Role of Genetic Algorithms and Artificial Neural Networks in Predicting the Phase Behavior of Colloidal Delivery Systems

Snezana Agatonovic-Kustrin[1,4] and Raid G. Alany[2,3]

Received December 15, 2000; accepted April 4, 2001

Purpose. A genetic neural network (GNN) model was developed to predict the phase behavior of microemulsion (ME), lamellar liquid crystal (LC), and coarse emulsion forming systems (W/O EM and O/W EM) depending on the content of separate components in the system and cosurfactant nature.

Method. Eight pseudoternary phase triangles, containing ethyl oleate as the oil component and a mixture of two nonionic surfactants and n-alcohol or 1,2-alkanediol as a cosurfactant, were constructed and used for training, testing, and validation purposes. A total of 21 molecular descriptors were calculated for each cosurfactant. A genetic algorithm was used to select important molecular descriptors, and a supervised artificial neural network with two hidden layers was used to correlate selected descriptors and the weight ratio of components in the system with the observed phase behavior.

Results. The results proved the dominant role of the chemical composition, hydrophile-lipophile balance, length of hydrocarbon chain, molecular volume, and hydrocarbon volume of cosurfactant. The best GNN model, with 14 inputs and two hidden layers with 14 and 9 neurons, predicted the phase behavior for a new set of cosurfactants with 82.2% accuracy for ME, 87.5% for LC, 83.3% for the O/W EM, and 91.5% for the W/O EM region.

Conclusions. This type of methodology can be applied in the evaluation of the cosurfactants for pharmaceutical formulations to minimize experimental effort.

KEY WORDS: GNNs; ANNs; phase behavior; microemulsion; cosurfactant; molecular descriptors.

INTRODUCTION

Pharmaceutical colloids are attracting increasing interest as vehicles for drug delivery and to provide bases for cosmetics, hence the need to understand the formulation and stabilization of these systems. Among the long line of colloidal systems that have been either examined or exploited as potential drug delivery systems are microemulsions (ME) and lyotropic liquid crystalline systems.

MEs are isotropic, optically transparent, and thermodynamically stable dispersions of oil and water stabilized by an interfacial film of amphiphile molecules, *i.e.,* suitable surfactant or a suitable combination of surfactants with or without a cosurfactant. They are spontaneously forming fluid systems of low viscosity. Most of the pharmaceutical surfactants do not lower the oil-water interfacial tension sufficiently to form microemulsions, nor are they of correct molecular structure (*i.e.,* hydrophile-lipophile balance [HLB]). Thus, a weaker amphiphile (cosurfactant) is added to further lower the interfacial tension between oil and water, influence the film curvature, and fluidize the hydrocarbon region of the interfacial film. The amphiphilic nature of low molecular weight alcohol cosurfactants enables them to distribute between the aqueous and oil phase thereby altering the chemical composition and hence the relative hydro-/lipophilicity (1). Penders and Strey pointed out the role of alcohol in microemulsions (2). They found the alcohol to act as "cosurfactant" and as "cosolvent." 1,2-Alkanediols were also used as cosurfactants to prepare nontoxic MEs with lecithin to replace the conventional aliphatic alcohols (3).

Lyotropic liquid crystalline systems are formed when surfactants are treated with a solvent. Weak interactions between the amphiphilic surfactant and solvent molecules govern the formation of association structures with long-range order known as lyotropic liquid crystals (4). The role of lyotropic liquid crystals in the stabilization of various types of pharmaceutical dispersal systems, as well as their application in drug delivery, is well established (5).

Before a microemulsion or a lyotropic liquid crystal-based system can be used for the delivery of therapeutic agents it is essential to establish the phase behavior of the particular combination of chosen components. One of the most common methods of studying the phase behavior of such systems is by constructing a phase diagram using Gibbs triangle. However, as the formulation may contain more than three components the complete phase behavior cannot be represented using a triangular diagram and becomes even more complicated as the number of the components increases (6). The phase behavior of the pseudoternary system—ethyl oleate: water: sorbitan monolaurate/polyoxyethlene 20 sorbitan monooleate—as well as the effect of the 8 n-alcohol and 1,2-alknediol cosurfactants on the phase behavior of the original pseudoternary system has been reported (7).

Artificial neural networks (ANNs) are computer programs designed to simulate the way in which the human brain processes information. ANNs learn (or are trained) through experience with appropriate examples, not from a preprogrammed set of rules. The behavior of a neural network is determined by the transfer functions of its neurons, by the learning rule, and by the architecture itself. We have used a supervised network with a back-propagation learning rule and multilayer perceptron (MLP) architecture. In this model the inputs are fully connected to the hidden layer, and hidden layer neurons are fully connected to the outputs. The presence of a hidden layer is a crucial feature that allows the network to make generalizations from the training data. Detailed description of this type of ANN model has been published (8–10). As testing large number of all possible combinations of variables would be a tedious task, we have used a genetic algorithm (GA) input selection (11–14). GA is a computational model of evolution, random search algorithm that uses selection and recombination processes to generate new population samples with higher fitness.

The goal of this research was to develop an ANN model

[1] School of Pharmaceutical Sciences, Universiti Sains Malaysia, Penang 11800, Malaysia.
[2] School of Pharmacy, University of Otago, PO Box 913, Dunedin, New Zealand.
[3] Current address: Faculty of Medical and Health Sciences, University of Auckland, Private Bag 92019, Auckland 1, New Zealand.
[4] To whom correspondence should be addressed. (e-mail: nena@usm.my)

0724-8741/01/0700-1049$19.50/0

to predict the phase behavior of quinary microemulsion-forming systems depending on the content of different components in the system and the cosurfactant nature. A nonlinear ANN model was used to correlate phase behavior of the investigated systems with cosurfactant descriptors that were preselected by a GA.

Pseudoternary phase diagrams were constructed following an accurate and invariable experimental protocol under identical laboratory conditions, thus fulfilling the requirements of successful QSAR modeling that parameters have values that are obtained in a consistent manner.

EXPERIMENTAL

Materials

Ethyl oleate (EO) was used as the oil component (Crodamol EO), sorbitan mono laurate (SML) as surfactant component 1 (Crill1, HLB = 8.6), and polyoxyethylene 20 sorbitan monooleate (PSMO) as surfactant component 2 (Crillet 4 super, HLB = 15). EO, SML, and PSMO were obtained form Croda Oleochemicals, New Zealand. Deionised water was used as the aqueous phase. 1-Propanol, 1-butanol, 1-hexanol, and 1-octanol were purchased from BDH (Poole, Dorset, UK) and were used as an alcohol cosurfactants. 1,2-Propandiol (May and Baker, Dagenham, UK), 1,2-pentanediol (Acros Organics, Geel, Belgium), 1,2-hexanediol, and 1,2-octanediol (Lancaster Synthesis, Morecomb, UK) were used as 1,2 alkanediol cosurfactants.

Apparatus and Software

Phase contrast and polarized light microscopy were performed using a Nikon Optiphot microscope (Nikon, Tokyo, Japan). MS-Windows®-based artificial neural network simulator software. Neural Networks™ (StatSoft Inc, Tulsa, OK) was used. For calculating drug properties from molecular structure, Pallas 2.1 (Compu Drug Int., San Francisco, CA) and ChemSketch 3.5 freeware (ACD Inc., Toronto, Canada) were used. Calculations were performed on a Pentium personal computer.

Construction of Pseudoternary Phase Diagrams

The detailed method used to construct the pseudoternary phase diagrams of the cosurfactant-free system and systems formulated with the 8 n-alcohol and 1,2-alkane diol cosurfactants have been reported (7).

Collecting Data for ANN Training and Testing (Sampling Theory)

The success of different sampling strategies (random, systematic, and stratified) in characterizing phase behavior was investigated using the pseudoternary phase diagram of the cosurfactant-free system (7). To evaluate simple random sampling, 171 samples were randomly picked by a scheme that ensures that each sample had an equal probability of being chosen. For the systematic sampling, the phase diagram was overlaid with a grid spacing of 5% (w/w) over each of the component axes. This gave a set of 171 vertices generated within the body of the phase diagram (Fig. 1). For stratified sampling, every fifth percentage change in component content was taken from the phase diagram, which was divided into four regions (strata): ME, lamellar mesophase (LC), water continuous coarse dispersions (O/W EM), and oil continuous coarse dispersions (W/O EM). Samples proportional to the percentage occupied by each stratum were chosen as follows: 38 samples from ME, 46 from LC, 72 from O/W EM, and 15 samples from the W/O EM stratum. The percentages occupied by the four different regions were ME 22%; LC 27%; O/W EM 42%; and W/O EM 9%.

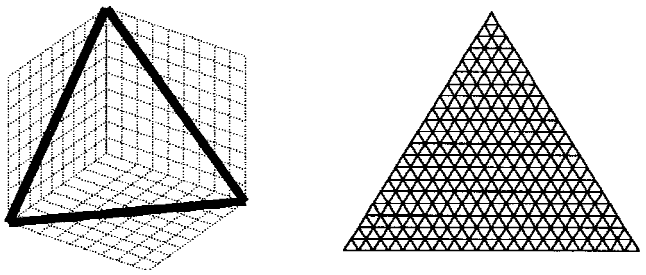

Fig. 1. Vertices and grid spacing of phase diagram for the systematic sampling strategy.

Each sample was labeled according to the proportions of surfactant blend, oil, and water in the mixture and matched with the nature of the phase structure found for this composition. Samples were coded as +1 to signify the presence of a particular system, and −1 to indicate its absence. Thus, a region consisting of ME would have values of +1, −1, −1, −1; pure LC was −1, +1, −1, −1; W/O EM was −1, −1, +1, −1; and O/W EM was categorized as −1, −1, −1, +1. ME and LC would have values of +1, +1, −1, −1; whereas LC and O/W EM were categorized as −1, +1, −1, +1. Values of −1, −1, +1, +1 would indicate the existence of an unstable coarse emulsion. A perfectly trained network should recover such values for perfect phase classification. Any deviation from +1 and −1 would reflect error in the classification process.

Descriptor Generation and Analysis

The physical and chemical properties of a compound are functions of its molecular structure. The major differences between behavior profiles of organic chemicals are attributable to physicochemical properties (15–19). A change in the structure of a molecule usually produces an associated change in its properties. Finding one or more molecular descriptors or attributes that explain variations in biological activity or physicochemical properties has resulted in the development of quantitative structure property relationships. Such relationships, once quantified, can be used to estimate the properties of other molecules from the sketches of their structures alone. Although some molecular descriptors can be determined experimentally, using computational methods to derive them is generally much faster and more convenient. Experimental determination of such properties is time-consuming and subject to large experimental variation and errors.

A total of 18 descriptors including chemical composition descriptors and calculated physicochemical descriptors for each of the cosurfactants was used for the initial ANN model (Table I). Descriptors were chosen based on their possible contribution to both interfacial and/or bulk effect of the investigated cosurfactants, hence the overall phase behavior of the system.

One of the difficulties with the large number of descrip-

Table I. Chemical Composition Descriptors and Physicochemical Descriptors Calculated from the Molecular Structure of Cosurfactants

Cosurfactant[a]	n-propanol	n-butanol	n-hexanol	n-octanol	1,2-propandiol	1,2-pentadiol	1,2-hexadiol	1,2-octadiol
HLB	7.48	7.00	6.05	5.10	9.38	8.43	7.95	7.00
N°C	3.00	4.00	6.00	8.00	3.00	5.00	6.00	8.00
L_C	5.30	6.56	9.09	11.62	5.30	7.83	9.09	11.62
V	108.10	135.00	188.80	242.60	108.10	161.90	188.80	242.60
C (%)	59.96	64.82	70.53	73.78	47.35	57.66	60.98	65.71
H (%)	13.42	13.60	13.81	13.93	10.60	11.61	11.94	12.41
O (%)	26.62	21.59	15.66	12.29	42.05	30.72	27.08	21.88
MR	17.48	22.11	31.38	40.64	18.97	28.24	32.87	42.14
V	75.50	92.00	125.00	158.00	73.40	106.40	122.90	155.90
γ	24.50	26.00	27.90	29.00	38.00	36.60	36.20	35.70
d_{20}	0.80	0.81	0.82	0.82	1.04	0.98	0.96	0.94
Log P	0.34	0.88	1.94	3.00	–1.34	–0.28	0.25	1.32
MV	125.48	152.85	207.86	262.67	121.93	176.77	204.13	259.06
Log D	0.53	1.04	2.05	3.07	–0.71	0.31	0.82	1.84
Log(1/S)	–0.40	0.20	1.50	2.80	–2.50	–1.20	–0.50	0.70

[a] HLB = hydrophile-lipophile balance value is a measure of the relative contributions of the hydrophilic and lipophilic regions of the molecule, calculated according to Davis (21). Lc = maximum chain length of the hydrocarbon chain (A). *V* = volume of the hydrocarbon chain (A^3) both estimated according to Tanford and Duke (8). C (%) = weight percentage of carbon content in molecular mass. H (%) = weight percentage of hydrogen content in molecular mass. O (%) = weight percentage of oxygen content in molecular mass. MR = molar refractivity can be calculated according to the Lorentz-Lorentz equation. V = (cm^3/mol) molar volume. γ = (dyne/cm) surface tension. d_{20n} = (g/cm^3) density. Log P = logarithm of the partition coefficient in octanol/water (22). *MV* = the molecular volumes (23) ($Å^3$). Log D = distribution coefficient at pH 7.0 in octanol/water. Log(1/S) = S represent aqueous solubility at 25°C of cosurfactant at pH 7.0.

tors is deciding which ones will provide the best regressions, considering both goodness of fit and the chemical meaning of the regression. Using a genetic algorithm for selection and following a unit penalty factor of 0.000–0.004, the number of inputs was reduced from 18 to 9. Input selection has reduced the size and complexity of the network and focused the training on the most important data. This also reduced the training time and improved network performance.

Data Analysis

Training Testing and Validation of the ANN Model

Before each training run both weights and biases were initialized with random values. During training, the performance of the ANN was evaluated with a testing data set. The training and testing data set consisted of the original data from phase diagrams containing 1-butanol, 1-hexanol, 1,2-propanediol, and 1,2-hexanediol as cosurfactants. The total number of data points consisted of 684 input/output sets and was split randomly into 548 training sets and 136 test sets. The results of the five runs were averaged. The training set was used to train the network, and the testing set was used to determine the level of generalization produced by the training set and to monitor overtraining the network, each with corresponding root mean squared (RMS) error. To calculate the RMS error, the individual errors were squared, summed, divided by the number of individual errors, and then square rooted.

For an unbiased estimate of the generalization error, the ANN was presented with a validation data set that was not used at all during the training process. Thus, the validity of the model was evaluated with the validation data set, by predicting phase behavior for systems containing 1-propanol, 1-octanol, 1,2-pentanediol, and 1,2-octanediol as cosurfactants.

Neural Network Analysis

The two forms of network analysis are model testing and sensitivity analysis. Both methods were performed concurrently with the training of the network. The testing and training set RMS errors were used to determine overall quality of a particular subset of descriptors. Training was stopped, at each run, once the error performance of the network began to deteriorate, based on the training and testing set errors, when the training RMS error failed to improve over a given number of epochs and the testing RMS error increased.

The second form of network analysis computes sensitivities of the network's outputs with respect to each of its inputs. ANNs compute the output as a sum of nonlinear transformations of linear combinations of the inputs. Sensitivity reports show the sensitivity of the output variables, as a percentage, to the changes in the input variables. If the direction of the change in the output variable is always the same as the change in investigated descriptor, then the average sensitivity is positive. The set of percentages reveals the effect that a change in a particular input has on output.

RESULTS AND DISCUSSION

Sampling Strategy

To investigate the different sampling strategies, the ANN with one hidden layer was trained by performing optimization of number and size of the weights for neuron interconnections. The lowest error was obtained with six hidden neurons. The highest coefficient of correlation was obtained with systematically sampled data. Traditionally, random sampling plans have been preferred over systematic sampling plans because that avoids subjective selection of sample locations, but systematic sampling more uniformly covered the entire sampling area. The trained ANNs reproduced the phase diagram

with a coefficient of multiple correlation of 0.95–0.99 and with the mean success of 0.975 (Table II).

ANN Models for Phase Behavior Prediction

In many problem domains, a range of input parameters is available that may be used to train a neural network, but it is not clear which of them is most significant or is needed at all. The problem is more complicated when there are interdependencies between input parameters. The only method guaranteed to select the best input set was to train the network with all possible input sets and all possible architectures, and then to select the best.

In this study the ANN model was constructed in several steps. The first step was to calculate physicochemical parameters as mathematical representations of chemical structure of investigated alcohols and 1,2-alkanediols. These parameters provided a description of the similarities and differences of investigated alcohols. A total of 15 physicochemical descriptors were calculated for the investigated cosurfactants. Comparison of the calculated descriptors of the homologue cosurfactants shows that they depend on the length of the straight hydrocarbon chain (R). As the length of R increases, the values are increased, except for the HLB values that decrease and for the density and the surface tension values of 1,2-alkanediols that decrease with the length of straight chain (Table I).

The subset of descriptors that best encodes the phase behavior of a given system was found to reduce the number of input variables, complexity and the size of the network, and to improve the network's prediction capabilities. MLP models compute the output as a sum of nonlinear transformations of linear combinations of the inputs. The number of weights and hidden units increases linearly with the number of inputs. The higher the dimensionality of the input space, the more training data sets are required. If the dimension of the input space is high, the network uses almost all its resources to represent irrelevant portions of the input space. Careful feature selection and scaling of the inputs reduces the complexity of the problem, as well as the selection of the best neural network model. To further reduce the amount of data and select the best ANN architecture, a pruning method was applied similarly to backward elimination in stepwise regression. Connections or units were eliminated during training based on unit penalty factor (Table III) and minimal generalization error.

The best nonlinear GNN model for the prediction of phase behavior was selected by comparing the prediction obtained from several high-scoring models (Table IV). The model has 14 inputs with unit penalty factor greater than 0.005. The results of ANN validation proved the dominant role of the chemical composition (%C, %H, and %O), HLB, number of carbon atoms, length of hydrocarbon chain, molecular volume, and hydrocarbon volume of the cosurfactant in prediction. Once a subset of descriptors was found, the descriptors were then correlated with the observed phase behavior using a nonlinear neural network. Network structure was optimized by heuristic search. The criterion for judging the best model was the rooted sum of mean squared error (RMS) of model predictions. Within investigated cosurfactants under identical conditions the phase behavior was approximated by a nonlinear combination of the percentages of oil and water and surfactants-cosurfactant blend and chemical composition, HLB, number of carbon atoms, length and volume of the hydrocarbon chain, molecular volume, and hydrocarbon volume of the cosurfactant.

The results suggest that a small number of chemically meaningful descriptors will provide the most predictive model. Better results were obtained with two hidden layers. A single hidden layer allows neurons to correlate improving ap-

Table II. Success of the Different Sampling Strategies in Characterizing Phase Behavior

Sampling strategy	ME	LC	O/W	W/O	
Random	0.98	0.97	0.95	0.96	0.965
Stratified	0.92	0.83	0.92	0.84	0.878
Systematic	0.99	0.98	0.98	0.95	0.975
	0.963	0.927	0.950	0.916	

Table III. Selection of the Inputs Based on the Unit Penalty Factor

Inputs	Unit penalty factor			
	0.0025	0.005	0.0075	0.01
Water (%)	1	1	1	1
Surfactant (%)	1	1	1	1
Oil (%)	1	1	1	1
Lc	1	1	1	
No_C	1	1	1	
H (%)	1	1	1	
C (%)	1	1	1	
O (%)	1	1	1	
HLB	1	1		
V	1	1		
MV	1	1		
logP	1	1		
γ	1	1		
LogD	1	1		
LogS	1			
Molar refraction	1			
V	1			
d_{20}	1			
Σ	18	14	10	3

Table IV. Overall Quality of Several High Scoring GNN Models Evaluated by a Cost Function

ANNs[a]	N	RMS_{tr}	RMS_{cv}	RMS_{val}	UP
Two hidden layers					
25-*20*-*12*-4	25	0.333	0.163	0.820	0.000
18-*17*-*10*-4	18	0.303	0.256	0.785	0.0025
14-*14*-*9*-4	14	0.235	0.231	**0.740**	0.0050
10-*6*-*4*-4	10	0.448	0.349	0.955	0.0075
One hidden layer					
25-*16*-4	25	0.306	0.486	0.832	0.000
18-*11*-4	18	0.227	0.297	0.827	0.0025
14-*9*-4	14	0.363	0.272	**0.807**	0.0050
10-*9*-4	10	0.375	0.595	0.858	0.0075
3-*3*-4	3	0.604	0.585	0.864	0.0100

Note. N – number of inputs. RMS – rooted mean squared error. UP – unit penalty factor. COST – cost function. Bold numbers are smallest validation RMS.

[a] Number of inputs – *hidden neurons* – outputs.

proximation at one point but worsening it elsewhere. On the other hand, an ANN with two hidden layers approximates the desired function by separating the decision space. For more complicated target functions, especially those with two hills or valleys (+1 and –1), it is useful to have several units in the second hidden layer. Each unit in the second hidden layer enables the net to fit a separate hill or valley. Consequently, an MLP with two hidden layers could yield an accurate approximation with fewer weights than an MLP with one hidden layer (20). Some of the neurons from the first layer partitioned the input space into regions. For each region a neuron in the second hidden layer combined the output of corresponding hidden layer neurons so that it computes the desired function within the region. In this way the effect of the neurons was isolated and the approximation in different regions was adjusted independently. Two hidden layers with 14 and 9 hidden neurons were enough to achieve good convergence on the validation data. The best model in this study had 14 inputs, 14 and 9 hidden neurons, and 4 outputs. For the systems included in the training set data, the different regions were reproduced with an accuracy of 98.69% for ME, 99.29% for LC, 98.59% for O/W EM, and 98.99% for W/O EM.

For the validation data set, various critical values (0, ± 0.5 and ± 0.75) were used to classify these data (Table V). That is, if an output value was greater than 0 the corresponding region was assumed to be present, and if less than 0 then absent. In the most stringent case, a region was assumed present if the output was greater than 0.75 and absent if less than 0.75. This approach was adapted in an attempt to improve the prediction reliability. Predictions were counted correct only when the phase behavior in a predicted region was in agreement with the experimental finding for that particular blend of components. For the best ANN model (14-14-9-4, UP 0.0050), if a critical value of zero was used, 82.5%–90.7% of data was classified correctly. Using critical value of 0.5, 82.5%–91.0% of data were classified correctly. With a more rigorous critical value of 0.75, 82.2%–91.5% data were classified correctly. Increasing the critical values of classification had little influence on the success of prediction for the systems with 25, 18, and 14 inputs, but significantly improved the reliability of prediction for the ANNs with 10 inputs. Nevertheless, the mean percentage success calculated over the validation data is very encouraging, given that only about 10% of the tetrahedron was sampled.

The best ANN model with a critical value of 0.75 predicted the phase behavior with accuracy of 82.2% for ME region, 87.5% for the LC region, 83.3% for the O/W EM, region, and 91.5% for W/O EM region.

Figure 2 shows the predicted phase triangles for systems formulated with 1-propanol, 1-octanol, 1,2-pentanediol, and 1,2-octanediol. For the 1-propanol triangle the network was successful in predicting a phase triangle with three regions only, namely O/W EM, W/O EM, and ME. The predicted phase triangle was in agreement with the experimental finding that incorporating a short chain alcohol with three to four carbon atoms perturbed the surfactant blend long-range order and subsequently substituted the LC region with a bicontinuous ME. Another interesting finding is that the predicted ME region showed that incorporation of 1-propanol balanced the surfactant blend favoring the formation of a balanced ME (24). For the 1-octanol phase triangle the boundary between the LC and the O/W EM was predicted with only slight discrepancy from the experimental boundary. The same holds true for the O/W-W/O EM boundary. The size of the pre-

Table V. Accuracy in Prediction; Percentage of Wrong Classified Data for Different Regions, Different Unit Penalty Factor, and Different Network Structure

			Critical value														
			0					0.5					0.75				
			Region					Region					Region				
ANN model	UP	Cosurfactant	ME	LC	O/W	W/O	%W	ME	LC	O/W	W/O	%W	ME	LC	O/W	W/O	%W
25-*20*-*12*-4	0.0000	n-C_3	26	0	29	4	14.8	21	0	30	7	14.5	20	0	34	9	15.8
		n-C_8	19	16	30	12	19.3	19	15	31	12	19.3	18	15	30	12	18.8
		1,2-C_5	16	17	8	9	12.5	14	17	8	9	12.0	15	17	8	9	12.3
		1,2-C_8	13	16	21	19	17.3	13	11	21	19	16.0	13	11	21	19	16.0
		%W	18.5	12.3	22.0	11.0	15.9	16.8	10.8	22.5	11.8	15.4	17.0	8.5	23.5	13.3	15.7
18-*17*-*16*-4	0.0025	n-C_3	22	0	30	6	14.5	20	0	30	6	14.0	20	0	31	6	14.3
		n-C_8	20	30	40	19	27.3	20	27	40	19	26.5	15	23	40	18	24.0
		1,2-C_5	12	6	9	24	12.8	13	16	9	23	15.3	13	16	10	22	15.3
		1,2-C_8	28	15	16	35	23.5	29	15	14	34	23.0	29	15	14	34	23.0
		%W	20.5	12.8	23.8	21.0	19.5	20.5	14.5	23.3	20.5	19.7	19.3	13.5	23.8	20.0	19.1
14-*14*-*9*-4	0.0050	n-C_3	18	1	21	6	11.5	18	10	19	6	13.3	18	10	19	6	13.3
		n-C_8	24	10	32	9	18.8	23	9	32	8	18.0	23	9	30	8	17.5
		1,2-C_5	11	16	10	10	11.8	11	16	9	10	11.5	11	16	9	10	11.5
		1,2-C_8	17	21	15	12	16.3	18	15	10	12	13.8	19	15	9	10	13.3
		%W	17.5	12.0	19.5	9.3	14.6	17.5	12.5	17.5	9.0	14.1	17.8	12.5	16.8	8.5	13.9
10-*6*-*4*-4	0.0075	n-C_3	77	100	49	68	73.5	24	0	32	5	15.3	22	0	34	5	15.3
		n-C_8	97	74	53	68	73.0	24	4	32	5	16.3	24	4	45	13	21.5
		1,2-C_5	67	83	75	65	72.5	6	17	21	13	14.3	6	17	19	13	13.8
		1,2-C_8	52	89	81	71	73.3	10	11	12	12	11.3	11	11	12	12	11.5
		%W	73.3	86.5	64.5	68.0	73.1	16.0	8.0	24.3	8.8	14.3	15.8	8.0	27.5	10.8	15.5

Note. n-C_3 = n-propanol. n-C_8 = n-octanol. 1,2-C_5 = 1,2-pentandiol. 1,2-C_8 = 1,2-octandiol. %W = % of wrong classified data.

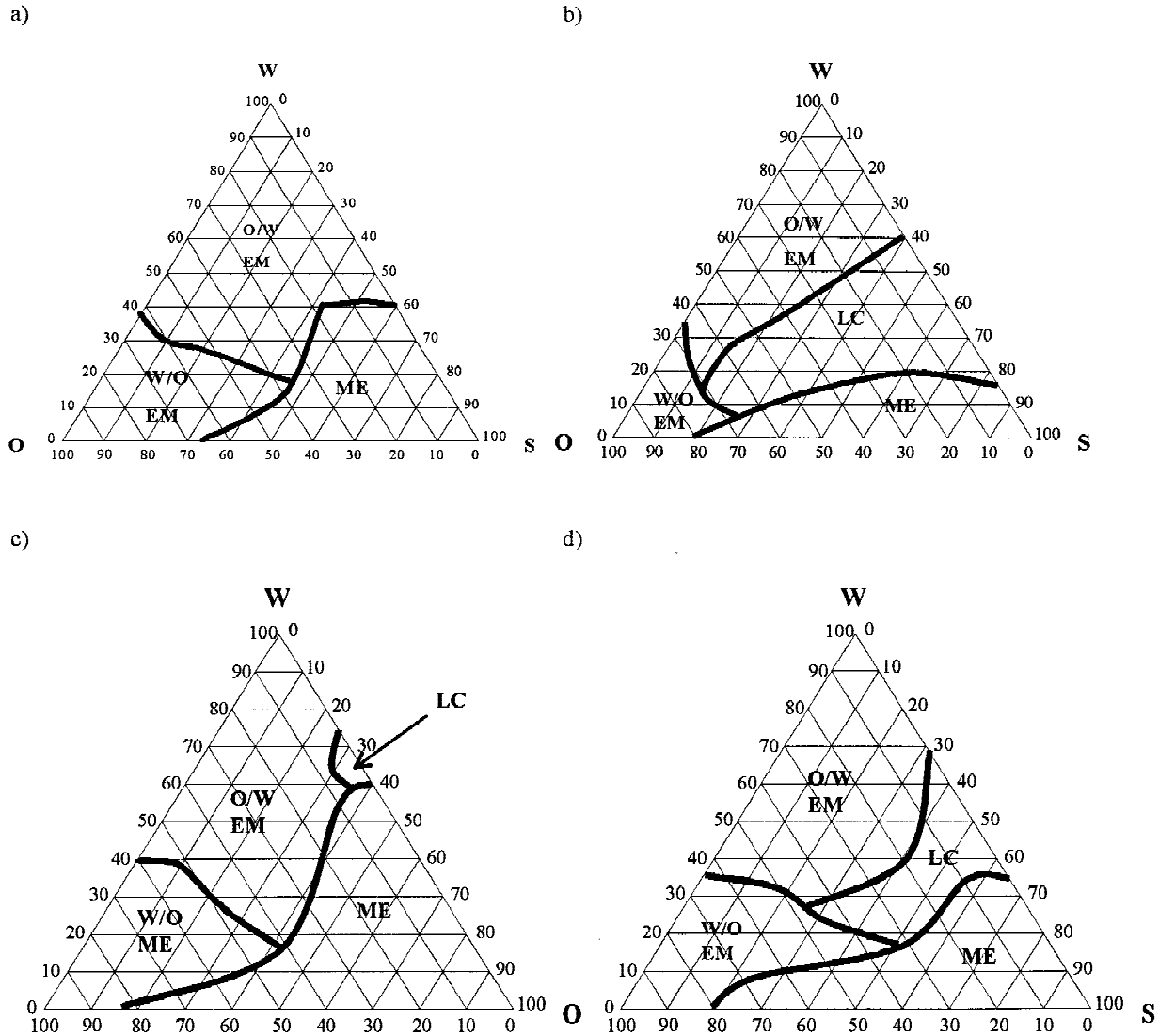

Fig. 2. GNN-predicted phase triangles for four surfactant/cosurfactant systems from the validation set with (a) n-propanol, (b) n-octanol, (c) 1,2-pentandiol, and (d) 1,2-octandiol as cosurfactants.

dicted ME region was significantly greater than that determined experimentally. Samples formulated within this region were found to behave as Winsor II ME were a water-in-oil ME exist in equilibrium with excess of water. Overall, with respect to the ME region, the network's prediction had reasonable success. The network successfully regenerated the LC region, indicating clearly that, unlike with short-chain alcohols (1-propanol), incorporation of a medium-chain alcohol failed to disturb the long-range order of the surfactant blend with subsequent emergence of a multiphase LC region.

For the 1,2-pentanediol phase triangle the network correctly classified most of the data points in the ME, W/O EM, and O/W EM regions with mistakes confined to the ME–LC boundary. The ME region was predicted to be balanced, with an almost equal preference of the surfactant blend for both the oily and aqueous phases. These predictions were in good agreement with the experimental findings. Predictions for the 1,2-octanediol system were encouraging, with the network successfully generating the four regions. The predicted ME region was shifted toward the oil apex of the phase triangle, suggesting that a cosurfactant with eight carbon atoms would be too lipophilic to alter the surfactant monolayer properties, but may instead decrease the lipophilicity of the oily phase. The experimentally produced ME region was also shifted toward the oil apex, but the ME had greater tendency to solubilize water.

CONCLUSION

This study evaluates the influence of the cosurfactant nature on the phase behavior of five component systems and can be applied as a preliminary evaluation of cosurfactants for pharmaceutically acceptable drug delivery systems to minimize experimental effort.

The used GNN has been proven to represent a statistically acceptable model to predict phase behavior depending on the content of separate components in the system and the cosurfactant nature. The training process required 171 data points form each pseudoternary phase diagram and 11 phys-

icochemical parameters related to the cosurfactant structure that can be easily calculated. The model was evaluated and the results have shown a fairly good degree of reliability. It can provide useful tools for the development of ME-based drug delivery systems.

REFERENCES

1. S. Cavalli. The effect of alcohols with different structures on the formation of warm o/w microemulsions. *J. Dispersion Sci. Tech*. **17**:717–734 (1996).
2. M. H. G. M. Penders and R. Strey. Phase behavior of the quaternary system H_2O/n-octane/C(8)E(5)/n-octanol—Role of the alcohol in microemulsions. *J. Phys. Chem*. **99**:10313–10318 (1995).
3. B. Faulhaber and M. Kahlweit. Preparing nontoxic microemulsions with alkyl monoglucosides and the role of alkanediols as cosolvents. *Langmuir* **12**:861–862 (1996).
4. G. J. T. Tiddy. Surfactant-water liquid crystal phases. *Physics Rep*. **57**:1–46 (1980).
5. P. Tyle. Liquid crystals and their application. In M. Rosoff (ed.), *Drug Delivery in Controlled Release of Drugs*, VCH Publishers, New York, 1988 pp. 125–162.
6. M. J. Schwuger, K. Stickdorn, and R. Schomacker. Microemulsions in technical processes. *Chem. Rev*. **95**:849–864 (1995).
7. R. Alany, T. Rades, S. Agatonovic-Kustrin , N. M. Davies, and I. G. Tucker. Effects of alcohols and diols on the phase behaviour of quaternary systems. *Int. J. Pharm.* **196**:141–145 (2000).
8. M. Egmont-Petersen, J. L. Talmon, and A. Hasman. Assessing the importance of features for multi-layer perceptrons. *Neural Networks* **11**:623–635 (1998).
9. G. Bologna and C. Pellegrini. Three medical examples in neural network rule extraction. *Phys. Med*. **13**:183–187 (1997).
10. T. F. Rathbun, S. K. Rogers, and M. P. DeSimio. MLP iterative construction algorithm. *Neurocomputing* **17**:195–216 (1997).
11. D. E. Goldberg. *Genetic Algorithms*, Addison Wesley, Reading, MA, 1989.
12. P. Willet. Genetic algorithms in molecular recognition and design. *Trends Biotechnol*. **13**:516–521 (1995).
13. S. S. So, and M. Karplus. Evolutionary optimization in quantitative structure-activity relationship: An application of genetic neural networks. *J. Med. Chem*. **39**:1521–1530 (1996).
14. S. S. So and M. Karplus. Three-dimensional quantitative structure-activity relationship from molecular similarity matrices and genetic neural networks. *J. Med. Chem*. **40**:4347–4359 (1997).
15. W. Lyman, W. Reehl, and D. Rosenblatt. *Handbook of Chemical Property Estimation*, American Chemical Society, Washington, DC, 1990.
16. R. W. Taft. *Steric Effects in Organic Chemistry*, John Wiley, New York, 1956.
17. J. Devillers, D. Domine, C. Guillon, and W. Karcher. Simulating lipophilicity of organic molecules with a back-propagation neural network. *J. Pharm. Sci.* **87**:1086–1090 (1998).
18. A. K. Saxena. Physicochemical significance of topological parameters—Molecular connectivity index and information. Content 2: Correlation studies with molecular refractivity and liphylicity. *Quant. Struct. Act. Rel*. **14**:142–148 (1995).
19. V. N. Viswandhan, A. K. Ghose, G. R. Revankar, and R.K. Robins. An estimation of the atomic contribution to octanol-water partition-coefficient and molar refractivity from fundamental atomic and structural properties—Its use in computer aided drug design. *Math. Comput. Model.* **14**:505–510 (1990).
20. D. L. Chester. *Why Two Hidden Layers are Better than One*, Lawrence Erlbaum, Mahwah, NJ, 1990 pp. 265–268.
21. H. T. Davis, Factors determining emulsion type–hydrophile-lipophile balance and beyond. *Colloid Surface A*. **91**:9–24 (1994).
22. C. Hansch, J. E. Quinlan, and G. L. Lawrence. The linear free-energy relationship between partition coefficient and the aqueous solubility of organic liquids. *J. Org. Chem*. **33**:347–50 (1968).
23. K. H. Kim, Effects with descriptors directly from 3D-structures using a comparative molecular-field analysis (COMFA) approach. *Quant. Struct. Act. Rel.* **11**:453–460 (1992).
24. K. Shinoda, M. Araki, A. Sadaghiani, A. Khan, and B. Lindman. Lecithin-based microemulsions. Phase behaviour and microstructure. *J. Phys. Chem.* **95**:989–993 (1991).