Application of QSPR to Mixtures

Subhash Ajmani,[†] Stephen C. Rogers,[‡] Mark H. Barley,[‡] and David J. Livingstone*[,†,§]

Centre for Molecular Design, Institute of Biomedical and Biomolecular Science, University of Portsmouth, King Henry 1 Street, Portsmouth PO1 2DY, U.K., ICI Strategic Technology Group, Wilton Centre, Wilton, Redcar TS10 4RF, U.K., and ChemQuest, Delamere House, 1 Royal Crescent, Sandown, Isle of Wight PO36 8LZ U.K.

Received December 29, 2005

In this paper we report an attempt to apply the QSPR approach for the analysis of data for mixtures. This is an extension of the conventional QSPR approach to the analysis of data for single molecules. The QSPR methodology was applied to a data set of experimental measured density of binary liquid mixtures compiled from the literature. The present study is aimed to develop models to predict the “delta” value of a mixture i.e., deviation of the experimental mixture density (MED) from the ideal, mole-weighted calculated mixture density (MCD). The QSPR was investigated in two perspectives (QMD-I and QMD-II) with respect to the creation of training and test sets. The study resulted in significant ensemble neural network and k-nearest neighbor models having statistical parameters r^2, q^2_{10cv} greater than 0.9, and pred_r^2 greater than 0.75. The developed models can be used to predict the delta and hence the density of a new mixture. The QSPR analysis shows the importance of hydrogen bond, polar, shape, and thermodynamic descriptors in determining mixture density, thus aiding in the understanding of molecular interactions important in molecular packing in the mixtures.

1. INTRODUCTION

In principle, the techniques of computer-aided drug design (CADD) such as molecular modeling and quantitative structure−activity relationships (QSAR) can be applied to the optimization of “performance” or “effect” chemicals such as flavors, fragrances, additives, catalysts, and so on. These applications can be grouped under the general heading of materials science and the fact that there is increasing use of CADD in this area can be seen in that recently a whole issue of the journal *QSAR and Combinatorial Science* was devoted to this topic (*QSAR Comb. Sci.* **2005**, *24*(1)). Where the chemical to be optimized is used essentially on its own or as the only “active” ingredient of a formulation, then the well-known CADD techniques may be used for the optimization. This circumstance, though, in materials science applications may be unusual, and it is more likely that the subject of optimization will be some form of more or less complicated formulation, in other words a mixture.

The application of CADD techniques to mixtures is, in principle, perfectly possible, but one of the practical problems that arise in their application is how to characterize a mixture. Considering the simplest case of a mixture, two components A and B in various mole ratios, then how should we compute and tabulate a set of descriptors for the mixture? One option is to use two values of each descriptor in the set, one for component A and one for B, and weight their contribution in some way according to their mole fraction. The problem with this approach, of course, is that at a stroke it doubles the number of descriptors to be considered for model building. With the huge choice of calculated parameters available today (Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Mannheim, 2000.) this can be a considerable disadvantage. Another option, as described by Sheridan,[1] is a centroid approximation of a mixture as the descriptor average of all the molecules. This method of characterizing mixtures was investigated by simulations using compounds from the MDL Drug Data Report where it was demonstrated that similarity searches were feasible. The advantage of this approach is that each descriptor is only considered once. The option that is considered in this report is to take a mole weighted average for each descriptor in the set.

To examine the feasibility of this approach to the construction of Quantitative Structure−Property Relationships (QSPR) models for mixtures, a set of measured density data for binary liquid mixtures has been compiled from the literature. The advantage of using liquid mixture density data is that they can be measured with high precision, are available for a wide range of chemical types, and represent a very simple property of a mixture which should be amenable to QSPR model construction.

The literature reveals an extensive interest in the densities and excess volumes of liquids and their mixtures which are used in a multitude of chemical processes. Accurate density values are essential for the design of chemical plants involved in the mixing, reacting, and separation of liquid products. The data measured reflect interactions between the molecules of the mixtures studied and are important from the theoretical viewpoint to understand the theories of the liquid state. As a result a lot of experimental data on the density of liquids and their mixtures have been collected.

* Corresponding author phone/fax: +44-1983-401793; e-mail: davel@chemquest.uk.com.

[†] University of Portsmouth.

[‡] ICI Strategic Technology Group.

[§] ChemQuest.

10.1021/ci050559o CCC: $33.50
Published on Web 07/18/2006

2. METHODS

2.1. Data Set. The data set consists of binary mixtures of 140 chemically diverse components viz. alcohols, esters, alkanes, ethers, ketones, amines, etc. The experimental density data of 271 binary mixtures was collected from the literature[2−78] and are reported in Table 1. The densities of 271 mixtures were reported at various mole fractions resulting in a total of 4679 mixture data points. The mole fractions of the two components in various mixtures were not the same i.e., mixtures have experimental density reported at different combinations (pentan-3-one 0.5203 and 1,1,2,2-tetrachloroethane 0.4797) and (acetonitrile 0.5256 and trichloromethane 0.4744), and number of combinations for a given mixture at various mole ratios as 10, 12, 15, 18, 22, etc. Density was measured at 25 °C with an accuracy of 10^{-5} g cm^{-3}.

The aim of the present study was to develop models to predict the deviation of the experimental mixture density (MED) from ideality. The ideal mixture density (referred to here as the calculated mixture density—MCD) was obtained by adding the product of the experimental density value of each component with its corresponding mole fraction in the mixture

$$\mathrm{MCD} = \mathrm{R1} \times \rho 1 + \mathrm{R2} \times \rho 2$$

where R1 and R2 are the mole fractions of the first and the second components in the mixture, and $\rho 1$ and $\rho 2$ are the experimental densities of the first and second components in the mixture.

Thus the dependent variable used in this study is “delta”, which is defined as

$$\mathrm{delta} = \mathrm{MED} - \mathrm{MCD}\ (\mathrm{g\ cm^{-3}}) = \text{deviation from ideality}$$

The delta values range from −0.1142 to 0.2287 g cm^{-3}. The data set also includes pure single components where delta = 0.

2.2. Descriptor Calculation. All 140 molecules were sketched and energy minimized using the Universal Force Field and Gasteiger charges with Cerius^2 software.[79] Various 2D and 3D molecular descriptors (∼1500) for each of the 140 components were calculated using Cerius^2 and E-Dragon 1.0 software.[80]

The main problem here is to decide how to calculate descriptors for mixtures. One method, which is used in the present work, is to simply calculate these as mole weighted average descriptors using the descriptor value and mole fraction of each component as follows

$$\mathrm{MD} = \mathrm{R1} \times \mathrm{D1} + \mathrm{R2} \times \mathrm{D2}$$

where MD is the mixture descriptor, R1 and R2 are the mole fractions of the first and second components in the mixture, and D1 and D2 are the descriptors of the first and second components.

2.3. Creation of the Training and the Test Sets. In the present study the data set was divided into training and test sets in two ways.

2.3.1. QMD-I. Where the same mixtures at different combinations of mole fractions were present in both training and test sets. The data set was divided into approximately 65% training and 35% test sets. The optimal training and test sets were generated using rational selection i.e., the sphere exclusion (SE) method.[81,92] This method allows the construction of a uniformly distributed subset of compounds as training and test sets from the whole data set. Random selection (RS) was also used for creating training and test sets to see the comparative effect of this method on the QSPR model.

2.3.2. QMD-II. Some mixtures were removed entirely from the training set, and these were then used to test the models developed. The total of 271 mixtures were divided into a training set of 232 and a test set of 39 mixtures by manual selection.

2.4. k-Nearest Neighbor (kNN) Method. The kNN methodology is a simple distance learning approach to classify an unknown member according to the classification of a majority of its k nearest neighbors in the training set. The nearness is measured by an appropriate distance metric (e.g., Euclidean distance using mixture descriptors). The standard kNN method involves the following steps:[82,83] (1) calculate distances between an unknown object (u) and all the objects in the training set; (2) select k objects from the training set closest to object u, according to the calculated distances; and (3) classify object u with the group to which the majority of the k objects belongs. The k value is optimized through the classification of a test set of samples or by internal validation. In this work the standard kNN method was modified so that the descriptors (or the principal components derived from the descriptors) and the optimal k values were chosen using a simulated annealing variable selection procedure as described below.

2.5. kNN-QSPR with Simulated Annealing. Simulated annealing (SA) is the simulation of a physical process, ‘annealing’, which involves heating the system to a high temperature and then gradually cooling it down to a preset temperature (e.g., room temperature).[90] During simulated annealing, the system samples all possible configurations with a probability determined by a Boltzmann distribution so that, as the “temperature” parameter in the Boltzmann distribution is reduced, the system is highly likely to settle into the lowest energy state (global minimum) rather than get stuck in a local minimum.

The kNN-QSPR method employs the kNN classification principle combined with a SA variable selection procedure. For each predefined number of variables (Vn) it seeks to optimize the following: (i) the number of nearest neighbors (k) used to estimate the delta of each mixture and (ii) selection of variables from the original pool of all mixture descriptors that are used to calculate similarities between mixtures.

Specifically, the kNN-QSPR procedure involves the following steps: (1) Select a subset of Vn descriptors randomly as a trial model. Vn is usually set to a value between 4 and 10. (2) For each trial model, compute r^2 (leave-one-out q^2) and q^2_{10cv} (leave-out-10% statistic) for a given range of k values (usually 2−5), by an internal-validation procedure as described below. (3) Repeat steps 1 and 2 until the maximum q^2_{10cv} for a given number of Vn is achieved. This optimization process is driven by generalized simulation using q^2_{10cv} as the objective function.

The implementation of kNN-QSPR as described below is similar to that described in refs 83 and 93 except that, of

Table 1: Binary Mixtures Used in QSPR Analysis[a]

component A	component B	ref	component A	component B	ref
dimethyl carbonate	cyclopentane	2	diethyl carbonate	methanol	27
diethyl carbonate	cyclopentane	2	diethyl carbonate	toluene	27
acetone	dimethyl carbonate	3	methanol	toluene	27
butanone	dimethyl carbonate[b]	3	water	n,n-dimethyl ethanolamine	28
pentanone	dimethyl carbonate	3	water	n,n-diethyl ethanolamine	28
ethanol	n-methyl piperazine[b]	4	cyclohexanone	benzene	29
ethyl acetate	pentanol	5	cyclohexane	hexane	30
ethanol	2-methyl propanol	5	acetone	water	31
methanol	ethyl butyrate	6	methanol	water	31
vinyl acetate	ethyl butyrate	6	acetone	butanol	31
methanol	ethyl propionate	7	methanol	butanol	31
vinyl acetate	ethyl propionate	7	ethyl formate	2,2,2-trifluoroethanol	32
methanol	n-methyl piperazine	8	ethyl formate	ethanol	32
propanol	n-methyl piperazine	8	ethyl formate	benzene	32
butanol	n-methyl piperazine	8	ethyl acetate	benzene	33
isobutanol	n-methyl piperazine	8	ethyl acetate	ethanol	33
phenylethylamine	toluene	9	ethyl acetate	2,2,2-trifluoroethanol	33
propionic acid	water	10	methanol	pentane	34
acetone	water[b]	10	methanol	hexane[b]	34
acetone	propionic acid	10	methanol	heptane	34
ethanol	butanone	11	methanol	octane	34
ethanol	benzene	11	ethanol	pentane	34
butanone	benzene	11	ethanol	hexane	34
methanol	propyl acetate	12	ethanol	heptane[b]	34
methanol	isopropyl acetate	12	ethanol	octane	34
vinyl acetate	propyl acetate[b]	12	propanol	pentane	34
vinyl acetate	isopropyl acetate	12	propanol	hexane	34
methanol	vinyl propionate	13	propanol	heptane[b]	34
vinyl acetate	vinyl propionate	13	propanol	octane	34
acetonitrile	methanol	14	methoxy ethanol	cyclohexane	35
acetonitrile	ethanol	14	n,n-dimethyl ethanolamine	1,4-dioxane	36
acetonitrile	propanol[b]	14	n,n-dimethyl ethanolamine	dimethyl formamide	36
acetonitrile	butanol	14	n,n-dimethyl ethanolamine	dimethyl acetamide[b]	36
2-methyl 2-propanol	toluene	15	n,n-dimethyl ethanolamine	dimethyl sulfoxide	36
2-methyl 2-propanol	methyl cyclohexane	15	acetonitrile	heptanol[b]	37
toluene	methyl cyclohexane	15	acetonitrile	octanol	37
2-methyl 2-propanol	isooctane	15	pentan-3-one	1,2-dichloroethane	38
diethylamine	water	16	pentan-3-one	1,3-dichloropropane	38
nitromethane	methanol	17	pentan-3-one	1,4-dichlorobutane	38
propylene carbonate	methanol	17	pentan-3-one	trichloromethane	38
dimethyl carbonate	dodecane	18	pentan-3-one	1,1,1-trichloroethane	38
diethyl carbonate	dodecane	18	pentan-3-one	1,1,2,2-tetrachloroethane	38
methyl methacrylate	benzene[b]	19	5-chloropentan-2-one	hexane	39
methyl methacrylate	1,4-xylene[b]	19	5-chloropentan-2-one	toluene	39
methyl methacrylate	diethyl ether	19	5-chloropentan-2-one	ethylbenzene	39
methyl methacrylate	dibutyl ether	19	benzene	2-methyl 2-butanol	40
methyl methacrylate	toluene	19	cyclohexane	2-methyl-2-butanol	40
methyl methacrylate	cyclohexane[b]	19	2-ethoxyethanol	octane	41
methyl methacrylate	diisopropyl ether	19	2-ethoxyethanol	cyclohexane	41
diethyl carbonate	hexane	20	2-ethoxyethanol	benzene	41
diethyl carbonate	heptane[b]	20	diethyl ether	methanol	42
diethyl carbonate	octane	20	diethyl ether	ethanol	42
diethyl carbonate	cyclohexane	20	diethyl ether	hexane	42
HFMP	hexane	21	diethyl ether	heptane[b]	42
HFMP	benzene	21	tetraglyme	dichloromethane	43
HFMP	ethanol[b]	21	tetraglyme	trichloromethane	43
HFMP	propanol	21	tetraglyme	tetrachloromethane	43
HFMP	butanone	21	2-methoxyethanol	dimethyl carbonate	44
TFTFE	hexane	21	2-methoxyethanol	diethyl carbonate	44
TFTFE	benzene	21	2-methoxyethanol	propylene carbonate	44
TFTFE	ethanol	21	DEGMME	dimethyl carbonate	44
TFTFE	propanol	21	DEGMME	diethyl carbonate[b]	44
TFTFE	butanone	21	DEGMME	propylene carbonate	44
1,2-dichloroethane	2-methoxyethanol	22	TEGMME	dimethyl carbonate	44
1,2-dichloroethane	1,2-dimethoxyethane	22	TEGMME	diethyl carbonate	44
2-amino-2-methyl propanol	water	23	TEGMME	propylene carbonate	44
butyl diethanolamine	water	23	2-ethoxyethanol	1,4-dioxane	45
propyl ethanolamine	water	23	2-ethoxyethanol	1,2-dimethoxyethane	45
monoethyl ethanolamine	water	23	2-methyl-2-butanol	pentanol	46
diethyl ethanolamine	water[b]	24	propyl ethanoate	ethanol	47
monomethyl ethanolamine	water[b]	25	propyl ethanoate	propanol[b]	47
dimethyl ethanolamine	water	25	propyl ethanoate	butanol	47
morpholine	water	26	pentanoic anhydride	hexane	48
methyl morpholine	water	26	pentanoic anhydride	octane[b]	48
dimethyl carbonate	toluene	27	pentanoic anhydride	dodecane	48

Table 1. (Continued)

component A	component B	ref	component A	component B	ref
hexanoic anhydride	hexane	48	acetonitrile	dichloromethane	62
hexanoic anhydride	octane	48	acetonitrile	trichloromethane^b	62
hexanoic anhydride	dodecane	48	acetonitrile	tetrachloromethane	62
2-isobutoxy ethanol	water	49	diethylamine	aceonitrile	63
1,3-dioxane	2,2,4-trimethylpentane	50	S_butylamine	aceonitrile	63
1,4-dioxane	2,2,4-trimethylpentane	50	DEGDEE	dichloromethane	64
1,3-dioxane	benzene	50	DEGDEE	trichloromethane^b	64
1,4-dioxane	benzene	50	DEGDEE	tetrachloromethane	64
1,3-dioxane	tetrachloroethene	50	1,4-dioxane	butyl acrylate	65
1,4-dioxane	tetrachloroethene	50	1,4-dioxane	ethyl acrylate	65
methyl methacrylate	propan-2-ol	51	1,4-dioxane	methyl methacrylate	65
methyl methacrylate	2-methyl propanol	51	1,4-dioxane	styrene	65
methyl methacrylate	butan-2-ol	51	benzene	butyl acrylate	66
methyl methacrylate	2-methyl 2-propanol	51	benzene	ethyl acrylate	66
pentane	1-chloropropane	52	benzene	methyl methacrylate	66
pentane	1-chlorobutane^b	52	benzene	styrene^b	66
pentane	1-chloropentane	52	1,2-dimethylbenzene	butyl acrylate	67
pentane	1-chlorohexane	52	1,2-dimethylbenzene	ethyl acrylate	67
hexane	1-chloropropane	52	1,2-dimethylbenzene	methyl methacrylate	67
hexane	1-chlorobutane	52	1,2-dimethylbenzene	styrene	67
hexane	1-chloropentane^b	52	benzene	isopropylbenzene	68
hexane	1-chlorohexane	52	benzene	1,2,4-trimethylbenzene	68
heptane	1-chloropropane	52	benzene	1,3,5-trimethylbenzene	68
heptane	1-chlorobutane	52	2-methoxyethanol	1,4-dioxane	69
heptane	1-chloropentane^b	52	1,2-dimethoxyethane	benzene	69
heptane	1-chlorohexane	52	propyl propanoate	1,2-dimethylbenzene	70
octane	1-chloropropane	52	propyl propanoate	1,3-dimethylbenzene^b	70
octane	1-chlorobutane	52	propyl propanoate	1,4-dimethylbenzene	70
octane	1-chloropentane	52	1,4-dimethylbenzene	butyl acrylate	71
octane	1-chlorohexane	52	1,4-dimethylbenzene	ethyl acrylate	71
1,2-ethanediol	propanol	53	1,4-dimethylbenzene	methyl methacrylate	71
1,2-ethanediol	butanol	53	1,4-dimethylbenzene	styrene	71
vinyl acetate	toluene	54	1,3-dioxolane	2-methylpropanol	72
vinyl acetate	ethylbenzene	54	1,3-dioxolane	2-methyl-2-propanol	72
vinyl acetate	*p*-xylene^b	54	1,4-dioxane	2-methylpropanol	72
vinyl acetate	isopropylbenzene^b	54	1,4-dioxane	2-methyl-2-propanol	72
vinyl acetate	mesitylene	54	toluene	propiophenone	73
vinyl acetate	butylbenzene	54	water	1,4-dioxane	74
vinyl acetate	isobutylbenzene	54	water	EGMME	74
vinyl acetate	*tert*-butylbenzene^b	54	water	TEGMME^b	74
1,2-propanediol	water	55	water	PEGMME	74
1,2-butanediol	water	55	water	EGDME	74
butanol	tetrahydrofuran	56	water	TEGDME	74
butan-2-ol	tetrahydrofuran	56	water	PEGDME	74
pentanol	cyclohexane	57	water	DEGDME^b	74
pentanol	benzene	57	mono ethanol amine	methanol	75
nonafluorobutyl methyl ether	ethanol	58	anisole	hexane	76
nonafluorobutyl methyl ether	methanol	58	anisole	heptane^b	76
methanol	vinyl acetate	59	anisole	octane	76
vinyl acetate	allyl acetate	59	anisole	nonane	76
ethyl acetate	benzene	60	anisole	decane^b	76
ethyl acetate	methylbenzene	60	anisole	dodecane	76
ethyl acetate	ethylbenzene^b	60	benzene	acetophenone	77
ethyl acetate	1,4-dimethylbenzene	60	toluene	acetophenone^b	77
ethyl acetate	1-methyethylbenzene	60	1,3-dimethylbenzene	acetophenone	77
ethyl acetate	1,3,5-trimethylbenzene	60	1,3,5-trimethylbenzene	acetophenone	77
ethyl acetate	1,1-dimethylethylbenzene	60	benzene	propiophenone	78
cyclohexane	butyl acrylate	61	toluene	propiophenone	78
cyclohexane	ethyl acrylate	61	ethylbenzene	propiophenone^b	78
cyclohexane	methyl methacrylate^b	61	butylbenzene	propiophenone	78
cyclohexane	styrene	61			

[a] Propylene carbonate = (4-methyl-1,3-dioxolan-2-one). Mesitylene = 1,3,5-trimethylbenzene. HFMP = 1,1,2,3,3,3-hexafluoromethoxypropane. TFTFE = 1,1,2,2-tetrafluoro-1-(2,2,2-trifuoroethoxy)ethane. EGDME = 1,2-dimethoxyethane, ethylene glycol dimethyl ether. TEGMME= triethyleneglycolmonomethyl ether,2-{2-(2-methoxyethoxy)ethoxy}ethanol. PEGMME = pentaethylene glycol monomethyl ether. TEGDME = triethylene glycol dimethyl ether. PEGDME = pentaethylene glycol dimethyl ether. DEGDME = diethylene glycol dimethyl ether. DEGMME = 2-(2-methoxyethoxy)ethanol,diethyleneglycolmonomethyl ether. *tert*-butylbenzene = 1,1-dimethylethylbenzene. Isopropylbenzene = 1-methylethylbenzene. DEGDEE = diethylene glycol diethyl ether. EGMME = 2-methoxyethanol. [b] Mixtures considered in test set for QMD-II.

course, it is not informative to calculate a leave-one-out q^2 (r^2) for a kNN model so we have computed a leave-out-10% statistic (q^2_{10cv})—see 2.8.1. The flowchart is shown in Figure 1. It may be summarized as the following steps: (1) Generate a trial solution to the underlying optimization problem; i.e., a QSPR model is built based on a random

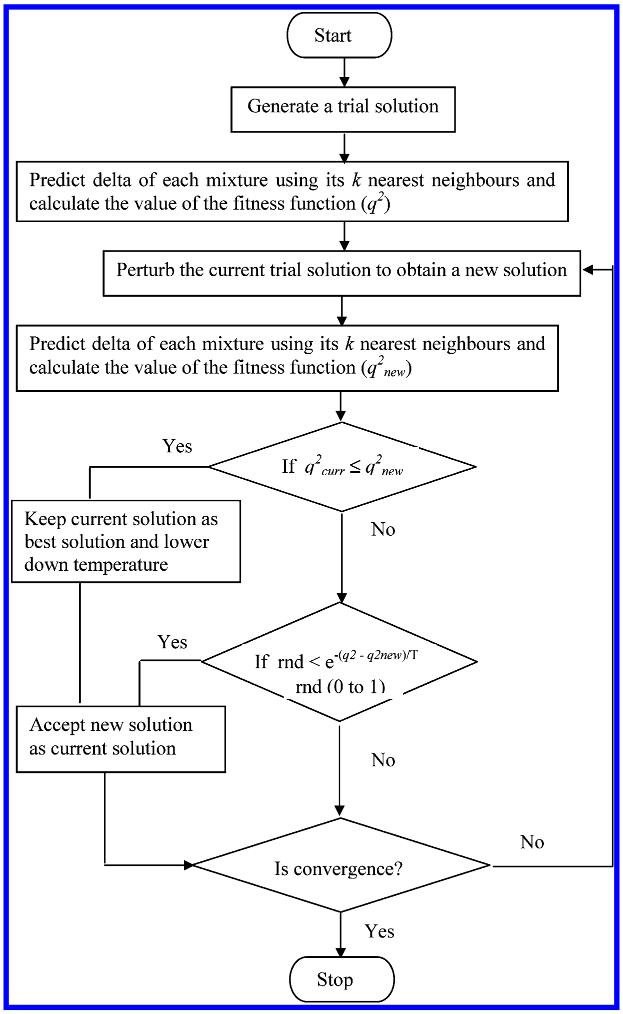
Start
Generate a trial solution
Predict delta of each mixture using its k nearest neighbours and calculate the value of the fitness function (q²)
Perturb the current trial solution to obtain a new solution
Predict delta of each mixture using its k nearest neighbours and calculate the value of the fitness function (q²new)
If q²curr ≤ q²new
Yes
No
Keep current solution as best solution and lower down temperature
If rnd < e^-(q2 - q2new)/T rnd (0 to 1)
Yes
No
Accept new solution as current solution
Is convergence?
No
Yes
Stop

Figure 1. Flowchart of kNN-QSPR.

selection of descriptors. (2) Calculate the value of the fitness function, which characterizes the quality of the trial solution to the underlying problem, i.e., the q^2 value for a QSPR model. (3) Perturb (i.e., slightly modify) the trial solution to obtain a new solution; i.e., change a fraction of the current trial solution descriptors to other randomly selected descriptors and build a new QSPR model for the new trial solution. (4) Calculate the new value of the fitness function (q^2_{new}) for the new trial solution. (5) Apply the optimization criteria: if $q^2_{\text{curr}} \leq q^2_{\text{new}}$ the new solution is accepted and used to replace the current trial solution; if $q^2_{\text{curr}} > q^2_{\text{new}}$, the new solution is accepted only if the following Metropolis criterion is satisfied; i.e., where rnd is a random number uniformly distributed between 0 and 1, and T is a parameter analogous to the temperature in Boltzmann distribution law.

$$\text{rnd} < e^{-(q^2\text{curr} - q^2\text{new})/T}$$

(6) Steps 3−5 are repeated until the termination condition is satisfied. The temperature-lowering scheme and convergence criteria used in this work have been adapted from Sun et al.[89] Thus, when a new solution is accepted or when a preset number of successive steps of generating trial solutions (20 steps) do not lead to a better result, the temperature is lowered by 10% (initial temperature is set to 1000 K). The calculations are terminated, when either the current temperature of simulations reaches 10^{-6} K or the ratio between the current temperature and the temperature corresponding to the best solution found equals 10^{-6}. The k value is optimized in the range of 2−5.

Using kNN-QSPR model the delta (property) of a mixture can be predicted using weighted average delta (eq 1) of the k most similar mixtures in the training set

$$\hat{y}_i = \sum w_i y_i \tag{1}$$

where y_i and $\hat{y}_i$ are the actual and predicted delta of the ith mixture respectively, and w_i are weights calculated using (eq 2).

$$w_i = \frac{\exp(-d_j)}{\sum_{j=1}^{k} \exp(-d_j)} \tag{2}$$

The similarities were evaluated as the inverse of Euclidean distances (d_j) between mixtures (eq 3) using only the subset of descriptors corresponding to the model

$$d_{i,j} = [\sum_{k=1}^{\text{Vn}} (\mathbf{X}_{i,k} - \mathbf{X}_{j,k})^2]^{1/2} \tag{3}$$

where **X** is the matrix of selected descriptors for the k-NN QSPR model.

2.6. Artificial Neural Networks. Artificial neural networks (NN) are a novel and powerful technique to build models that can effectively solve complex real world problems. NN are systems emulating the function of the brain where a very high number of information-processing neurons are interconnected in several ways to form various types of networks. Probably the most widely used architecture is a feed-forward multilayer neural network.[84] In this form of network the neurons are arranged in three layers (an input layer, one hidden layer, and an output layer). Each neuron in any layer is fully connected with the neurons of an adjacent layer, and there are no connections between neurons belonging to the same layer.

The input layer consists of n neurons (one for each descriptor used in the model) that serve as distribution points. The hidden layer of neurons computes the weighted sum of all of the inputs according to (eq 4) and then transforms it using a nonlinear activation function [85] (eq 5) before sending the result to the output neurons

$$\text{net} = \sum_{i=1}^{n} x_i w_i + \theta \tag{4}$$

where w_i (i = 1, n) represents the connection weights, and θ is called the bias.

The nonlinear activation function transforms the function *net* and provides the ability of the neural network to model nonlinear relationships. In this study, the tan sigmoid function given by eq 5 was used as the activation function.

$$y(\text{final output}) = f(\text{net}) = \frac{2}{1 + e^{-\text{net}}} - 1 \tag{5}$$

The networks were taught by giving them examples of input data and the corresponding results (target outputs). Through an iterative process, the connection weights are modified until the network gives the desired results for the training set of data. A back-propagation (BP) algorithm was used to minimize the error function in the present study.[86]

In this work hidden neurons had a tan sigmoidal transfer activation function, whereas a linear transfer function was used for the output neuron. The network was trained using the scaled conjugate gradient algorithm. The optimal network architecture (the number of units in the hidden layer) was determined by allowing the size of the hidden layer to vary from 4 to 30 hidden neurons. Starting from a neural network with 4 hidden neurons and adding one hidden neuron at a time, the trend of increase in training set r^2 was observed, and the optimal number was decided when adding another neuron made no significant increase in training set r^2.

2.7. Ensemble Neural Networks. Ensemble neural network formalism was used in the present work wherein 100 neural networks were built, and from these an ensemble or committee of 10 networks with the smallest cross-validation standard error in the training set was selected. The mean of the outputs from the ensemble networks was used to make predictions on the test data, while the variance of the outputs provides a measure of confidence in the predictions.

2.8. Model Validation. Validation was required to test both the internal stability and predictive ability of the QSPR models. Models were validated by internal and external validation procedures.

2.8.1. Internal Validation. Internal validation was carried out using leave 10% out (q^2_{10cv}) method. For q^2_{10cv}, the training set was divided into 10 equal groups. Each group was eliminated, and the delta of each mixture in that group was predicted by using the remaining nine groups. The q^2_{10cv} was calculated using eq 6 which describes the internal stability of a model. For kNN models the r^2 (leave-one-out q^2) was also computed using eq 6 by calculating a delta value for each mixture in the training set

$$q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{\text{mean}})^2} \qquad (6)$$

where y_i and y_i are the actual and predicted deltas of the ith mixture in the training set, respectively, and y_{mean} is the average delta of all mixtures in the training set.

The goodness of the correlation was also tested by standard error of estimate (SEE) computed by the following formula

$$\text{SEE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (7)$$

where n is the number of mixtures in the training set.

2.8.2. External Validation. For external validation, the delta of each mixture in the test set was predicted using the model developed by the training set. The pred_r^2 value is calculated as follows

$$\text{pred\_}r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - y_{\text{mean}})^2} \qquad (8)$$

where y_i and $\hat{y}$ are the actual and predicted deltas of the ith mixture in the test set, respectively, and y_{mean} is the average delta of all mixtures in the training set. Both summations are over all mixtures in the test set. Thus the pred_r^2 value is indicative of the predictive power of the current model for the external test set.

2.8.3. Randomization Test. To evaluate the statistical significance of the QSPR model for an actual data set, one-tail hypothesis testing was used.[83,87] The robustness of the models for training sets was examined by comparing these models to those derived for random data sets. Random sets were generated by rearranging the delta of mixtures in the training set. The significance of the models hence obtained was derived based on a calculated Zscore.[83,87] A Zscore value is calculated by formula Zscore = $(h - \mu)/\sigma$ where h is the q^2 value for the actual data set, μ is the average q^2, and σ is its standard deviation for random data sets. The calculated Zscore is compared with tabulated Zscore (critical) values at different levels of significance to determine the level (α) at which the hypothesis (H_0) should be rejected. For example, a Zscore value greater than 3.10 indicates that there is a probability (α) of less than 0.001 and that the QSPR model constructed for the real data set is random.

2.8.4. Evaluation of the QSPR Models. The QSPR models were evaluated using following statistical measures: n, number of observations (mixtures); k, number of nearest neighbors; Vn, number of descriptors; r^2, coefficient of determination; q^2_{10cv}, cross-validated r^2 (by leave 10% out); pred_r^2, r^2 for external test set; Zscore, Z score calculated by the randomization test; best_ran_q^2, highest q^2 value in the randomization test; best_ran_r^2, highest r^2 value in the randomization test; α, statistical significance parameter obtained by the randomization test; and SEE, standard error of estimate.

2.9. NN and kNN Method Implementation and Computation Environment. All work was carried out using an AMD Athlon 64 processor 3400+ with 1GB RAM. The kNN-QSPR method was implemented in a MATLAB script.[88] The Neural Network toolbox in MATLAB was used for building NN models.

3. RESULTS AND DISCUSSION

Models were developed to predict the delta i.e., deviation of the experimental mixture density (MED) from the “ideal” calculated mixture density (MCD). Figure 2 shows a plot of the relationship of MED versus MCD where it can be seen that this is, in general, linear but that there are both positive and negative deviations from ideality.

Before starting the QSPR analysis, descriptors with zero variance were removed from the data set, and the remaining descriptors were autoscaled. Here quantitative structure−property relationship was investigated in two perspectives with respect to creation of training and test sets i.e., (1) where the same mixtures at different combinations of mole fractions were present in both training and test set (QMD-I) and (2) where some mixtures were removed entirely (all combinations) from the training set and these were then used to test the developed models (QMD-II).

In QMD-I, the data set was divided into training and test sets using rational (sphere exclusion (SE) algorithm) and random selection (RS) methods. Principal component analysis was carried out using all the descriptors. The first 20

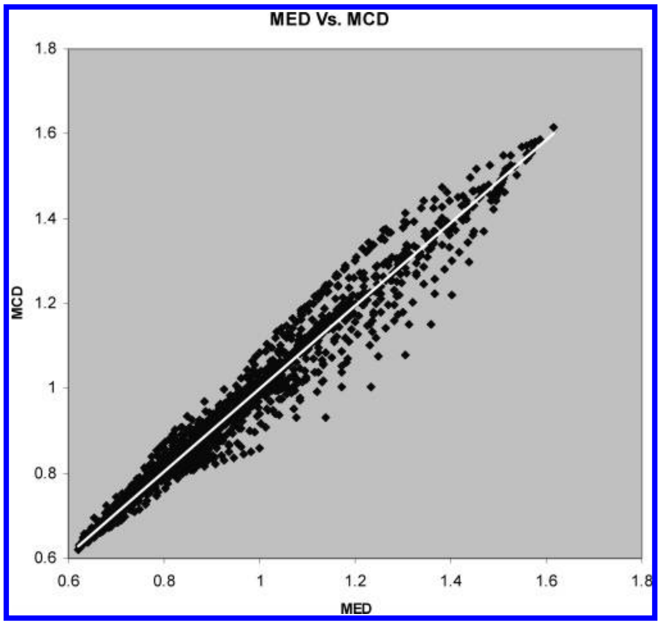

Figure 2. Plot of linear relationship of MED versus MCD.

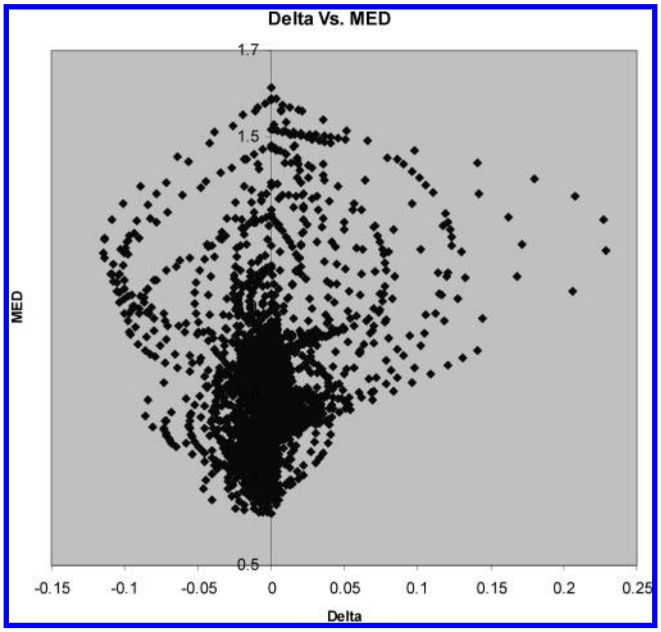

Figure 3. Plot of MED versus delta.

principal components explaining ∼88% variance of the data were used as independent variables for the SE method.

At first, various linear regression methods i.e., multiple regression, principal component regression, and partial least squares regression coupled with a simulated annealing (SA) variable selection procedure were applied to both the SE and RS selected training and test sets. These led to no statistically significant models (best $r^2 \sim 0.20$). Inspection of a plot of delta vs MED (Figure 3) reveals the nonlinearity of delta as a "response" variable, and thus it is not surprising that linear models give such a poor fit.

Considering this nonlinearity, it was decided to apply the k-nearest neighbor method. Initially, the kNN QSPR model was built using principal components ranging from the first 5−25 components in steps of 5. This led to statistically significant models (Table 2) with the first 20 principal components (out of 25 principal components) using both the RS and SE selected training and test sets. These results indicate that the statistical significance of the model is independent of the method of training and test set selection i.e., SE and RS. This is presumably a consequence of the fact that these are relatively large sets. Randomization test results in $\alpha < 0.001$ for r^2/q^2_{10cv} which further confirms statistical significance of all the models. The Zscore r^2/q^2_{10cv} value was calculated by generating 10 random sets, and this value (greater than 4 in all the models) suggested that the QSPR model was better than random with greater than 99.9% confidence.

Table 2: kNN QSPR Model Using 20 Principal Components, QMD-I

	SE	RS
k	2	2
training set	2964	2948
test set	1715	1731
r^2	0.957	0.956
q^2_{10cv}	0.952	0.952
pred_r^2	0.961	0.960
r^2 SEE	0.0051	0.0051
q^2_{10cv} SEE	0.0054	0.0053
pred_r^2 SEE	0.0048	0.0049
best_ran_ r^2	−0.423	−0.460
best_ran_q^2_{10cv}	−0.453	−0.454
Zscore_ r^2	32.249	48.543
Zscore_q^2_{10cv}	29.632	40.870
α_ r^2	<0.001	<0.001
α_q^2_{10cv}	<0.001	<0.001

To understand the mechanism of packing of molecules in mixtures, we have applied the kNN QSPR method using two different sets of descriptors calculated by Cerius2 and Dragon software to build individual QSPR models. The analysis resulted in the models revealing the descriptors important in explaining variation in delta. Table 3 reports models using 8 Cerius^2, 7 Dragon, and a combined set of 15 descriptors. All the models have r^2, q^2_{10cv}, and pred_r^2 greater than 0.9.

The following descriptors were found to play important roles in determining delta of the mixtures:

Cerius^2 Descriptors. LUMO: energy of lowest unoccupied molecular orbital calculated by the CNDO/2 method. This is important in governing molecular reactivity and frontier properties.

HBA: number of hydrogen bond acceptors in a molecule.

HBD: number of hydrogen bond donors in a molecule.

AtypeH47: number of hydrogens attached to sp^3, sp^2 carbon atoms in a molecule.

Superdelocalizability (Sr): an index of reactivity in aromatic hydrocarbons (AH), proposed by Fukui

$$\mathrm{Sr} = 2\sum_{j=1}^{m} \frac{{c_{jr}}^2}{e_j}$$

where Sr = superdelocalizability at position r, e_j = bonding energy coefficient in the jth MO (eigenvalue), c_{jr} = molecular orbital coefficient at position r in the HOMO, and m = index of the HOMO.

Molecular flexibility index (PHI): This is a descriptor based on structural properties that restrict a molecule from being "infinitely flexible", the model for which is an endless chain of C (sp^3) atoms. The structural features considered as preventing a molecule from attaining infinite flexibility are as follows: (a) fewer atoms, (b) the presence of rings,

Table 3: kNN QSPR Model Using Cerius2 and Dragon Descriptors, QMD-I

	descriptors					
	Cerius2 (8)		Dragon (7)		Cerius2 and Dragon (15)	
	SE	RS	SE	RS	SE	RS
k	2	2	2	2	2	2
training set	2964	2948	2964	2948	2964	2948
test set	1715	1731	1715	1731	1715	1731
r^2	0.936	0.924	0.923	0.911	0.954	0.951
q^2_{10cv}	0.930	0.918	0.919	0.909	0.950	0.945
pred_r^2	0.932	0.943	0.944	0.904	0.961	0.964
r^2 SEE	0.0062	0.0067	0.0068	0.0073	0.0052	0.0054
q^2_{10cv} SEE	0.0065	0.0069	0.0070	0.0073	0.0055	0.0057
pred_r^2 SEE	0.0063	0.0058	0.0057	0.0076	0.0048	0.0046
best_ran_ r^2	−0.466	−0.479	−0.432	−0.473	−0.470	−0.463
best_ran_q^2_{10cv}	−0.478	−0.485	−0.452	−0.466	−0.473	−0.469
Zscore_ r^2	43.743	42.131	42.379	45.752	33.071	39.073
Zscore_q^2_{10cv}	36.267	41.211	39.378	46.032	42.720	36.499
α_ r^2	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
α_q^2_{10cv}	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

(c) branching, and (d) the presence of atoms with covalent radii smaller than those of C (sp^3). These features are encoded in the index as follows

$$\mathrm{PHI} = {k_1}^{\alpha}{k_2}^{\alpha}/N$$

where N = number of vertices, and k_1^{α} and k_2^{α} are alpha modified Kier's first- and second-order shape indices, respectively.

Kier's third-order shape index (kappa3): These indices compare the molecule graph with "minimal" and "maximal" graphs, where the meaning of "minimal" and "maximal" depends on the order n. This is intended to capture different aspects of the molecular shape

$$\mathrm{kappa3} = (N-1)(N-3)^2/P^2 \quad \text{for } N \text{ odd}$$

$$\mathrm{kappa3} = (N-3)(N-2)^2/P^2 \quad \text{for } N \text{ even}$$

where N = number of vertices, and P = number of paths of length 3 in the graph.

Foct: This is the 1-octanol desolvation free energy derived from a hydration shell model developed by Hopfinger, where Foct is in kcal mol^{-1}. Foct is a physiochemical property associated with linear free energy models of a property.

Dragon Descriptors.[91] **EEig04d:** eigenvalue 04 from edge adjacent matrix weighted by dipole moments.

BEHv3: highest eigenvalue *n*. 3 of Burden matrix weighted by atomic van der Waals volumes.

Mor15m: 3D-MoRSE − signal 15 weighted by atomic masses.

H0e: H autocorrelation of lag 0 weighted by atomic Sanderson electronegativities.

GATS1e: Geary autocorrelation of lag 1 weighted by atomic Sanderson electronegativities.

E2e: second component accessibility directional WHIM index weighted by atomic Sanderson electronegativities.

Mor10e: 3D-MoRSE − signal 10 weighted by atomic Sanderson electronegativities.

The presence of hydrogen bond acceptor and donor descriptors in the QSPR model suggests the importance of inter- and/or intramolecular hydrogen bonding for liquid mixture density. Kier shape index and molecular flexibility index shows the role of molecular flexibility for two components to adjust to accommodate each other in a mixture. Desolvation free energy in octanol (Foct) and AlogP hydrogen atom type (AtypeH47) reveals the importance of thermodynamic equilibrium and the hydrophobic nature of molecules in determining the mixture density. The energy of the lowest unoccupied molecular orbital and superdelocalizibility indicate the role of electrophilic/nucleophilic interactions between the components in the mixture density. They may also represent interactions of a charge-transfer type.

Table 4: Ensemble NN QSPR Model Using 20 Principal Components, QMD-I

	SE	RS
NN architecture	20-20-1	20-20-1
training set	2964	2948
test set	1715	1731
r^2	0.988	0.989
q^2_{10cv}	0.988	0.987
pred_r^2	0.964	0.970
r^2 SEE	0.0026	0.0026
q^2_{10cv} SEE	0.0027	0.0027
pred_r^2 SEE	0.0045	0.0042
best_ran_r^2	0.268	0.266
best_ran_q^2_{10cv}	−0.315	−0.328
Zscore_r^2	48.494	32.689
Zscore_q^2_{10cv}	51.788	43.754
α_r^2	<0.001	<0.001
α_q^2_{10cv}	<0.001	<0.001

Although interpretation of Dragon descriptors is not straightforward, the presence of H0e, GATS1e, E2e, Mor10e (matrix weighted by atomic Sanderson electronegativities), and EEig04D (matrix weighted by dipole moment) descriptors in the QSPR model suggests the importance of polarity and dipole moment in molecular packing in a mixture. The presence of BEHv3 and Mor15m reveals the role of size, shape, and atomic weight in determining the mixture density.

It can be noted that all the descriptors found to be important in the QSPR models represent very basic physical/chemical properties of molecules which play a role in determining one of the main physicochemical properties (density) used to characterize mixtures.

The neural network approach was also used to build QSPR models. The ensemble NN models were built using the first

Table 5: Ensemble NN QSPR Model Using Cerius2 and Dragon Descriptors, QMD-I

	descriptors					
	Cerius2 (8)		Dragon (7)		Cerius2 and Dragon (15)	
	SE	RS	SE	RS	SE	RS
NN architecture	8-24-1	8-24-1	7-21-1	7-21-1	15-30-1	15-30-1
training set	2964	2948	2964	2948	2964	2948
test set	1715	1731	1715	1731	1715	1731
r^2	0.960	0.959	0.930	0.932	0.989	0.990
q^2_{10cv}	0.930	0.929	0.902	0.905	0.978	0.977
pred_r^2	0.933	0.925	0.910	0.898	0.977	0.974
r^2 SEE	0.0049	0.0049	0.0065	0.0064	0.0026	0.0024
q^2_{10cv} SEE	0.0065	0.0065	0.0077	0.0075	0.0036	0.0037
pred_r^2 SEE	0.0062	0.0067	0.0072	0.0078	0.0036	0.0039
best_ran_r^2	0.099	0.122	0.105	0.110	0.262	0.229
best_ran_q^2_{10cv}	−0.112	−0.121	−0.081	−0.088	−0.302	−0.281
Zscore_r^2	127.840	58.895	103.630	80.378	33.183	85.902
Zscore_q^2_{10cv}	63.051	139.650	61.143	49.200	73.968	39.004
α_r^2	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
α_q^2_{10cv}	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

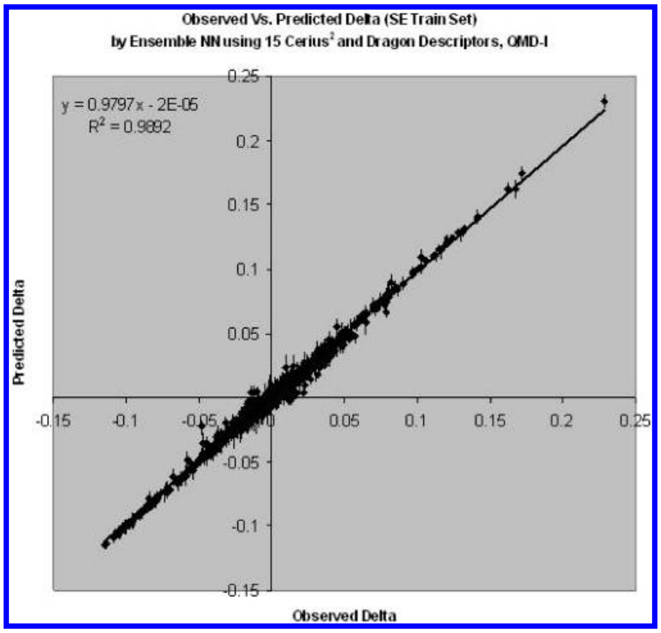

Figure 4. Plot of observed versus predicted delta in the training set (SE) by ensemble NN using 15 descriptors, in QMD-I.

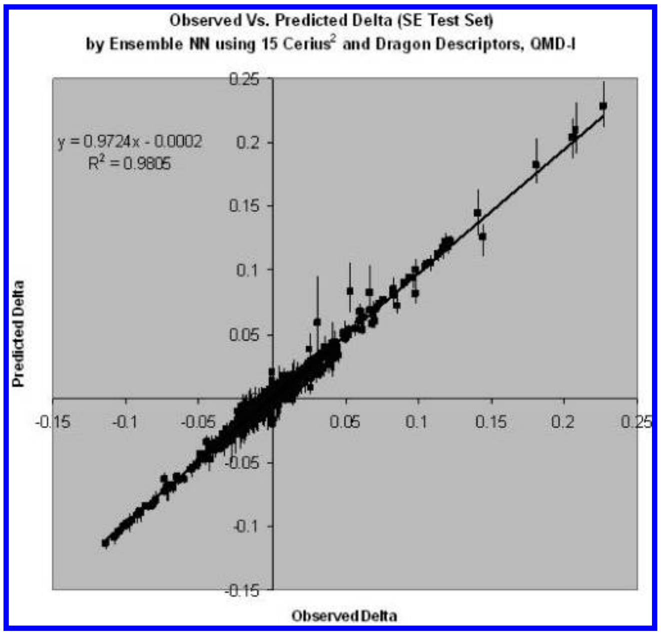

Figure 5. Plot of observed versus predicted delta in the test set (SE) by ensemble NN using 15 descriptors, in QMD-I.

20 principal components with 20 hidden neurons leading to statistically significant and better models as compared with the kNN models using both the RS and SE training and test sets (Table 4). The ensemble NN models were also built using both the descriptor sets found to be important in kNN models and resulted in statistically significant models. Table 5 reports ensemble NN models built using 8 Cerius^2, 7 Dragon, and 15 combined descriptors with 24, 21, and 30 hidden neurons, respectively. Randomization test results in $\alpha < 0.001$ for r^2 and q^2_{10cv} which further confirms the statistical significance of all the models. All the ensemble NN models have r^2, q^2_{10cv}, and pred_r^2 greater than 0.9.

Figures 4−6 shows the plot of observed versus predicted delta values for training, test, and cross-validated training set, respectively, for the best QSPR model i.e., the ensemble NN using both selected Cerius^2 and Dragon descriptors. Error bars in Figures 4−6 show the range of the maximum and minimum delta value predicted by the 10 NN models used for building ensemble NN model.

In QMD-II, the data set was divided into training (232 mixtures, 4018 observations) and test sets (39 mixtures, 661 observations) using manual selection in a way to consider a wide range of diverse mixtures for prediction in test set. This QSPR analysis was done especially to find out how well the QSPR model predicts property (delta) of new mixtures which have been kept entirely away while building the model. The same two sets of descriptors (Cerius^2 and Dragon) and two approaches i.e., ensemble NN and kNN as described above, were used to build the QSPR model.

Initially the ensemble neural network approach was used to build QSPR models as it resulted in better models as compared to k-NN models for QMD-I. Table 6 reports ensemble NN models built using 8 Cerius^2, 7 Dragon, and 15 combined descriptors with 24, 21, and 30 hidden neurons, respectively. All the ensemble NN models have r^2, q^2_{10cv} and pred_r^2 greater than 0.85. Randomization test results in $\alpha < 0.001$ for r^2 and q^2_{10cv} which further confirms the statistical significance of all the models. Figure 7 shows the plot of observed versus predicted delta values for training and test

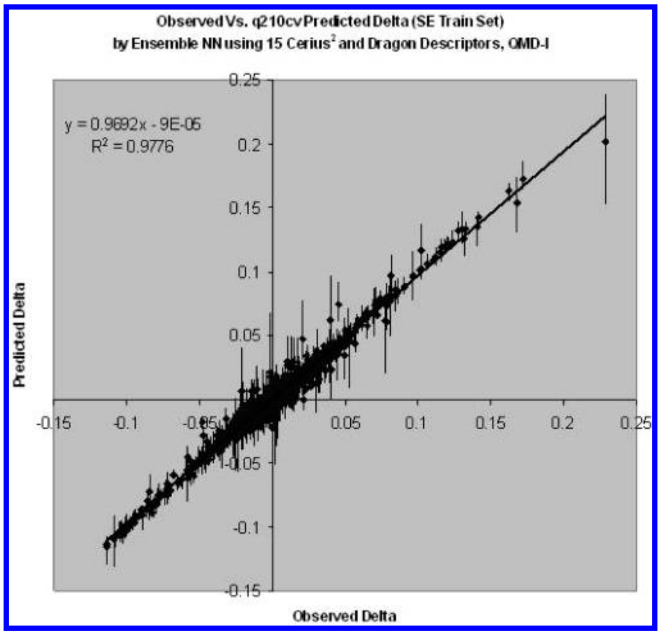

Figure 6. Plot of observed versus predicted delta (internal validation) in the training set (SE) by ensemble NN using 15 descriptors, in QMD-I.

Table 6: Ensemble NN QSPR Model Using Cerius2 and Dragon Descriptors, QMD-II

	descriptors		
	Cerius2 (8) SE	Dragon (7) SE	Cerius2 and Dragon (15) SE
NN architecture	8-24-1	7-21-1	15-30-1
training set	4018	4018	4018
test set	661	661	661
r^2	0.957	0.931	0.989
q^2_{10cv}	0.888	0.867	0.955
pred_r^2	0.954	0.863	0.976
r^2 SEE	0.0050	0.0064	0.0025
q^2_{10cv} SEE	0.0061	0.0071	0.0035
pred_r^2 SEE	0.0053	0.0091	0.0038
best_ran_r^2	0.117	0.084	0.236
Zscore_r^2	73.162	100.800	51.079
α_r^2	<0.001	<0.001	<0.001

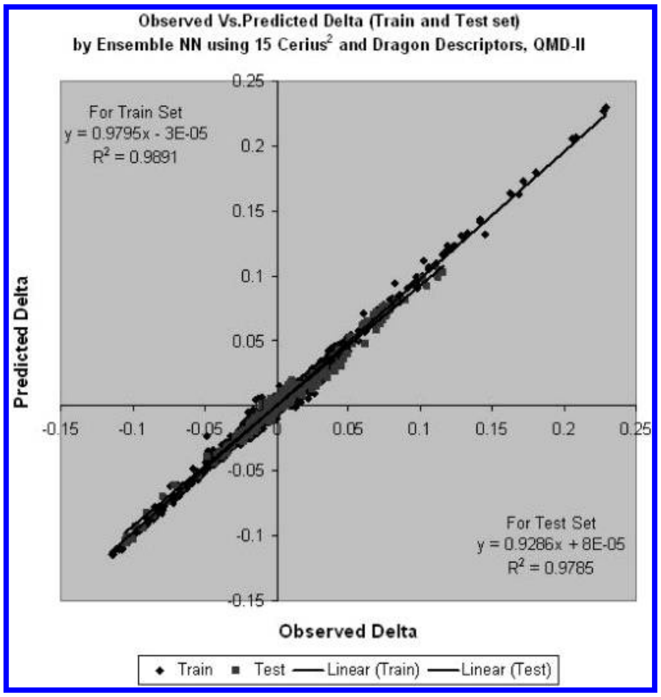

Figure 7. Plot of observed versus predicted delta in training and test sets by ensemble NN using 15 descriptors, in QMD-II.

Table 7: kNN QSPR Model Using Cerius2 and Dragon Descriptors, QMD-II

	descriptors		
	Cerius2 (8)	Dragon (7)	Cerius2 and Dragon (15)
k	2	2	2
training set	4064	4064	4064
test set	615	615	615
r^2	0.944	0.930	0.951
q^2_{10cv}	0.930	0.916	0.944
pred_r^2	0.754	0.762	0.865
r^2 SEE	0.0057	0.0064	0.0054
q^2_{10cv} SEE	0.0064	0.0070	0.0058
pred_r^2 SEE	0.0091	0.0090	0.0068
best_ran_ r^2	−0.434	−0.454	−0.426
Zscore_ r^2	45.094	85.368	43.267
α_ r^2	<0.001	<0.001	<0.001

sets for the ensemble NN using both selected Cerius^2 and Dragon descriptors.

In QMD-II also the kNN approach was used to build QSPR models. When a kNN-QSPR model was built using 39 mixtures in the test set, it was observed that a few combinations of 2 mixtures (acetonitrile and trichloromethane, HFMP and ethanol) were not predicted well. Therefore these two mixtures (all combinations) were considered in the training set while rebuilding the kNN-QSPR model. Table 7 reports the kNN-QSPR models (train 234 and test 37 mixtures) built using 8 Cerius^2, 7 Dragon, and 15 combined descriptors with $k = 2$. All the models have r^2, q^2_{10cv} greater than 0.9, and pred_r^2 greater than 0.75. Randomization test results in $\alpha < 0.001$ for r^2 and q^2_{10cv}.

In QMD-I, it can be seen from Tables 2−5 that both the kNN and ensemble NN models using 15 descriptors are comparable to the models using principal components with respect to all statistical parameters r^2, q^2_{10cv}, and pred_r^2. Ensemble NN models were found to be better than kNN models in both QMD-I and QMD-II. Also it can be noted (Tables 5 and 6) that there is not a significant difference in ensemble NN models for both QMD-I and QMD-II. However there is a significant difference in kNN models in both QMD-I and QMD-II w.r.t their predictive ability (pred_r^2) (Tables 3 and 7). These results suggest that the ensemble neural network approach is better as compared to kNN for dealing with mixture QSPR analysis.

Finally, the developed models can be used for predicting the delta and hence the density of a new mixture provided that the experimental density and mole fraction of both the components in the mixture are known. The obtained models use descriptors that pertain to the main structural features and interactions involved in packing of molecules in mixtures.

4. CONCLUSION

In the present work, we have shown that it is possible to apply the QSPR approach to the analysis of mixture data. In this case, the property modeled was a set of experimental density data for binary liquid mixtures compiled from the literature. One of the major problems in developing QSPR models for mixtures lies in the definition of descriptors for mixtures. In the approach adopted here, mixtures were characterized by mole weighted average descriptors. A disadvantage of this procedure is that there is no mechanistic basis for this approach, other than the obvious expectation that single molecular descriptors may be significant in the explanation of a simple property such as liquid mixture density. A further disadvantage is that this does not allow the examination of interaction effects. An advantage, however, is that this is a simple approach to the characterization of mixtures while minimizing the number of descriptors used. A further advantage is that it was possible to build good predictive models which can be used to predict the delta and hence the density of a new mixture.

Currently, we are studying various other mixing rules to derive new descriptors for mixtures for QSPR analysis, which may aid in understanding the molecular interactions important in mixtures. The findings of these studies will be published in our next paper.

REFERENCES AND NOTES

(1) Sheridan, R. P. The Centroid Approximation for Mixtures: Calculating Similarity and Deriving Structure−Activity Relationships. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1456−1469.

(2) Pereiro, A. B.; Rodriguez A.; Canosa, J.; Tojo, J. Density, Viscosity, and Speed of Sound of Dialkyl Carbonates with Cyclopentane and Methyl Cyclohexane at Several Temperatures. *J. Chem. Eng. Data* **2004**, *49*, 1392−1399.

(3) Pereiro, A. B.; Rodriguez A.; Canosa, J.; Tojo, J. Measurement of the Isobaric Vapor-Liquid Equilibria of Dimethyl Carbonate with Acetone, 2-Butanone, and 2-Pentanone at 101.3 kPa and Density and Speed of Sound at 298.15 K. *J. Chem. Eng. Data* **2005**, *50*, 481−486.

(4) Chen, S.; Fang, W.; Yao, J.; Zong, H. Density and Refractive Index at 298.15 K and Vapor-Liquid Equilibria at 101.3 kPa for Binary Mixtures of Ethanol + *N*-Methylpiperazine. *J. Chem. Eng. Data* **2001**, *46*, 596−600.

(5) Resa, J. M.; Gonzalez, C.; Goenaga, J. M.; Iglesias, M. Density, Refractive Index, and Speed of Sound at 298.15 K and Vapor-Liquid Equilibria at 101.3 kPa for Binary Mixtures of Ethyl Acetate + 1-Pentanol and Ethanol + 2-Methyl-1-propanol. *J. Chem. Eng. Data* **2004**, *49*, 804−808.

(6) Resa, J. M.; Gonzalez, C.; Landaluce, S. D.; Lanz, J. Density, Refractive Index, and Speed of Sound at 298.15 K, and Vapor-Liquid Equilibria at 101.3 kPa for Binary Mixtures of Methanol + Ethyl Butyrate and Vinyl Acetate + Ethyl Butyrate. *J. Chem. Eng. Data* **2002**, *47*, 1123−1127.

(7) Resa, J. M.; Gonzalez, C.; Landaluce, S. D.; Lanz, J. Density, Refractive Index, Speed of Sound, and Vapor-Liquid Equilibria for Binary Mixtures of Methanol + Ethyl Propionate and Vinyl Acetate + Ethyl Propionate. *J. Chem. Eng. Data* **2002**, *47*, 435−440.

(8) Chen, S.; Lei, Q.; Fang, W. Density and Refractive Index at 298.15 K and Vapor-Liquid Equilibria at 101.3 kPa for Four Binary Systems of Methanol, n-Propanol, n-Butanol, or Isobutanol with N-Methylpiperazine. *J. Chem. Eng. Data* **2002**, *47*, 811−815.

(9) Sporzynski A., Hofman, T.; Miskiewicz, A.; Strutynska, A.; Synoradzki, L. Vapor-Liquid Equilibrium and Density of the Binary System 1-Phenylethylamine + Toluene. *J. Chem. Eng. Data* **2005**, *50*, 33−35.

(10) Estrada-Baltazar, A.; Leon-Rodriguez, A. D.; Hall, K. R.; Ramos-Estrada, M.; Iglesias-Silva, G. A. Experimental Densities and Excess Volumes for Binary Mixtures Containing Propionic Acid, Acetone, and Water from 283.15 to 323.15 K at Atmospheric Pressure *J. Chem. Eng. Data* **2003**, *48*, 1425−1431.

(11) Grguric, I. R.; Serbanovic, S. P.; Kijevcanin, M. L.; Tasic, A. Z.; Djordjevic, B. D. Volumetric properties of the ternary system ethanol + 2-butanone + benzene by the van der Waals and Twu-Coon-Bluck-Tilton mixing rules: experimental data, correlation and prediction. *Thermochim. Acta* **2004**, *412*, 25−31.

(12) Resa, J. M.; Gonzalez, C.; Landaluce, S. D.; Lanz, J. Vapor-Liquid Equilibrium of Binary Mixtures Containing Methanol + Propyl Acetate, Methanol + Isopropyl Acetate, Vinyl Acetate + Propyl Acetate, and Vinyl Acetate + Isopropyl Acetate at 101.3 kPa. *J. Chem. Eng. Data* **2001**, *46*, 1338−1343.

(13) Resa, J. M.; Gonzalez, C.; Landaluce, S. D.; Goenaga J. M. Density, Refractive Index, Speed of Sound, and Vapor-Liquid Equilibria for Binary Mixtures of Methanol + Vinyl Propionate and Vinyl Acetate + Vinyl Propionate. Vapor Pressures of Vinyl Propionate. *J. Chem. Eng. Data* **2005**, *50*, 319−324.

(14) Torres, R. B.; Francesconi, A. Z.; Volpe, P. L. O. Experimental study and modelling using the ERAS-Model of the excess molar volume of acetonitrile-alkanol mixtures at different temperatures and atmospheric pressure. *Fluid Phase Equilib.* **2003**, *210*, 287−306.

(15) Pena, M. P.; Martinez-Soria, V.; Monton, J. B. Densities, refractive indices, and derived excess properties of the binary systems tert-butyl alcohol + toluene, + methylcyclohexane, and + isooctane and toluene + methylcyclohexane, and the ternary system tert-butyl alcohol + toluene + methylcyclohexane at 298.15 K. *Fluid Phase Equilib.* **1999**, *166*, 53−65.

(16) Lampreia, I. M. S.; Dias, F. A.; Mendonc, A. F. S. S. Volumetric study of (diethylamine + water) mixtures between (278.15 and 308.15) K. *J. Chem. Thermodyn.* **2004**, *36*, 993−999.

(17) Piekarski, H.; Pietrzak, A.; Waliszewski D. Heat capacities and volumes of nitromethane-methanol and propylene carbonate-methanol mixtures at 298.15 K. *J. Mol. Liq.* **2005**, *121*, 41−45.

(18) Rivas, M. A.; Pereira, S. M.; Banerji, N.; Iglesias, T. P. Permittivity and density of binary systems of {dimethyl or diethyl carbonate} + n-dodecane from T = (288.15 to 328.15) K. *J. Chem. Thermodyn.* **2004**, *36*, 183−191.

(19) George, J.; Sastry, N. V.; Prasad, D. H. L. Excess molar enthalpies and excess molar volumes of methyl methacrylate + benzene, + toluene, + p-xylene, + cyclohexane and + aliphatic diethers (diethyl, diisopropyl and dibutyl). *Fluid Phase Equilib.* **2003**, *214*, 39−51.

(20) Rodriguez, A.; Canosa, J.; Tojo, J. Physical properties of the binary mixtures (diethyl carbonate + hexane, heptane, octane and cyclohexane) from T = 293.15 K to T = 313.15 K. *J. Chem. Thermodyn.* **2003**, *35*, 1321−1333.

(21) Ogawa, H.; Karashima, S.; Takigawa, T.; Murakami, S. Excess molar enthalpies and volumes of binary mixtures of two hydrofluoroethers with hexane, or benzene, or ethanol, or 1-propanol, or 2-butanone at T = 298.15 K. *J. Chem. Thermodyn.* **2003**, *35*, 763−774.

(22) Manfredini, M.; Marchetti, A.; Sighinolfi, S.; Tassi, L.; Ulrici, A.; Vignali, M. Densities and excess molar volumes of binary mixtures containing 1,2-Dichloroethane + 2-methoxy ethanol or 1,2-dimethoxy ethane at different temperatures. *J. Mol. Liq.* **2002**, *100*, 163−181.

(23) Chan, C.; Maham, Y.; Mather, A. E.; Mathonat, C. Densities and volumetric properties of the aqueous solutions of 2-amino-2-methyl-1-propanol, n-butyldiethanolamine and n-propylethanolamine at temperatures from 298.15 to 353.15 K. *Fluid Phase Equilib.* **2002**, *198*, 239−250.

(24) Lebrette, L.; Maham, Y.; Teng, T. T.; Hepler, L. G.; Mather A. E. Volumetric properties of aqueous solutions of mono, and diethylethanolamines at temperatures from 5 to 80 °C II. *Thermochim. Acta* **2002**, *386*, 119−126.

(25) Maham, Y.; Teng, T. T.; Hepler, L. G.; Mather A. E. Volumetric properties of aqueous solutions of monoethanolamine, mono- and dimethylethanolamines at temperatures from 5 to 80 °C I. *Thermochim. Acta* **2002**, *386*, 111−118.

(26) Maham, Y.; Boivineau M.; Mather A. E. Density and excess molar volumes of aqueous solutions of morpholine and methylmorpholine at temperatures from 298.15 K to 353.15 K. *J. Chem. Thermodyn.* **2001**, *33*, 1725−1734.

(27) Rodriguez, A.; Canosa, J.; Tojo, J. Density, refractive index on mixing, and speed of sound of the ternary mixtures (dimethyl carbonate or diethyl carbonate + methanol + toluene) and the corresponding binaries at T = 298.15 K. *J. Chem. Thermodyn.* **2001**, *33*, 1383−1397.

(28) Zhang, F. U.; Li, H. P.; Dai, M.; Zhao, J. P. Volumetric properties of binary mixtures of water with ethanolamine alkyl derivatives. *Thermochim. Acta* **1995**, *254*, 347−357.

(29) Nishikawa, K.; Ohomuro, K.; Tamura, K.; Murakami, S. Excess thermodynamic properties of mixtures of cyclohexanone and benzene at 298.15 and 308.15 K and the effect of excess expansion factor. *Thermochim. Acta* **1995**, *267*, 323−332.

(30) Beg, S. A.; Tukur, N. M.; A1-Harbi, D. K. Densities and excess volumes of cyclohexane + hexane between 298.15 K and 473.15 K. *Fluid Phase Equilib.* **1995**, *113*, 173−184.

(31) Iglesias, M.; Orge, B.; Tojo, J. Refractive indices, densities and excess properties on mixing of the systems acetone + methanol + water and acetone + methanol + 1-butanol at 298.15 K. *Fluid Phase Equilib.* **1996**, *126*, 203−223.

(32) Hu, J.; Tamura, K.; Murakami, S. Excess thermodynamic properties of binary mixtures of ethyl formate with benzene, ethanol, and 2,2,2-trifluoroethan-1-ol at 298.15 K. *Fluid Phase Equilib.* **1997**, *131*, 197−212.

(33) Hu, J.; Tamura, K.; Murakami, S. Excess thermodynamic properties of binary mixtures of ethyl acetate with benzene, ethanol, and 2,2,2-trifluoroethan-1-ol at 298.15 K. *Fluid Phase Equilib.* **1997**, *134*, 239−253.

(34) Orge, B.; Iglesias, M.; Rodriguez, A.; Canosa, J. M.; Tojo, J. Mixing properties of (methanol, ethanol, or 1-propanol) with (n-pentane, n-hexane, n-heptane and n-octane) at 298.15 K. *Fluid Phase Equilib.* **1997**, *133*, 213−227.

(35) Nishimoto, M.; Tabata, S.; Tamura, K.; Murakami, S. Thermodynamic properties of the mixtures of methoxyethanol and cyclohexane: Measurements at the temperatures 293.15, 298.15 and 303.15 K above and below UCST. *Fluid Phase Equilib.* **1997**, 136, 235−247.

(36) Dai, M.; Zhang, F. Q.; Li, H. P.; Zhao, J. P. Excess enthalpies and excess volumes of N, N-dimethylethanolamine + 1,4-dioxane, + DMF, + DMA or + DMSO. *Fluid Phase Equilib.* **1997**, *138*, 231−239.

(37) Pina, C. G.; Francesconi, A. Z. New applications of the ERAS-Model: excess volumes of binary liquid mixtures of 1-alkanols with acetonitrile. *Fluid Phase Equilib.* **1998**, *143*, 143−152.

(38) Teodorescu, M.; Linek, J. Densities and excess volumes of pentan-3-one + 1,2-dichloroethane, + 1,3-dichloropropane, + 1,4-dichlorobutane, + trichloromethane, + 1,1,1-trichloroethane, + 1,1,2,2-tetrachloroethane binary mixtures. *Fluid Phase Equilib.* **1998**, *146*, 155−160.

(39) Teodorescu, M.; Linek, J.; Wichterle, I. Isothermal vapour-liquid equilibria and densities for the 5-chloropentan-2-one + n-hexane, + toluene and + ethylbenzene binary mixtures. *Fluid Phase Equilib.* **1998**, *149*, 127−138.

(40) Iglesias, M.; Pineiro, M. M.; Marino, G.; Orge, B.; Domınguez, M.; Tojo, J. Thermodynamic properties of the mixture benzene + cyclohexane + 2-methyl-2-butanol at the temperature 298.15 K: excess molar volumes prediction by application of cubic equations of state. *Fluid Phase Equilib.* **1999**, *154*, 123−138.

(41) Ohji, H.; Ogawa, H.; Murakami, S.; Tamura, K.; Grolier J. E. Excess volumes and excess thermal expansivities for binary mixtures of 2-ethoxyethanol with non-polar solvents at temperatures between 283.15 and 328.15 K. *Fluid Phase Equilib.* **1999**, *156*, 101−114.

(42) Canosa, J.; Rodriguez, A.; Tojo, J. Binary mixture properties of diethyl ether with alcohols and alkanes from 288.15 K to 298.15 K. *Fluid Phase Equilib.* **1999**, *156*, 57−71.

(43) Pal, A.; Dass, G. Excess molar volumes and viscosities of binary mixtures of tetraethylene glycol dimethyl ether (tetraglyme) with chloroalkanes at 298.15 K. *J. Mol. Liq.* **2000**, *84*, 327−337.

(44) Pal, A.; Kumar, H.; Kumar, A.; Dass, G. Excess molar volumes and viscosities of n-alkoxyethanols with dialkyl carbonates at 298.15 K. *Fluid Phase Equilib.* **1999**, *166*, 245−258.

(45) Ohji, H. Excess thermodynamic properties of (2-ethoxy ethanol + 1,4-dioxane or 1,2-dimethoxyethane) at temperatures between (283.15 and 313.15) K. *J. Chem. Thermodyn.* **2000**, *32*, 319−328.

(46) Orge, B.; Iglesias, M.; Marino, G.; Dominguez, M.; Pineiro, M. M.; Tojo, J. Mixing properties of benzeneq2-methyl-2-butanol + 1-pentanol at 298.15 K. Experimental results and comparison between ERAS model and cubic EOS estimations for excess molar volumes. *Fluid Phase Equilib.* **2000**, *170*, 151−163.

(47) Ortega, J.; Gonzalez, C.; Pena, J.; Galvan, S. Thermodynamic study on binary mixtures of propyl ethanoate and an alkan-1-ol (C2−C4) Isobaric vapor-liquid equilibria and excess properties. *Fluid Phase Equilib.* **2000**, *170*, 87−111.

(48) Lugo, L.; Lopez, E. R.; Garcia, J.; Comunas, M. J. P.; Fernandez, J. Analysis of the molecular interactions of organic anhydride + alkane binary mixtures using the Nitta-Chao model. *Fluid Phase Equilib.* **2000**, *170*, 69−85.

(49) Doi, H.; Tamura, K. Thermodynamic properties of aqueous solution of 2-isobutoxyethanol at T = (293.15, 298.15, and 303.15) K, below and above LCST. *J. Chem. Thermodyn.* **2000**, *32*, 729−741.

(50) Takigawa, T. Thermodynamic properties of (1,3-dioxane, or 1,4-dioxane + a non-polar liquid) at T = 298:15 K; speed of sound, excess isentropic and isothermal compressibilities and excess isochoric heat capacity. *J. Chem. Thermodyn.* **2000**, *32*, 1045−1055.

(51) Sastry, N. V.; Patel, S. R. Excess volumes and dielectric properties for (methyl methacrylate + a branched alcohol) at T = 298.15 K and T = 308.15 K. *J. Chem. Thermodyn.* **2000**, *32*, 1669−1682.

(52) Kovacs, E.; Aim, K.; Linek, J. Excess molar volumes of (an alkane + 1-chloroalkane) at T = 298.15 K. *J. Chem. Thermodyn.* **2001**, *33*, 33−45.

(53) Jimenez, E.; Cabanas, M.; Segade, L.; Garcia-Garabal, S.; Casas, H. Excess volume, changes of refractive index and surface tension of binary 1,2-ethanediol + 1-propanol or 1-butanol mixtures at several temperatures. *Fluid Phase Equilib.* **2001**, *180*, 151−164.

(54) Resa, J. M.; Iglesias, M.; Gonzalez, C.; Lanz, J.; Mtz de Ilarduya, J. A. Excess volumes of binary mixtures of vinyl acetate and aromatic hydrocarbons. *J. Chem. Thermodyn.* **2001**, *33*, 723−732.

(55) Geyer, H.; Ulbig, P.; Gornert, M.; Susanto, A. Measurement of densities and excess molar volumes for (1,2-propanediol, or 1,2-butanediol + water) at the temperatures (288.15, 298.15, and 308.15) K and at the pressures (0.1, 20, 40, and 60) MPa. *J. Chem. Thermodyn.* **2001**, *33*, 987−997.

(56) Valen, A.; Lopez, M. C.; Urieta, J. S.; Royo, F. M.; Lafuente, C. Thermodynamic study of mixtures containing oxygenated compounds. *J. Mol. Liq.* **2002**, *95*, 157−165.

(57) Hiroyuki, O. Excess volumes of (1-pentanol + cyclohexane or benzene) at temperatures between 283.15 and 328.15 K. *J. Chem. Thermodyn.* **2002**, *34*, 849−859.

(58) Takigawa, T.; Minamihounoki, T.; Tamura, K. Excess enthalpies and excess volumes of binary mixtures of hydrofluoroether with alcohols. *J. Chem. Thermodyn.* **2002**, *34*, 841−847.

(59) Resa, J. M.; Gonzalez, C.; Ortiz de Landaluce, S., Lanz, J. (Vapour + liquid) equilibria, densities, excess molar volumes, refractive indices, speed of sound for (methanol + allyl acetate) and (vinyl acetate + allyl acetate). *J. Chem. Thermodyn.* **2002**, *34*, 1013−1027.

(60) Resa, J. M.; Gonzalez, C.; Ortiz de Landaluce, S., Lanz, J. Densities, excess molar volumes, and refractive indices of ethyl acetate and aromatic hydrocarbon binary mixtures. *J. Chem. Thermodyn.* **2002**, *34*, 995−1004.

(61) Peralta, R. D.; Infante, R.; Cortez, G.; Villarreal, L.; Wisniak, J. Volumetric properties of cyclohexane with ethyl acrylate, butyl acrylate, methyl methacrylate, and styrene at 298.15 K. *Thermochim. Acta* **2002**, *390*, 47−53.

(62) Torres, R. B.; Francesconi, A. Z.; Volpe, P. L. O. Excess molar volumes of binary mixtures of acetonitrile and chloroalkanes at 298.15 K and atmospheric pressure with application of the ERAS-Model. *Fluid Phase Equilib.* **2002**, *200*, 1−10.

(63) Torres, R. B.; Francesconi, A. Z. Application of the ERAS-Model to binary mixtures of diethylamine and s-butylamine with acetonitrile in the temperature range (288.15−303.15) K. *Fluid Phase Equilib.* **2002**, *200*, 317−328.

(64) Pal, A.; Bhardwaj, R. K. Excess molar volumes and viscosities of binary mixtures of diethylene glycol diethyl ether with chloroalkanes at 298.15 K. *J. Mol. Liq.* **2003**, *102*, 197−211.

(65) Peralta, R. D.; Infante, R.; Cortez, G.; Ramirez, R. R.; Wisniak, J. Densities and excess volumes of binary mixtures of 1,4-dioxane with either ethyl acrylate, or butyl acrylate, or methyl methacrylate, or styrene at T = 298.15 K. *J. Chem. Thermodyn.* **2003**, *35*, 239−250.

(66) Peralta, R. D.; Infante, R.; Cortez, G.; Cisneros, A.; Wisniak, J. Densities and excess volumes of benzene with ethyl acrylate, butyl acrylate, methyl methacrylate, and styrene at 298.15 K. *Thermochim. Acta* **2003**, *398*, 39−46

(67) Peralta, R. D.; Infante, R.; Cortez, G.; Torres-Lubian, J. R.; Wisniak, J. Volumetric properties of 1,2-dimethylbenzene + ethyl acrylate, butyl acrylate, methyl methacrylate, and styrene at 298.15 K. *Thermochim. Acta* **2003**, *402*, 247−252.

(68) Moravkova, L.; Linek, J. Excess molar volumes of (benzene + isopropylbenzene, or 1,3,5-trimethylbenzene, or 1,2,4-trimethylbenzene) at temperatures between 298.15 K and 328.15 K. *J. Chem. Thermodyn.* **2003**, *35*, 1139−1149.

(69) Ohji, H.; Tamura, K. Excess enthalpies and excess volumes of (2-methoxyethanol + 1,4-dioxane) and (1,2-dimethoxyethane + benzene) at temperatures between 283.15K and 313.15K. *J. Chem. Thermodyn.* **2003**, *35*, 1591−1599.

(70) Dominguez-Perez, M.; Jimenez de Llano, J.; Segade, L.; Cabeza, O.; Franjo, C.; Jimenez, E. Excess molar volumes and enthalpies for the binary systems propyl propanoate + o-xylene, m-xylene, and p-xylene at 298.15K. *Thermochim. Acta* **2004**, *420*, 45−49.

(71) Peralta, R. D.; Infante, R.; Cortez, G.; Elizalde, L. E.; Wisniak, J. Density, excess volumes and partial volumes of the systems of p-xylene + ethyl acrylate, butyl acrylate, methyl methacrylate, and styrene at 298.15K. *Thermochim. Acta* **2004**, *421*, 59−68.

(72) Villares, A.; Martin, S.; Haro, M.; Giner, B.; Artigas, H. Densities and speeds of sound for binary mixtures of (1,3-dioxolane or 1,4-dioxane) with (2-methyl-1-propanol or 2-methyl-2-propanol) at the temperatures (298.15 and 313.15) K. *J. Chem. Thermodyn.* **2004**, *36*, 1027−1036.

(73) Moravkova, L.; Wagner, Z.; Linek, J. (P, Vm, T) Measurements of (toluene + propiophenone) at temperatures from 298.15 K to 328.15 K and at pressures up to 40 MPa. *J. Chem. Thermodyn.* **2005**, *37*, 658−666.

(74) Schrodle, S.; Hefter, G.; Buchner, R. Effects of hydration on the thermodynamic properties of aqueous ethylene glycol ether solutions. *J. Chem. Thermodyn.* **2005**, *37*, 513−522.

(75) Valtz, A.; Coquelet, C.; Richon, D. Volumetric properties of the monoethanolamine-methanol mixture at atmospheric pressure from 283.15 to 353.15 K. *Thermochim. Acta* **2005**, *428*, 185−191.

(76) Al-Jimaz, A. S.; Al-Kandary, J. A.; Abdul-latif, A. H. M.; Al-Zanki, A. M. Physical properties of {anisole + *n*-alkanes} at temperatures between (293.15 and 303.15) K. *J. Chem. Thermodyn.* **2005**, *37*, 631−642.

(77) Moravkova, L.; Linek, J. Excess molar volumes of (acetophenone + benzene, or toluene, or 1,3-xylene, or 1,3,5-trimethylbenzene) at temperatures (298.15 and 328.15) K. *J. Chem. Thermodyn.* **2005**, *37*, 814−819.

(78) Moravkova, L.; Linek, J. Excess molar volumes of (propiophenone + benzene, or toluene, or ethylbenzene, or butylbenzene) at temperatures (298.15 and 328.15) K. *J. Chem. Thermodyn.* **2005**, *37*, 1023−1028.

(79) *Cerius2 software, version 4.8*; Accelrys Inc.. http://www.accelrys.com/.

(80) *E-Dragon 1.0 On-line software*; Virtual Computational Chemistry Laboratory. http://146.107.217.178/lab/edragon/index.html.

(81) Golbraikh, A.; Tropsha, A. QSAR Modeling Using Chirality Descriptors Derived from Molecular Topology. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 144−154.

(82) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*; Wiley: New York, 1986.

(83) Zheng, W.; Tropsha, A. Novel Variable Selection Quantitative Structure−Property Relationship Approach Based on the k-Nearest-Neighbour Principle. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185−194.

(84) Zupan, J.; Gasteiger, J. *Neural Networks for Chemists*; VCH-Wiley: Weinheim, Germany, 1993; p 207.

(85) Tambe, S. S.; Kulkarni, B. D.; Deshpande, P. B. *Elements of Artificial Neural Networks with Selected Applications in Chemical Engineering, Chemical and Biological Sciences*; Simulation and Advanced Controls Inc.: Louisville, KY, 1996.

(86) Andrea, T. A.; Kalayeh, H. Application of Neural Networks in Quantitative Structure−Activity Relationships of Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1991**, *34*, 2824−2836.

(87) Gilbert, N. *Statistics*; W.B. Saunders, Co.; Philadelphia, PA, 1976.

(88) *MATLAB, version 6.1*; The MathWorks Inc.: Natick, MA.

(89) Sun, L.; Xie, Y.; Song, X.; Wang, J.; Yu, R. Cluster Analysis By Simulated Annealing. *Comput. Chem.* **1994**, *18*, 103−108.

(90) Aarts, E.; Korst, J. *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*; John Wiley & Sons: Chichester, 1989.

(91) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Mannheim, 2000.

(92) Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on rational division of experimental datasets into diverse training and test sets. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 357−369.

(93) Shen, M.; Xiao, Y.; Golbraikh, A.; Gombar, V. K.; Tropsha, A. Development and Validation of k-Nearest-Neighbor QSPR Models of Metabolic Stability of Drug Candidates. *J. Med. Chem.* **2003**, *46*, 3013−3020.

CI050559O