikuron 200, Epikuron 170 and Ovithin 200) can be represented as a single virtual molecule of average physicochemical properties derived from the properties of its phospholipid components. Furthermore, we were prompted to virtually split each lecithin surfactant into hydrophilic and hydrophobic segments to further simplify descriptor calculation (see Section 2.4.1). The cleavage was performed in such a way that the hydrophobic segments included hydrocarbon chains only, while the hydrophilic heads included the remaining polar and charged residues (i.e., phosphates, sugars, esters, etc.) (Taha et al., 2002).

On the other hand, the fact that all cosurfactants and most oils were incorporated in their respective microemulsions as pure well-defined compounds allowed direct uncomplicated calculation of their descriptors. Nevertheless, soybean oil and miglyol 812, both are complex oily mixtures, were modeled as virtual oil molecules of mean physicochemical properties derived from the descriptors of their individual components (see Section 2.4.3).

The assumption that complex oil or surfactant mixtures can be treated as virtual average molecules of mean physicochemical properties is not without precedence in physical pharmacy. One of the unequivocal examples on this concept is the calculation of the hydrophile–lipophile balance (HLB) values of surfactant mixtures as the average of the HLB values of the corresponding components according to their individual ratios (Attwood and Florence, 1983). The validity of this approach is supported by the fact that it allowed us to successfully derive statistically significant QSPR models for W/O microemulsions. Moreover, this concept was successfully implemented in the derivation of significant QSPR models for O/W and W/O microemulsions stabilized by non-ionic surfactants (Taha et al., 2002).

The resulting QSPR models were validated employing the leave-20%-out crossvalidation protocol (Ramsey and Schafer, 1997). Both Eqs. (1) and (2) illustrated good predictive potential against three sets of randomly selected test compounds (Tables 9 and 10). Furthermore, both equations were experimentally validated by comparing their predicted microemulsion areas with corresponding experimentally determined microemulsion domains for novel oil/lecithin/cosurfactant combinations (Abdel-Halim, 2002). However, we intend to publish our experimental findings and their relation to the statistical/molecular models latter.

4.2. Interpretation of the modeling results

4.2.1. Interpretation of W/O models

Despite the apparent differences between Eqs. (1) and (2), both models seem to encode similar information about the factors that affect the stability of W/O microemulsions.

The two QSPR equations exhibit the "surfactant ratio" descriptor (SR) indicating the necessity of certain optimal surfactant/cosurfactant combination for microemulsion stability. Nevertheless, Eq. (2) displays only one surfactant-related descriptor, namely LS-ABSQ (the sum of absolute values of charges on each atom of the lipophilic segment of surfactant molecule), while Eq. (1) lacks any surfactant-related descriptor. This noticeable under-representation of surfactant descriptors in both QSPR models suggests that the stability of W/O microemulsions is relatively independent of the type of stabilizing lecithin. This conduct is probably related to the steric bulkiness of phospholipids, which is

expected to minimize the influence of other surfactant physicochemical properties on interfacial packing, e.g., electrostatic attraction or repulsion. This conclusion is supported by the well-known fact that lecithins fail in forming stable microemulsions when used solely, indicating the importance of cosurfactant molecules for tight molecular packing at the interface.

Unsurprisingly, a variety of cosurfactant descriptors emerged in both QSPR models. In addition to the gap-filling role of interfacial cosurfactant molecules, it is established that cosurfactants contribute significantly in lowering the interfacial tension to the necessary extent required for spontaneous microemulsion formation (Kreuter, 1994).

The emergence of cosurfactant-related topological indices in Eqs. (1) and (2), i.e., Co-$^1\chi$, Co-$^3\chi$ and Co-$\kappa_{\alpha 3}$ in Eq. (1) and Co-WI in Eq. (2), suggests a significant relationship tying the molecular shape of cosurfactants with microemulsion existence areas. Both equations suggest that optimal W/O microemulsion stability requires the cosurfactant molecules to be branched and short, such that they can occupy sphere-like gaps between interfacial surfactant molecules. This is consistent with experimental findings. For example, the amount of solubilized water in oleic acid-based microemulsions increased from 32% to 45% upon changing the cosurfactant from *n*-hexanol to *n*-butanol (Aboofazeli et al., 1994b). The effect of cosurfactant branching is illustrated by comparing water solubilization in IPM-based microemulsions stabilized by *n*-butanol (36%, system 50, Table 1) and *sec*-butanol (44%, system 46, Table 1) at the same surfactant ratio.

The appearance of cosurfactant-related charge descriptors in Eqs. (1) and (2), i.e., Co-MaxQ$^+$, Co-ABSQon, Co-MaxQ$^-$ and Co-ABSQ, further supports the interfacial gap-filling role proposed for cosurfactants. These descriptors suggest that polar cosurfactants enhance interfacial molecular packing, probably through electrostatic attraction with the zwitterionic phospholipid heads, and consequently enhance the stability of corresponding microemulsions.

The emergence of Co-log *P* in Eqs. (1) and (2) emphasizes the effect of cosurfactant lipophilicity on microemulsion stability. Lipophilic cosurfactants readily migrate from the aqueous bulk towards the interface leading to better interfacial packing.

The effect of oil molecules on the stability of W/O microemulsions is evident through the appearance of oil-related descriptors in Eqs. (1) and (2). Nevertheless, the two QSPR models exhibit apparent differences in their oil-related descriptors.

The appearance of O-MaxQ$^-$ in Eq. (1) and O-ABSQon in Eqs. (1) and (2) emphasizes the influence of the electrostatic properties of oils on W/O microemulsion stability. Both descriptors are negatively correlated with microemulsion existence areas suggesting that charged oils (e.g., fatty acids) disfavor W/O microemulsion formation.

On the other hand, the emergence of O-$^0\chi^V$, O-$^3\chi$ and O-$\kappa_{\alpha 3}$ in Eq. (2) emphasizes the effect of the molecular topology of oils on W/O microemulsion stabilities. The three topological descriptors suggest that short or 3D curbed (e.g., *cis*-unsaturated fatty acids) oils promote the stability of W/O microemulsions. The overall combination of oil-related electrostatic and topology descriptors in both models indicate that charged oils, which tend to migrate from the lipophilic bulk to the interface, disturb the interfacial packing in an extent proportional to their structural elongation. This conclusion is supported by the fact that low-grade soybean lecithins, of high free fatty acid contents, generally fail in forming stable W/O microemulsions (Aboofazeli et al., 1994a).

4.2.2. Unsuccessful modeling of O/W microemulsion systems

Despite extensive exploration, all attempts to produce statistically significant QSPR model(s) for O/W microemulsions were futile. We believe this failure is related to the fact that lecithins, being more hydrophobic than hydrophilic, tend to form aqueous micellar solutions rather than O/W microemulsions. Actually, a recent article failed to unequivocaly proof the existence of lecithin-based O/W microemulsions (Aboofazeli et al., 2000). Accordingly, it is reasonable to assume that many systems in Table 1 were erroneously reported as O/W microemulsions, while they were in fact micellar solutions. Such heterogenous data mixture is probably responsible for the failure in modeling O/W microemulsions. However, failure in modeling O/W microemulsions seems to validate our W/O QSPR models as it suggests that our modeling approaches are highly sensitive to inadequate assumptions such as those we faced in O/W microemulsions. Consequently, one can conclude that the successful W/O models correspond to reasonable overall assumptions and approximations.

In conclusion, this study has shown that data mining, molecular modeling, descriptor calculation, followed by multiple linear regression analysis and genetic algorithm were successful in producing statistically significant and predictive QSPR models for W/O microemulsions. Furthermore, statistical cross-validation strengthened the significance of the resulting model. Still, the employed statistical and molecular modeling techniques failed completely in deriving any useful QSPR model connecting O/W microemulsion areas with the physicochemical properties of the corresponding components.

The resulting W/O QSPR models allowed better undertanding of the factors governing lecithin-based microemulsion formation and stability. Furthermore, these QSPR models should shorten the trial time required for the preparation of W/O microemulsions.

Acknowledgments

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®.

References

Abdel-Halim H.R., 2002. Ionic-surfactants microemulsions: prediction of their formation and stability using in vaccu molecular modeling and data mining, MSc Thesis, Jordan University, Jordan.

Aboofazeli, R., Lawrence, J.M., 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., Patel, N., Thomas, M., Lawrence, J.M., 1994a. Investigations into the formation and characterization of phospholipid microemulsions. II. Pseudo-ternary phase diagrams of systems containing water–lecithin–isopropyl myristate and alcohol: the influence of purity of lecithin. Int. J. Pharm. 106, 51–61.

Aboofazeli, R., Lawrence, B.C., Wicks, R.S., Lawrence, J.M., 1994b. Investigation into the formation and characterization of phospholipid microemulsions. III. Pseudo-ternary phase diagrams of systems containing water–lecithin–alcohol–isopropyl myristate and either an alkanoic acid, amine, alkanediol, polyethylene glycol alkyl ether or alcohol as cosurfactant. Int. J. Pharm. 111, 63–72.

Aboofazeli, R., Patel, N., Thomas, M., Lawrence, J.M., 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.

Aboofazeli, R., David, J., Barlow, M., Lawrence, J., 2000. Particle size analysis of concentrated phospholipid microemulsion. I. Total intensity light scattering. AAPS Pharm. Sci. 2, 1–13.

Angeline, P.J., 1995. Evolution revolution: An introduction to the special track on genetic and evolutionary programming. IEEE Expert Intell. Syst. Appl. 10, 6–10.

Attwood, D., Florence, T.A., 1983. Surfactant Systems, their Chemistry, Pharmacy and Biology. Chapman & Hall, New York.

Attwood, D., Mallon, G., Taylor, J.C., 1992. Phase studies on oil-in-water microemulsions. Int. J. Pharm. 84, R5–R8.

Bodor, N., Gabanyi, Z., Wong, C.K., 1989. A new method for estimation of partition coefficient. J. Am. Chem. Soc. 111, 3783–3786.

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.

Butter, H., 1993. Poucher's Perfumes, Cosmetics and Soaps, 9th ed. Chapman & Hall, UK.

Constantinides, P.P., 1995. Lipid microemulsions for improving drug dissolution and oral absorption: physical and biopharmaceutical aspects. Pharm. Res. 12, 1516–1572.

Dearden, C.J., 1994. Application of quantitative structure property relationships to pharmaceutics. Chemometr. Intell. Lab. Syst. 24, 77–87.

Fiume, Z., 2001. Final report on the safety assessment of Lecithin and Hydrogenated Lecithin. Int. J. Toxicol. 20, 21–45.

Friberg, E.S., 1990. Micelles, microemulsions, lipid 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.

Gasteiger, J., Marsili, M., 1980. Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. Tetrahedron 36, 3219–3228.

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.

Hall, L., Kier, L., Hall, M.L., 2001. QSAR Development with QsarIS. The Guide for Development of QSAR Models with QsarIS. Academic Press, UK.

Ho, H.-O., Chih-Chuan, H., Ming-Thau, S., 1996. Preparation of microemulsions using polyglucerol 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. II 72, 1525–1567.

Katritzky, A.R., Gordeeva, E.V., 1993. Traditional topological indices vs. 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. Relat. 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, 1st ed. Marcel Dekker, New York, USA.

Maran, U., Karelson, M., Katritzky, R.A., 1999. A comprehensive QSAR treatment of genotoxicity of heteroaromatics and aromatic amines. Quant. Struct. Act. Relat. 18, 3–10.

Martin, A., 1993. Physical Chemistry, 4th ed. Lea and Febriger, Philadelphia, USA.

Mathew, C., Holde, K.E., 1991. Biochemistry, 2nd ed. The Benjamin/Cummings publishing Company Inc., USA.

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, 1st ed. Wadesworth Publishing Company, USA.

Rosano, L.H., Lan, T., Weiss, A., 1979. Transparent dispersions: an investigation of some of the variables affecting their formation. J. Colloid Interf. Sci. 72, 233–244.

Schmuhl, N., Davis, E., Cheung, H.M., 1998. Morphology of thermally polymerized microporous polymer materials prepared from methyl methacrylate and 2-hydroxyethyl methacrylate microemulsions. Langmuir 14, 757–761.

SciVision, 1999. SciQSAR 3.0 User Guide. Academic Press, Massachusetts.

SciVision, 2000. QSARIS® Reference Guide: Statistical Analysis and Molecular Descriptors. Academic Press, Massachusetts.

Selassie, C.D., Mekapati, S.B., Verma, R.P., 2002. QSAR: then and now. Curr. Top. Med. Chem. 2, 1357–1379.

Taha, M., Al-Ghazawi, M., Abu-Amara, H., Khalil, E., 2002. Development of quantitative structure–property relationship models for pseudoternary microemulsions formulated with nonionic surfactants and cosurfactants: application of data mining and molecular modeling. Eur. J. Pharm. Sci. 15, 461–478.

Tenjarla, S., 1999. Microemulsions: An Overview and Pharmaceutical Applications. Crit. Rev. Ther. Drug. Carrier Syst. 16, 461–521.

Thevenin, M.A., Grossiord, J.L., Poelman, M.C., 1996. Sucrose esters/cosurfactant microemulsion systems for transdermal delivery: assessment of bicontinues structures. Int. J. Pharm. 137, 177–186.

Tripos Inc., Alchemy 2000 Reference Manual. USA, October 1998.

Wade, A., Weller, J.P., 1994. Handbook of Pharmaceutical Excepients, 2nd ed. Pharmaceutical Press, London.

Watnasirichaikul, S., Davies, M.N., Rades, T., Tucker, G.I., 2000. Preparation of Biodegradable Insulin Nanocapsules from Biocompatible Microemulsions. Pharm. Res. 17, 684–689.

Wiener, H., 1947. Correlation of Heats of Isomerization, and Differences in Heat Vaporization of Isomers, among the Paraffin Hydrocarbons. J. Am. Chem. Soc. 69, 2636–2644.

Windhole, M., 1983. The Merk Index of Chemicals and drugs, 10th ed. Merk Inc., UK.

Xu, W., Siong, K., Gao, Z., Lee, S.Y., Chow, P.Y., Gan, L.M., 1999. Microporous polymeric composite electrolyte microemulsion polymerization. Langmuir 15, 4812–4819.