Research article

Drug Testing and Analysis

Received: 23 February 2011 Revised: 10 March 2011 Accepted: 10 March 2011 Published online in Wiley Online Library: 3 May 2011

(www.drugtestinganalysis.com) DOI 10.1002/dta.288

Prediction of polar surface area of drug molecules: A QSPR approach

H. Noorizadeh,[a]* A. Farmany,[a] M. Noorizadeh[b] and M. Kohzadi[b]

A quantitative structure-property relationship (QSPR) study based on an artificial neural network (ANN) was carried out for the prediction of the microemulsion liquid chromatography polar surface area (PSA) of a set of 32 drug compounds. The genetic algorithm-kernel partial least squares (GA-KPLS) method was used as a variable selection tool. A KPLS method was used to select the best descriptors and the selected descriptors were used as input neurons in neural network model. For choosing the best predictive model from among comparable models, square correlation coefficient Q^2 for the whole set calculated based on leave-group-out predicted values of the training set and model-derived predicted values for the test set compounds is suggested to be a good criterion. Finally, to improve the results, structure-property relationships were followed by nonlinear approach using artificial neural networks and consequently better results were obtained. Also this demonstrates the advantages of ANN.

Keywords: drug molecules; microemulsion liquid chromatography; polar surface area; QSPR; genetic algorithm-kernel partial least squares; Levenberg-Marquardt artificial neural network

Introduction

The modern pharmaceutical industry is constantly in search of new, innovative chemical entities. Today, combinatorial chemistry and high throughput methods enable large numbers of drug candidates to be screened for activity. Substances identified by these methods are often characterized by high lipophilicity and low solubility, since potential drug candidates are selected primarily by their receptor affinity *in vitro*. Poor biopharmaceutical properties are a major reason for the failure of new chemical entities (NCEs) in pharmaceutical drug discovery. For example, poor aqueous solubility often results in poor bioavailability after oral dosing. But the gastrointestinal tract is often not the only barrier that the drug must pass through in order to reach its site of action. For drugs that are intended to act in the central nervous system, the ability to cross the blood–brain barrier (BBB) is a further, crucial prerequisite for success. Conversely, low BBB penetration may be desirable in order to minimize CNS-related side effects for drugs that are designed to act at peripheral sites.[1–3] Thus, the ability to cross the BBB has emerged as an issue of primary interest, even in the early stages of drug discovery. Tight junctions in the BBB form a physical barrier, which restricts the transportation of macromolecules between the blood and the brain for the maintenance of cerebral homeostasis. These tight junctions are composed of membrane-associated accessory proteins, such as occludin and claudin, and intracellular proteins.[4,5] The BBB restricts the transport of many therapeutically important drugs from the blood stream into the brain. This barrier is formed by the endothelial cells of the cerebral capillaries, which essentially comprise the major exchange interface between the blood and the brain. Different mechanisms, such as passive transcellular diffusion, paracellular diffusion, and active transport, have been proposed for uptake across this barrier,[6–8] with the tight endothelium of brain capillaries constituting the principal permeability barrier for the passive transport of substances across the barrier. An established method to reflect the distribution of drugs between blood and brain tissue and hence their ability to penetrate the BBB is to determine the ratio of the steady state concentration of the drug in the brain to its concentration in the blood.

The polar surface area (PSA) is defined as the surface sum over all polar atoms (usually oxygen and nitrogen), including also attached hydrogens. PSA is a commonly used medicinal chemistry metric for the optimization of cell permeability. Molecules with a PSA of greater than 140 angstroms squared are usually believed to be poor at permeating cell membranes. For molecules to penetrate the BBB (and thus acting on receptors in the central nervous system), PSA should be less than 60 angstroms squared.[9,10]

It has been more than 10 years since microemulsion was used as the mobile phase in a high performance liquid chromatography (HPLC) system.[11,12] This HPLC mode, namely microemulsion liquid chromatography (MELC), has received growing interest and is showing great potential in separation sciences due to its unique selectivity and high efficiency compared to conventional HPLC.[13–17] The MELC, which utilizes microemulsions as a mobile phase and has been shown to be suitable for the separation of a range of pharmaceutical compounds using both isocratic and gradient elution modes [18] and for validated determinations of fosinoprilat in human plasma and simvastatin and its impurities in bulk drug and tablet formulations.[19] Also biopartitioning micellar chromatography (BMC) has been used in modelling many biopartitioning processes, including human drug absorption,[20] BBB penetration,[21] ocular tissue permeability,[22] skin permeability,[23] etc. Using a microemulsion as the mobile phase alters the solutes partitioning between the mobile and stationary phase characteristics in comparison with conventional HPLC because a layer

* *Correspondence to: H. Noorizadeh, Islamic Azad University, Ilam Branch, Ilam, Iran. E-mail: hadinoorizadeh@yahoo.com*

a *Islamic Azad University, Ilam Branch, Ilam, Iran*

b *Members of Young Researchers Club, Islamic Azad University, Ilam Branch, Ilam, Iran*

of surfactant molecules adsorbs on the surface of the stationary phase. The solutes also partition from the aqueous phase or the stationary phase onto the microemulsion droplets. The eluent parameters that mainly affect separations in MELC are surfactant concentration, percent of a small alcohol as a cosurfactant, and percent of organic oily solvent, temperature and pH of the mobile phase.[24] When there are many parameters and especially if the factors interact, application of experimental designs has been proposed to optimized separation condition and reduces the number of experiments and amounts of chemicals. MELC separations are affected by more than two factors as compared with hydro-organic reversed-phase HPLC. For this reason, it is necessary to simultaneously optimize the factors affecting the separation. Application of optimization algorithms result in the desired optimized experimental conditions with faster convergence.

Quantitative structure-property relationship (QSPR) techniques based on different molecular descriptors have been successfully used to model organic chemicals properties.[25] A number of reports that deals with QSPR calculation of several compounds have been published in the literature.[26–28] The QSPR models can be applied to partial least squares (PLS) methods often combined with genetic algorithms (GA) for feature selection.[29–31] Because of the complexity of relationships between the property of molecules and structures, nonlinear models are also used to model the structure-property relationships. Levenberg-Marquardt artificial neural network (L-M ANN) is nonparametric, nonlinear modelling technique that has attracted increasing interest. In the recent years, nonlinear kernel-based algorithms as kernel partial least squares (KPLS) have been proposed.[32,33] The basic idea of KPLS is first to map each point in an original data space into a feature space via nonlinear mapping and then to develop a linear PLS model in the mapped space. According to Cover's theorem, nonlinear data structure in the original space is most likely to be linear after high-dimensional nonlinear mapping.[34] Therefore, KPLS can efficiently compute latent variables in the feature space by means of integral operators and nonlinear kernel functions. Compared to other nonlinear methods, the main advantage of the kernel-based algorithm is that it does not involve nonlinear optimization. It essentially requires only linear algebra, making it as simple as the conventional linear PLS. In addition, because of its ability to use different kernel functions, KPLS can handle a wide range of nonlinearities. In the present study, GA-PLS, GA-KPLS, and L-M ANN were employed to generate QSPR models that correlate the structure of some drugs; with observed PSA.

Experimental

Data set

In the current research, the data set was taken from Liu *et al*.[35] The drug set comprises 32 commercially available drugs. Those which are reported to transport across cell membrane with the participation of some transport protein and which have molecular weight less than 200 and are hydrophilic at the physiological conditions, were excluded from the drug set in advance, for example, verapamil, quinidine, doxorubicin, vinblastine, risperidone and salicylic acid, caffeine, theophylline, acetaminophen. The chromatograph was equipped with an LC-10AT pump (SHIMADZU, Japan), a SPD-10A UV–vis detector (SHIMADZU, Japan), a manual injection valve (7725i, USA) and a HT-130 column heater (HITACHI, Japan). The flow rate was set at 0.8 ml/min. The detection wavelength was set at 210 nm for all drugs. The column temperature was maintained at 35 °C. The injection volume was 10 µl. An at ChromC_{18} column (150 mm × 4.6 mm, 5 µm) (Lanzhou Institute of Chemical Physics, Lanzhou, China) equipped with a C_{18} guard column (12.5 mm × 4.6 mm, 5 µm) was used for separation. The pore size of packing materials in the analytical column is 10 nm. Chromatographic signals were acquired and processed by Anastar chromatography data system (Version 5.2) (Autoscience Instrument Co. Ltd, Tianjin, China). A complete list of the drugs' names and their corresponding experimental PSA is given in Table 1.

Table 1. The data set and the corresponding observed and predicted PSA values by L-M ANN for training and test set

No	Name	PSA_{Exp}	PSA_{Cal}	RE	SE
Training set					
1	Amitriptyline	3.24	2.84	12.3	1.9
2	Imipramine	6.48	5.97	7.9	1.8
3	Mirtazapine	19.4	17.8	8.2	1.3
4	Midazolam	25.3	26.9	6.3	0.9
5	Aminopyrine	26.8	27	0.7	0.9
6	Diazepam	32.7	31.6	3.4	0.7
7	Trifluoperazine	35	38	8.6	0.5
8	Clonidine	36.4	39.1	7.4	0.4
9	Ibuprofen	37.3	35.3	5.4	0.6
10	Alprazolam	38.1	40.5	6.3	0.4
11	Haloperidol	40.5	45.6	12.6	0.1
12	Propranolol	41.5	45.7	10.1	0.1
13	Codeine	41.9	42.3	1.0	0.3
14	Physostigmine	44.8	45.6	1.8	0.1
15	Morphine	52.4	47.8	8.8	0.1
16	Fluphenazine	55.3	51.6	6.7	0.1
17	Thioridazine	57.1	65.2	14.2	0.7
18	Oxazepam	61.7	67.8	9.9	0.8
19	Indomethacin	63.6	63.9	0.5	0.6
20	Salbutamol	72.7	65.4	10.0	0.7
21	Amobarbital	75.3	78.5	4.2	1.2
22	Atenolol	84.6	91.2	7.8	1.8
23	Ranitidine	112	100.6	10.2	2.1
24	Cimetidine	114	118.5	3.9	2.9
Test Set					
25	Mianserin	6.48	5.64	13.0	5.4
26	Chlorpromazine	31.8	26	18.2	2.9
27	Tibolone	37.3	34.7	7.0	1.8
28	Phenylbutazone	40.6	48.2	18.7	0.1
29	Carbamazepine	46.3	50.1	8.2	0.1
30	Phenytoin	58.2	55.4	4.8	0.8
31	Pentobarbital	75.3	63.5	15.7	1.8
32	Omeprazole	96.3	115.6	20.0	8.3

The data set was randomly divided into training (calibration and prediction sets) and test sets after sorting based on the PSA values. The training set consisted of 24 molecules; the test set consisted of 8 molecules. The training set was used for model development, while the test set – in which its molecules have no role in model building – was used for evaluating the predictive ability of the models for external set.

Computer hardware and software

All calculations were run on an HP laptop computer with an AMD Turion64X2 processor and a Windows XP operating system. The

Table 2. Parameters of the genetic algorithm

Population size: 30 chromosomes
On average, five variables per chromosome in the original population
Regression method: PLS, KPLS
Cross validation: leave-group-out
Number subset: 4
Maximum number of variables selected in the same chromosome: (PLS, 30)
Elitism: True
Crossover: multi Point
Probability of crossover: 50%
Mutation: multi Point
Probability of mutation: 1%
Maximum number of components: (PLS, 10)
Number of runs: 100

optimizations of molecular structures were done by HyperChem 7.0 (AM1 method) and descriptors were calculated by Dragon Version 3.0 software.[36] MINITAB software Version 14 was used for simple PLS analysis. Cross validation, GA-PLS, GA-KPLS, L-M ANN and other calculation were performed in MATLAB (Version 7, Mathworks, Inc.) environment.

Genetic algorithm

A detailed description of the GA can be found in the literature.[37–39] Genetic algorithm is simulated methods based on ideas from Darwin's theory of natural selection and evolution (the struggle for life). In GA, a chromosome (or an individual) can be defined as an enciphered entity of a candidate solution, which is expressed as a set of variables. GA consist of the following basic steps: (1) A chromosome is represented by a binary bit string and an initial population of chromosomes is created in a random way; (2) A value for the fitness function of each chromosome is evaluated; (3) Based on the values of the fitness functions, the chromosomes of the next generation are produced by selection, crossover and mutation operations. The fitness function was proposed by Depczynski *et al*.[40] The parameters for the algorithm are reported in Table 2.

Results and discussion

Linear model

Results for the GA-PLS model

To reduce the original pool of descriptors to an appropriate size, the objective descriptor reduction was performed using various criteria. Reducing the pool of descriptors eliminates those descriptors which contribute either no information or whose information content is redundant with other descriptors present in the pool. After this process, 1003 descriptors remained. These descriptors were employed to generate the models with the GA-PLS and GA-KPLS program. The best GA-PLS model contains 11 selected descriptors in 4 latent variables space. These descriptors were obtained constitutional descriptors (sum of atomic Sanderson electronegativities (scaled on Carbon atom) (Se) and number of Oxygen atoms (nO)), 3D-MoRSE descriptors (3D-MoRSE - signal 05/weighted by atomic masses) (Mor05m), WHIM descriptors (1st component symmetry directional WHIM index/weighted by atomic electrotopological states (G1s) and (1st component size directional WHIM index/weighted by atomic polarizabilities (L1p)), molecular properties (topological polar surface area using N, O, S, P polar contributions (TPSA(Tot)), functional group counts (number of donor atoms for H-bonds (N and O) (nHDon)), atom-centred fragments (phenol/enol/carboxyl OH (O-057) and quantum chemical descriptors (dipole moment, polarizibility and lowest unoccupied molecular orbital (LUMO)). The obtained statistic parameters of the GA-PLS model are shown in Table 3. For this in general, the number of components (latent variables) is less than the number of independent variables in PLS analysis. The PLS model uses higher number of descriptors that allow the model to extract better structural information from descriptors to result in a lower prediction error.

Table 3. The statistical parameters of different constructed QSPR models

	Training set					Test set			
Model	R^2	Q^2	RE	RMSE	N	R^2	RE	RMSE	N
GA-PLS	0.834	0.831	19.40	10.06	24	0.663	31.85	17.34	8
GA-KPLS	0.891	0.893	10.12	7.58	24	0.822	19.01	12.67	8
L-M ANN	0.964	0.966	7.00	4.43	24	0.921	13.2	8.89	8

Nonlinear models

Results for the GA-KPLS model

The leave-group-out cross validation (LGO-CV) has been performed. In this paper a radial basis kernel function, $k(x,y)=\exp(||x-y||^2/c)$, was selected as the kernel function with $c = rm\sigma^2$, where r is a constant that can be determined by considering the process to be predicted (here r was set to be 1), m is the dimension of the input space and σ^2 is the variance of the data.[41] It means that the value of c depends on the system under the study. The eight descriptors in three latent variables space chosen by GA-KPLS feature selection methods were contained. These descriptors were obtained constitutional descriptors (number of Hydrogen atoms (nH) and number of Nitrogen atoms (nN)), WHIM descriptors 1st (G1s), molecular properties (topological polar surface area using N, O polar contributions) (TPSA (NO)), atom-centred fragments (phenol/enol/carboxyl OH (O-057), H attached to C_1 (sp^3)/C_0 (sp^2) (H-047)) and quantum chemical descriptors (dipole moment and LUMO). Plots of predicted PSA values by GA-KPLS versus experimental PSA values for training and test set are shown in Figure 1. The square correlation coefficient LGO-CV (Q^2), relative error (RE) and root mean square error (RMSE) of this model are represented in Table 3.

The data presented in Table 3 indicate that the GA-PLS linear model has good statistical quality with low prediction error, while the corresponding errors obtained by the GA-KPLS model are lower. Consequently, this GA-KPLS approach currently constitutes the most accurate method for predicting the partitioning of the drug components than that of the GA-PLS method. This suggests that GA-KPLS hold promise for applications in choosing of variable for L-M ANN systems. This result indicates that the log K_s of drug molecules possesses some nonlinear characteristics.

Results for the L-M ANN model

With the aim of improving the predictive performance of nonlinear QSPR model, L-M ANN modelling was performed. Descriptors of

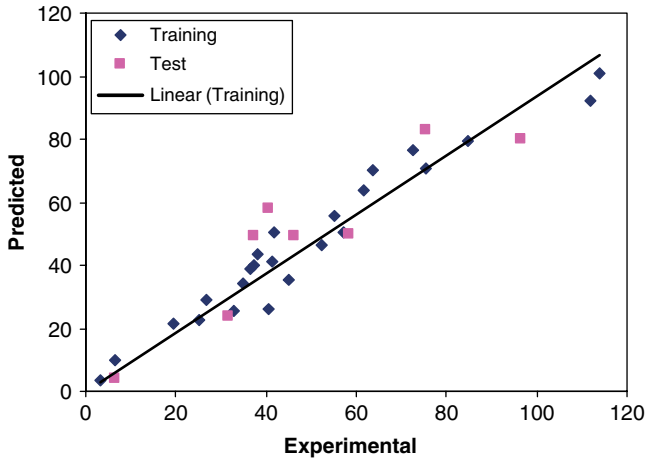

Figure 1. Plot of predicted PSA obtained by GA-KPLS against the experimental values.

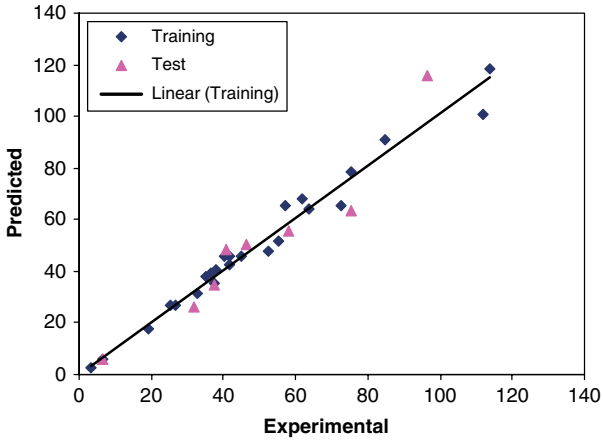

Figure 2. Plot of predicted PSA obtained by L-M ANN against the experimental values.

GA-KPLS model were selected as inputs in L-M ANN model. The network architecture consisted of seven neurons in the input layer corresponding to the seven mentioned descriptors. The output layer had one neuron that predicts the PSA. The number of neurons in the hidden layer is unknown and needs to be optimized. In addition to the number of neurons in the hidden layer, the learning rate, the momentum and the number of iterations also should be optimized. In this work, the number of neurons in the hidden layer and other parameters except the number of iterations were simultaneously optimized. A MATLAB program was written to change the number of neurons in the hidden layer from two to seven, the learning rate from 0.001 to 0.1 with a step of 0.001 and the momentum from 0.1 to 0.99 with a step of 0.01. The root mean square errors for training set were calculated for all of the possible combination of values for the mentioned variables in LGO-CV. It was realized that the RMSE for the training set was minimum when two neurons were selected in the hidden layer and the learning rate and the momentum values were 0.4 and 0.3, respectively. Finally, the number of iterations was optimized with the optimum values for the variables. It was realized that after 13 iterations, the RMSE for prediction set was minimum. The values of experimental, calculated, and percent relative error and standard error are shown in Table 1. The statistical parameters obtained by LGO-CV for L-M ANN, GA-KPLS and the linear QSPR models are compared in Table 3. Each of the statistical parameters mentioned above were used for assessing the statistical significance of the QSPR model. Inspection of the results reveals a higher Q^2 and lowers other parameter values for the training set compared with their counterparts for GA-KPLS and GA-PLS. Plots of predicted PSA values by L-M ANN versus experimental PSA values for training and test set are shown in Figure 2. Obviously, there is a close agreement between the experimental and predicted PSA and the data represent a very low scattering around a straight line with respective slope and intercept close to one and zero. This clearly shows the strength of L-M ANN as a nonlinear feature selection method. The key strength of L-M ANN is their ability to allow for flexible mapping of the selected features by manipulating their functional dependence implicitly. Neural network handles both linear and nonlinear relationship without adding complexity to the model. This capacity offset the large computing time required and complexity of L-M ANN model with respect other models.

Interpretation of descriptors

PSA is formed by polar atoms of a molecule. It is a descriptor that shows good correlation with passive molecular transport through membranes, and so allows estimation of transport properties of drugs. Drug absorption depends on the lipid solubility of the drug, its formulation, and the route of administration. A drug needs to be lipid soluble to penetrate membranes unless there is an active transport system or it is so small that it can pass through the aqueous channels in the membrane. Drug penetration may be attributed mostly to the un-ionized form. Distribution of an ionisable drug across a membrane at equilibrium is determined by the drug's PSA.[42–44]

Constitutional descriptors are the simplest and most commonly used descriptors, reflecting the molecular composition of a compound without any information about its molecular geometry. The most common constitutional descriptors are number of atoms, number of bound, absolute and relative numbers of specific atom type, absolute and relative numbers of single, double, triple, and aromatic bound, number of ring, number of ring divided by the number of atoms or bonds, number of benzene ring, number of benzene ring divided by the number of atom, molecular weight and average molecular weight.

Different hydrogen bond donors and acceptors are two important parameters introduced to describe molecular properties important for a drug's pharmacokinetics in the human body. The availability to form H bonds is an important parameter to define the physical-chemical properties of a drug.

3D-MoRSE (3D-MOlecule Representation of Structures based on Electron diffraction) descriptors are based on the idea of obtaining information from the 3D atomic coordinates by the transform used in electron diffraction studies. These descriptors are calculated by summing atom weights viewed by a divergent angular scattering function.

The WHIM descriptors are built in such a way as to capture the relevant molecular 3-D information regarding the molecular size, shape, symmetry, and atom distribution with respect to some invariant reference frame. These descriptors are quickly computed from the atomic positions of the molecule atoms (hydrogen included). WHIM descriptors are based on principal component analysis of the weighted covariance matrix obtained from the atomic Cartesian coordinates.

Different hydrogen bond donors and acceptors are two important parameters introduced by Lipinski *et al*. [45] to describe molecular properties important for a drug's pharmacokinetics in the human body. The availability to form H bonds is an important parameter to define the physicochemical properties of a drug.

Although constitutional descriptors, functional group, atom-centred fragments and molecular properties are often successful in PSA of drug molecules, they cannot account for conformational changes and they do not provide information about electronic influence through bonds or across space. For that reason, quantum chemical descriptors are used in developing QSPR.

Quantum chemical descriptors can give great insight into structure and reactivity and can be used to establish and compare the conformational stability, chemical reactivity, and intermolecular interactions. They include thermodynamic properties (system energies) and electronic properties (LUMO energy). Quantum chemical descriptors were defined in terms of atomic charges and used to describe electronic aspects both of the whole molecule and of particular regions, such atoms, bonds, and molecular fragments. Electronic properties may play a role in the magnitude in a biological activity, along with structural features encoded in indexes. LUMO as an electron acceptor represents the ability to obtain an electron. The energy of the LUMO is directly related to the electron affinity and characterizes the susceptibility of the molecule toward attack by nucleophiles. The LUMO energy can be interpreted as a measure of charge transfer interactions and/or of hydrogen bonding effects. Electron affinity was also shown to greatly influence the chemical behaviour of compounds, as demonstrated by its inclusion in the QSPR.

TPSA (NO) of a molecule is defined as the surface sum over of polar atoms. This molecular descriptor explains the electrostatic and polarization interactions between the solute and the solvent. All the interactions are obviously weak interactions such as higher multipole, dipole and induced-dipole interactions. So, TPSA (NO) can be considered an important electrostatic descriptor during a QSPR study to understand the charge distribution of the molecules and use this information to project new drugs with desired properties.[46]

Conclusion

In this research, an accurate QSPR model for estimating the PSA was developed by employing the one linear model (GA-PLS) and two nonlinear models (GA-KPLS and L-M ANN). Three models have good predictive capacity and excellent statistical parameters. A comparison between these models revealed the superiority of the L-M ANN to other models. It is easy to notice that there was a good prospect for the L-M ANN application in the QSPR modelling. This indicates that PSA of drug molecules possesses some nonlinear characteristics. The results showed that the L-M ANN model can be effectively used to describe the molecular structure characteristic of these compounds. It can also be used successfully to estimate the PSA for new compounds or for other compounds whose experimental values are unknown.

References

[1] X. C. Fu, C. X. Chen, WQ. Liang, Q. S. Yu. Predicting blood-brain barrier penetration of drugs by polar molecular surface area and molecular volume. *Acta Pharmacol. Sin*. **2001**, *22*, 663.
[2] B. Diamond, P. T. Huerta, P. Mina-Osorio, C. Kowal, B. T. Volpe. Losing your nerves? Maybe it's the antibodies. *Nat. Rev. Immunol.* **2009**, *9*, 449.
[3] A. Alonso, E. Reinz, M. Fatar, J. Jenne, M. G. Hennerici, S. Meairs. Neurons but not glial cells overexpress ubiquitin in the rat brain following focused ultrasound-induced opening of the blood–brain barrier. *Neuroscience* **2010**, *169*, 116.
[4] L. L. Mitic, J.M. Anderson. Molecular Architecture of Tight Junctions. *Annu. Rev. Physiol.* **1998**, *60*, 121.
[5] B. V. Zlokovic. The Blood-Brain Barrier in Health and Chronic Neurodegenerative Disorders. *Neuron* **2008**, *57*, 178.
[6] D. J. Begley. Delivery of therapeutic agents to the central nervous system: the problems and the possibilities. *Pharmacol. Ther.* **2004**, *104*, 29.
[7] W. M. Pardridge. Blood–brain barrier delivery. *Drug Discov. Today* **2007**, *12*, 54.
[8] S. Ohtsuki, T. Terasaki. Contribution of Carrier-Mediated Transport Systems to the Blood–Brain Barrier as a Supporting and Protecting Interface for the Brain; Importance for CNS Drug Discovery and Development. *Pharm. Res.* **2007**, *24*, 1745.
[9] P. Ertl, B. Rohde, P. Selzer. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J.Med. Chem*. **2000**, *43*, 3714.
[10] P. Ertl, Polar surface area, in *Molecular Drug Properties*, (ed: R. Mannhold), Wiley-VCH: Weinheim, Germany, **2007**.
[11] A. Berthod, O. Nicolas, M. Porthault. Water in oil microemulsions as mobile phase in liquid chromatography. *Anal. Chem.* **1990,** *62,* 1402.
[12] A. Berthod, M. D. Carvalho. Oil in water microemulsions as mobile phases in liquid chromatography. *Anal. Chem.* **1992**, *64*, 2267.
[13] D. T. M. El-Sherbiny, S. M El- Ashry, M. A. Mustafa, A. A. El-Emam, S. H. Hansen. Evaluation of the use of microemulsions as eluents in high-performance liquid chromatography. *J. Sep. Sci*. **2003**, *26*, 503.
[14] A. Marsh, B. Clark, K. *Altria. Oil-in-Water Microemulsion High Performance Liquid Chromatographic Analysis of Pharmaceuticals. Chromatographia*. **2004**, *59*, 531.
[15] A. Malenovic, D. Ivanovic, M.Medenica, B. Jancic, S. Markovic. Retention modelling in liquid chromatographic separation of simvastatin and six impurities using a microemulsion as eluent. *J. Sep. Sci.* **2004**, *27*, 1087.
[16] A. Marsh, B. J. Clark, K. D. Altria. Oil-In-Water Microemulsion LC Determination of Pharmaceuticals Using Gradient Elution. *Chromatographia* **2005**, *61*, 539.
[17] K. D. Altria, A. Marsh, B. J. Clark. High Performance Liquid Chromatographic Analysis of Pharmaceuticals Using Oil-In-Water Microemulsion Eluent and Monolithic Column. *Chromatographia* **2006**, *63*, 309.
[18] A. Marsh, B. J. Clark, K. D. Altria. Oil-In-Water Microemulsion LC Determination of Pharmaceuticals Using Gradient Elution. *Chromatographia* **2005**, *61*, 539.
[19] B. Jancic, D. Ivanovic, M. Medenica, A. Malenovic, N. Dimkovic. Development of liquid chromatographic method for fosinoprilat determination in human plasma using microemulsion as eluent. *J. Chromatogr. A* **2005**, *1088*, 187.
[20] M. Molero-Monfort, L. Escuder-Gilabert, R. M. Villanueva-Camanas, S. Sagrado, M. J. Medina-Hernandez. Biopartitioning micellar chromatography: an in vitro technique for predicting human drug absorption. *Biomed. Sci. Appl.* **2001**, *753*, 225.
[21] L. Escuder-Gilabert, M. Molero-Monfort, R. M. Villanueva-Camanas, S. Sagrado, M. J. Medina-Hernandez. Potential of biopartitioning micellar chromatography as an in vitro technique for predicting drug penetration across the blood–brain barrier. *Anal. Technol. Biomed. Life Sci.* **2004**, *807*, 193.
[22] Y. Martın-Biosca, M. Molero-Monfort, S. Sagrado, R. M. Villanueva-Camanas, M. J. Medina-Hernandez. Rapid in vitro test to predict ocular tissue permeability based on biopartitioning micellar chromatography. *Eur. J. Pharm. Sci.* **2003**, *20*, 209.
[23] J. J. Martınez-Pla, Y. Martın-Biosca, S. Sagrado, R. M. Villanueva-Camanas, M. J. Medina-Hernandez. Biopartitioning micellar chromatography to predict skin permeability. *Biomed. Chromatogr.* **2003**, *17*, 530.
[24] A. Marsh, B. J. Clark, K. D. Altria. A review of the background, operating parameters and applications of microemulsion liquid chromatography (MELC). *J. Sep. Sci.* **2005**, *28*, 2023.
[25] G. Schuurmann, R. U. Ebert, R. Kuhne. Prediction of the Sorption of Organic Compounds into Soil Organic Matter from Molecular Structure. *Environ. Sci. Technol.* **2006,** *40*, 7005.
[26] A. Talevi, M. Goodarzi, E. V. Ortiz, P. R. Duchowicz, C. L. Bellera, G. Pesce, E. A. Castro, L. E. Bruno-Blanch. Prediction of drug intestinal absorption by new linear and non-linear QSPR. *Eur. J.Med.Chem.* **2011**, *46*, 218.
[27] A. C. Lee, G.M. Crippen. Predicting pK_a. *J Chem Inf Model*, **2009**, *49*, 2013.
[28] Y. B. Liou, H. O. Ho, C. J. Yang, Y. K. Lin, M. T. Sheu. Construction of a quantitative structure-permeability relationship (QSPR) for the transdermal delivery of NSAIDs. *J. Controlled Release* **2009**, *138*, 260.
[29] A. Gajewicz, M. Haranczyk, T. Puzyn. Predicting logarithmic values of the subcooled liquid vapor pressure of halogenated persistent organic pollutants with QSPR: How different are chlorinated and brominated congeners? *Atmos. Environ*. **2010**, *44*, 1428.

[30] H. Noorizadeh, A. Farmany. QSRR models to predict retention indices of cyclic compounds of essential oils. *Chromatographia* **2010**, *72*, 563.
[31] H. Golmohammadi, M. Safdari. Quantitative structure–property relationship prediction of gas-to-chloroform partition coefficient using artificial neural network. *Microchemical J* **2010** *95*, 140.
[32] H. Noorizadeh, A. Farmany, A. Khosravi, *J. Chin. Chem. Soc.* **2010,** 57, 1.
[33] N. Krämer, A. L. Boulesteix, G. Tutz. Penalized Partial Least Squares with applications to B-spline transformations and functional data. *Chemom. Intell. Lab. Syst.* **2008**, *94,* 60.
[34] S. Haykin, *Neural Networks*, Prentice-Hall: New Jersey, **1999.**
[35] J. Liu, J. Sun, X. Sui, Y. J. Wang, Y. Hou, Z. He. Predicting blood–brain barrier penetration of drugs by microemulsion liquid chromatography with corrected retention factor. *J. Chromatogr A*, **2008**, *1198*, 164.
[36] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, *DRAGON-Software for the Calculation of Molecular Descriptors. Version 3.0 for Windows,* **2003**.
[37] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley–Longman: Reading, MA, **2000**.
[38] S. Riahi, E. Pourbasheer, R. Dinarvand, M. R. Ganjali, P. Norouzi. Exploring QSARs for Antiviral Activity of 4-Alkylamino-6-(2-hydroxyethyl)-2-methylthiopyrimidines by Support Vector Machine. *Chem. Biol. Drug Des.* **2008**, *72*, 205.
[39] H. Noorizadeh, A. Farmany. *J. Chin. Chem. Soc.* **2010**, *57,* 1.
[40] H. Noorizadeh, A. Farmany, M. Noorizadeh. Application of GA-PLS and GA-KPLS calculations for the prediction of the retention indices of essential oils. *Quim. Nova* **2011**, in press.
[41] K. Kim, J. M. Lee, I. B. Lee. A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction. *Chemom. Intell. Lab. Syst.* **2005**, *79*, 22.
[42] S. Rozou, S. Michaleas, E. Antoniadou-Vyza. Study of structural features and thermodynamic parameters, determining the chromatographic behaviour of drug–cyclodextrin complexes. *J. Chromatogr. A*, **2005**, *1087*, 86.
[43] M. Meloun, S. Bordovska, A. Vrana. The thermodynamic dissociation constants of the anticancer drugs camptothecine, 7-ethyl-10-hydroxycamptothecine, 10-hydroxycamptothecine and 7-ethylcamptothecine by the least-squares nonlinear regression of multiwavelength spectrophotometric pH-titration data. *Anal. Chim. Acta* **2007**, *584*, 419.
[44] J. P. Barbara. Factors affecting drug absorption and distribution. *Anaesth. Intensive CareMed*. 2005, 6, 135.
[45] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3.
[46] R. Todeschini, V. Consonni. Drug. *Handbook of Molecular Descriptors*, Wiley-VCH: Weinheim, Germany, **2000**.