Received: 16 January 2011 Revised: 16 February 2011 Accepted: 19 February 2011 Published online in Wiley Online Library: 16 April 2011

(www.drugtestinganalysis.com) DOI 10.1002/dta.279

pK_a modelling and prediction of drug molecules through GA-KPLS and L-M ANN

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

Genetic algorithm and partial least square (GA-PLS), kernel PLS (GA-KPLS) and Levenberg- Marquardt artificial neural network (L-M ANN) techniques were used to investigate the correlation between dissociation constant (pK_a) and descriptors for 60 drug compounds. The applied internal (leave-group-out cross validation (LGO-CV)) and external (test set) validation methods were used for the predictive power of models. Descriptors of GA-KPLS model were selected as inputs in L-M ANN model. The results indicate that L-M ANN can be used as an alternative modeling tool for quantitative structure–property relationship (QSPR) studies. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: drug molecules; immobilized liposome chromatography; QSPR; genetic algorithm-kernel partial least squares; Levenberg-Marquardt artificial neural network.

Introduction

Liposomes possess a lipid bilayer structure which simulates biological membranes, and therefore they are frequently used as models to study interactions between membranes and biological molecules, such as proteins, peptides, and drugs. For chromatographic analysis, liposomes and biomembranes have been immobilized in gel beads by steric entrapment,[1] hydrophobic binding,[2] avidin–biotin affinity binding,[3] or covalent attachment.[4] Immobilized liposome chromatography (ILC), developed in recent years, is regarded as a powerful tool to study drug–membrane interactions *in vitro*.[5,6] Liposome [7–9] formed by phosphatidylcholine, the main component found in cell membrane, or unilaminar phospholipids, [10–13] were non-covalently or covalently immobilized on soft gel particles or silica particles as chromatographic stationary phase to probe the penetration ability of compounds through biological membranes, which has been considered as one of most important parameters to evaluate their bioactivity.[14–16] In comparison to unilaminar phospholipids stationary phase, an immobilized liposome is structurally more similar to biological membranes because of its lipid bilayer structure and better fluidity of lipid molecules. Variations in retention volume depend on the extent of solute–membrane interaction and can be precisely measured. However, it is often difficult to detect weak interactions of proteins and peptides with membranes from retention volume using ILC analysis. These interactants have relatively small molecular mass and thus show little conformational change under stress conditions. ILC is a convenient method, since the membranes are stable for months and the analysis is quite rapid, which enables extensive substance and parameter screening. The reproducibility and precision of the data is also high [17–19] and competes favourably with other methods used for membrane interaction studies.

The pK_a or dissociation constant is a measure of the strength of an acid or a base. The pK_a allows you to determine the charge on a molecule at any given pH. pK_a measurement is useful parameter for use in understanding the behaviours of drug molecules. Different ionic species of a molecule differ in physical, chemical, and biological properties and so it is important to be able to predict which ionic form of the molecule is present at the site of action. The Partition Coefficient is a very useful parameter which may be used in combination with the pK_a to predict the distribution of a drug compound in a biological system. Factors such as absorption, excretion, and penetration may be related to the Log P and pK_a values of drug and in certain cases predictions can be made. Drug absorption is determined by physicochemical properties of drugs, their formulations, and routes of administration.[20–23]

Using chemometrics tools to predict drugs and chemical tissue distribution, membrane permeability or biphasic system partition is of major importance in physicochemical, environmental, and life sciences. Chemical distribution phenomena depend not only on molecular structure but also on the properties of the system in question.[24] Quantitative structure–property relationship (QSPR) techniques based on different molecular descriptors have been successfully used to model organic chemicals properties.[25]

Computational drug design is a rapidly growing field and an important component of the medicinal chemistry discipline. It is aimed at shortening the drug discovery process which otherwise may be long and expensive. Computationally determined retention parameters have become crucial in identifying potential drug candidates, and this technique is used in lead and clinical candidate optimization as well as in the selection of new compounds for screening. High throughput screening of large combinatorial libraries has increased pressure to obtain drug pharmacokinetic and metabolism data as early as possible. A number of reports that deal with QSPR calculation of several compounds have been published in the literature.[26–28]

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

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

a *Faculty of Science, Islamic Azad University, Ilam Branch, Ilam, Iran*

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

between the property of molecules and structures, non-linear models are also used to model the structure–property relationships. Levenberg-Marquardt artificial neural network (L-M ANN) is a non-parametric non-linear modelling technique that is attracting increasing interest. In the recent years, non-linear kernel-based algorithms such 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 non-linear mapping and then to develop a linear PLS model in the mapped space. According to Cover's theorem, non-linear data structure in the original space is most likely to be linear after high-dimensional non-linear mapping.[34] Therefore, KPLS can efficiently compute latent variables in the feature space by means of integral operators and non-linear kernel functions. Compared to other non-linear methods, the main advantage of the kernel-based algorithm is that it does not involve non-linear 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 non-linearities. 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 pK_a. The present study is a first research on QSPR of the drug molecules against the pK_a, using GA-KPLS.

Theory

Data set

In the current research, the data set was taken from the reference.[35] These data belong to 60 drug compounds. The drugs comprise a homologous series of compounds such as β-adrenoceptor blockers, local aesthetics, and steroids as well as a set of chemically diverse compounds. A complete list of drug names and their corresponding experimental pK_a is given in Table 1. The pK_a of these compounds was decreased in the range of 12 and 1.45 for both desmethyldiazepam and antipyrine, respectively.

Descriptor calculation

All structures of the compounds were drawn with the HyperChem 6.0 program. The pre-optimization of all molecules was performed using MM+ molecular mechanics force field. A more precise optimization was done with the semi empirical AM1 method in HyperChem. The molecular structures were optimized using the Fletcher-Reeves algorithm until the root mean square gradient was 0.01, since the calculated values of the quantum chemical features of molecules will be influenced by the related conformation. In the current research, an attempt was made to use the most stable conformations. Some quantum chemical descriptors such as orbital energies of LUMO (lowest unoccupied molecular orbital) and HOMO (high occupied molecular orbital) were calculated by using the HyperChem program. The output files were transferred into the DRAGON 3.0 program to calculate 1497 molecular descriptors.[36]

Genetic algorithm

A detailed description of the genetic algorithm (GA) can be found in the literature.[37–39] GA comprises 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 consists 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.

Software and programs

A Pentium IV personal computer (CPU at 3.06 GHz) with windows XP operational system was used. Geometry optimization was performed by HyperChem (Version 7.0 Hypercube, Inc.); Dragon software was used to calculate of descriptors. MINITAB software (version 14, MINITAB) was used for the simple PLS analysis. Cross validation, GA-PLS, GA-KPLS, L-M ANN and other calculation were performed in the MATLAB (Version 7, Mathworks, Inc.) environment.

Model validation

Validation is a crucial aspect of any QSPR/QSRR modelling.[41] The accuracy of the proposed models was illustrated using the evaluation techniques such as leave-group-out cross validation (LGO-CV) procedure and validation through an external test set.

Cross validation is a popular technique used to explore the reliability of statistical models.[42] In particular, the LGO procedure was utilized in this study. The statistical significance of the screened model was judged by the correlation coefficient (Q^2). The predictive ability was evaluated by the cross validation coefficient (Q^2 or $R^2{}_{cv}$) which is based on the prediction error sum of squares (PRESS) and was calculated by following equation:

$$R^2_{cv} \equiv Q^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i^{\wedge})^2}{\sum_{i=1}^{n}(y_i - y^{-})^2} \qquad (1)$$

where y_i, $y_i^{\wedge}$ and y^{-} were respectively the experimental, predicted, and mean pK_a values of the samples. The accuracy of cross validation results is extensively accepted in the literature considering the Q^2 value. In this sense, a high value of the statistical characteristic ($Q^2 > 0.5$) is considered proof of the high predictive ability of the model.[43]However, this assumption is in many cases incorrect and can be that exist the lack of the correlation between the high Q^2 and the high predictive ability of QSPR/QSRR models has been established and corroborated recently. [41] Thus, the high value of Q^2 appears to be a necessary but not sufficient condition for the models to have a high predictive power. As a next step, further analysis was also followed for chemical property of the new set of compounds using the developed QSPR model.

Validation through the external test set

Validating QSPR with external data (i.e. data not used in the model development) is the best method of validation. However, the availability of an independent external test set of several compounds is rare in QSPR. Thus, the predictive ability of a QSPR model with the selected descriptors was further explored by

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

No.	Name	pK$_{a\ Exp}$	pK$_{a\ Cal}$	RE	AbsE
Training set					
1	Antipyrine	1.45	1.37	5.52	0.11
2	Oxazepam	1.7	1.6	5.88	0.10
3	Proxicromil	1.93	2.08	7.77	0.09
4	Tetracaine	2.4	2.1	12.50	0.09
5	Olsalazine	2.85	2.72	4.56	0.08
6	Salicylic acid	2.97	2.76	7.07	0.08
7	Diflunisal	3	3	0.00	0.07
8	Diazepam	3.3	3.3	0.00	0.07
9	Tolmetin	3.5	3.8	8.57	0.06
10	Flufenamic acid	3.9	4.2	7.69	0.05
11	Naproxen	4.2	4.1	2.38	0.05
12	Flurbiprofen	4.27	4.1	3.98	0.05
13	Omeprazole	4.4	4.6	4.55	0.04
14	Diclofenac	4.5	4.7	4.44	0.04
15	Sulindac	4.5	4.8	6.67	0.04
16	Fenbufen	4.51	4.91	8.87	0.03
17	Ketoprofen	4.6	4.9	6.52	0.03
18	Gemfibrozil	4.7	4.6	2.13	0.04
19	d-DAVP	4.8	4.4	8.33	0.05
20	Warfarin	4.9	4.7	4.08	0.04
21	Ibuprofen	5.2	4.8	7.69	0.04
22	Piroxicam	5.46	5.84	6.96	0.02
23	Enalapril	5.5	5.7	3.64	0.02
24	Indoprofen	5.8	6.1	5.17	0.01
25	Cephalexin	7.3	6.7	8.22	0.00
26	Amoxicillin	7.4	6.8	8.11	0.01
27	Inogatran	7.6	7.1	6.58	0.01
28	Prilocaine	7.9	7.5	5.06	0.02
29	Bupivacaine	8.1	8.3	2.47	0.04
30	Phenytoin	8.3	8.6	3.61	0.04
31	5-Hydroxyquinoline	8.56	8.55	0.12	0.04
32	Theophylline	8.6	8.7	1.16	0.04
33	Pindolol	8.8	9.0	2.27	0.05
34	Verapamil	8.92	9.24	3.59	0.06
35	Amlodipine	9.02	9.47	4.99	0.06
36	Promethazine	9.1	9	1.10	0.05
37	Sulpiride	9.12	8.95	1.86	0.05
38	Acebutolol	9.2	8.7	5.43	0.04
39	Nadolol	9.39	9.02	3.94	0.05
40	Propranolol	9.45	9.2	2.65	0.06
41	Oxprenolol	9.5	9.8	3.16	0.07
42	Atenolol	9.6	10.1	5.21	0.07
43	Alprenolol	9.65	9.71	0.62	0.07
44	Metoprolol	9.7	10.2	5.15	0.08
45	4-Phenylbutyl amine	10.42	10.03	3.74	0.07
46	Terbutaline	11.2	10.8	3.57	0.09
47	Sulphasalazine	11.8	10.7	9.32	0.09
48	Desmethyldiazepam	12	13	8.33	0.13
Test Set					
49	Procaine	2.3	2.5	8.70	0.34
50	Acetylsalicylic acid	3.5	3.1	11.43	0.29
51	Tolfenamic acid	4.2	3.9	7.14	0.22
52	Indomethacin	4.5	4.0	11.11	0.21
53	5-Phenylvaleric acid	4.88	5.39	10.45	0.10
54	Ciprofloxacin	6	5	16.67	0.13
55	Lidocaine	7.9	6.8	13.92	0.02

Table 1. (*Continued*)

No.	Name	$pK_{a\ Exp}$	$pK_{a\ Cal}$	RE	AbsE
56	Tramadol	8.3	7.5	9.64	0.08
57	Loperamide	8.7	8.1	6.90	0.13
58	Salmeterol	9.3	9.8	5.38	0.27
59	Practolol	9.5	10.8	13.68	0.35
60	Metolazone	9.7	9.0	7.22	0.20

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

dividing the full data set. The predictive power of the models developed on the selected training set is estimated on the predicted values of test set chemicals. The data set was randomly divided into two groups including training set (calibration and prediction sets) and test set, which consists of 48 and 12 molecules, respectively. The calibration set was used for model generation. The prediction set was applied to deal with over fitting of the network, whereas the test set, whose molecules have no role in model building, was used for the evaluation of the predictive ability of the models for external set.

Results and discussion

Linear model

Results of 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 model was selected on the basis of the highest multiple correlation coefficient (LGO-CV) (Q^2), the least root mean squares error (RMSE), standard error (SE), and relative error (RE) of prediction and simplicity of the model. These parameters are probably the most popular measure of how well a model fits the data. The best GA-PLS model contained 8 selected descriptors in 3 latent variables space. These descriptors were obtained: constitutional descriptors (number of Hydrogen atoms (nH) and number of bonds (nBT)), RDF descriptors (Radial Distribution Function – 4.1/weighted by atomic Sanderson electronegativities (RDF041e)), molecular properties (TPSA (NO)), charge descriptors (relative positive charge (RPCG)), atom-centred fragments (phenol/enol/carboxyl OH (O-057)), and quantum chemical descriptors (HOMO and LUMO). For this in general, the number of components (latent variables) is less than the number of independent variables in PLS analysis. The obtained statistic parameters of the GA-PLS model are shown in Table 3. The PLS model uses a higher number of descriptors that allow the model to extract better structural information from descriptors to result in a lower prediction error.

Non-linear models

Results of the GA-KPLS model

LGO-CV was 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.[44] It means that the value of c depends on the system under the study. The five descriptors in two latent variables space chosen by GA-KPLS feature selection methods were contained. These descriptors were obtained: constitutional descriptors (mean atomic van der Waals volume (scaled on Carbon atom (Mv) and sum of atomic van der Waals volumes (scaled on Carbon atom (Sv)), charge descriptors (RPCG), molecular properties ((Moriguchi octanol-water partition coeff. (logP)) (MLOGP) and quantum chemical descriptors (HOMO). The statistical parameters of this model, constructed by the selected descriptors, are depicted in Table 3. The RMSE values of the GA-KPLS model for the training and test sets were much lower than GA-PLS model. From these results, it can be noticed that the GA-KPLS model gives the highest Q^2 values, so this model provides the most satisfactory results, compared with the results obtained from the GA-PLS model. 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 pK_a of the drug compounds 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.

Results of the L-M ANN model

With the aim of improving the predictive performance of non-linear QSPR model, L-M ANN modelling was performed. Descriptors of GA-KPLS model were selected as inputs in the L-M ANN model. The network architecture consisted of five neurons in the input layer corresponding to the five mentioned descriptors. The output

Table 3. The statistical parameters of different constructed QSPR models

	Training set						Test set				
Model	R^2	Q^2	RE	RMSE	SE	N	R^2	RE	RMSE	SE	N
GA-PLS	0.761	0.750	14.07	0.98	0.33	48	0.703	28.19	3.53	0.66	12
GA-KPLS	0.921	0.917	7.64	0.76	0.17	48	0.863	15.91	1.24	0.33	12
L-M ANN	0.984	0.983	4.9	0.37	0.05	48	0.929	10.18	0.73	0.19	12

layer had 1 neuron that predicts the pK_a. 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 2 to 7, 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 is minimum when two neurons were selected in the hidden layer and the learning rate and the momentum values were 0.5 and 0.3, respectively. Finally, the number of iterations was optimized with the optimum values for the variables. It was realized that after 15 iterations, the RMSE for prediction set was minimum. The values of experimental, calculated and percent relative error are shown in Table 1. The obtained statistic parameters of the L-M ANN model are shown in Table 3. For the constructed model, five general statistical parameters were selected to evaluate the prediction ability of the model for the pK_a. The statistical parameters R^2, Q^2, SE, RE, and RMSE were obtained for the proposed models. 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 other parameter values which were lower for the training and test sets compared with their counterparts for GA-KPLS and GA-PLS. Plots of predicted pK_a versus experimental pK_a values by L-M ANN for training and test set are shown in Figure 1. Obviously, there is a close agreement between the experimental and predicted pK_a 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 non-linear feature selection method. The key strength of L-M ANN is its ability to allow for flexible mapping of the selected features by manipulating its functional dependence implicitly. Neural network handles both linear and non-linear relationships 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

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 ionizable drug across a membrane at equilibrium is determined by the drug's pK_a and the pH gradient, when present.[45–47]

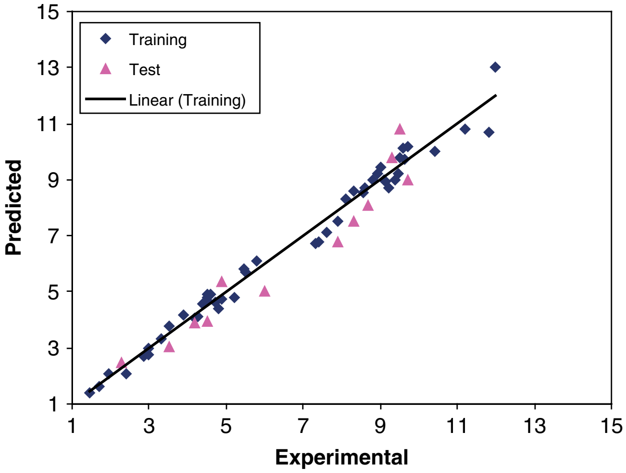

Figure 1. Plot of predicted pK_a obtained by L-M ANN against the experimental values.

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.

Log P as a molecular properties descriptor estimates the propensity of a neutral compound to differentially dissolve in two immiscible phases. Nowadays, log P is commonly used in QSAR/QSPR studies and drug design since it relates to drug absorption, bioavailability, metabolism, and toxicity.

The radial distribution function of an ensemble of atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of certain radius, employing different atomic properties such as atomic polarizabilities, volumes, masses, or atomic electro negativities in order to differentiate the contribution of atoms to the property being analyzed.

Although constitutional descriptors, functional group and atom-centred fragments are often successful in pK_a 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 or HOMO energy). Quantum chemical descriptors are 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. The eigenvalues of LUMO and HOMO and their energy gap reflect the chemical activity of the molecule. LUMO as an electron acceptor represents the ability to obtain an electron, while HOMO as an electron donor represents the ability to donate an electron. The HOMO energy plays a very important role in the nucleophylic behaviour and it represents molecular reactivity as a nucleophile. Good nucleophiles are those where the electron residue is high lying orbital. The energy of the LUMO is directly related to the electron affinity and characterizes the susceptibility of the molecule towards 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.

Charge 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. Charge-based descriptors have been widely employed as chemical reactivity indices or as measures of weak intermolecular interactions. Many quantum chemical descriptors are derived from the partial charge distribution in a molecule or from the electron densities on particular atoms. Relative positive charge (RPCG) is the quotient between maximum atomic positive charge in the molecule and positive atomic charge in the molecule. It contains electronic information to describe the molecule, and therefore it encodes features responsible for interaction between molecules and the modified reversed stationary phase.[48]

Conclusion

In this research, an accurate QSPR model for estimating the pK_a was developed by employing the one linear model (GA-PLS) and two non-linear 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 modeling. This indicates that pK_a of drug molecules possesses some non-linear 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 pK_a for new compounds or for other compounds whose experimental values are unknown.

References

[1] Q. Yang, P. Lundahl. Steric immobilization of liposomes in chromatographic gel beads and incorporation of integral membrane proteins into their lipid bilayers. *Anal. Biochem.* **1994**, *218*, 210.

[2] Y. Zhang, C.-M. Zeng, Y.-M. Li, S. Hjerten, P. Lundahl. Immobilized liposome chromatography of drugs on capillary continuous beds for model analysis of drug-membrane interactions. *J. Chromatogr. A* **1996**, *749*, 13.

[3] Q. Yang, X.-Y. Liu, S.-I. Ajiki, M. Hara, P. Lundahl, J. Miyake. Avidin-biotin immobilization of unilamellar liposomes in gel beads for chromatographic analysis of drug-membrane partitioning. *J. Chromatogr. B* **1998**, *707*, 131.

[4] Q. Yang, X.-Y. Liu, M. Yoshimoto, R. Kuboi, J. Miyake. Covalent immobilization of unilamellar liposomes in gel beads for chromatography. *Anal. Biochem.* **1999**, *268*, 354.

[5] Y. Wang, L. Kong, X. Lei, L. Hu, H. Zou, E. W. Beck, *et al*. Comprehensive two-dimensional high-performance liquid chromatography system with immobilized liposome chromatography column and reversed-phase column for separation of complex traditional Chinese medicine Longdan Xiegan Decoction. *J. Chromatogr. A* **2009**, *1216*, 2185.

[6] C. Huang, J. T. Mason. Geometric packing constraints in egg phosphatidylcholine vesicles. *Proc. Natl. Acad. Sci.* **1978**, *75*, 308.

[7] S. Ong, H. Liu, C. Pidgeon. Immobilized-artificial-membrane chromatography: measurements of membrane partition coefficient and predicting drug membrane permeability. *J. Chromatogr. A* **1996**, *728*, 113.

[8] P. Lundahl, F. Beigi. Immobilized liposome chromatography of drugs for model analysis of drug-membrane interactions. *Adv. Drug Deliv. Rev.* **1997**, *23*, 221.

[9] X. Liu, Q. Yang, N. Kamo, J. J. Miyake. Effect of liposome type and membrane fluidity on drug-membrane partitioning analyzed by immobilized liposome chromatography. *J. Chromatogr. A* **2001**, *913*, 123.

[10] C. Pidgeon, S. Ong, H. Chol, H. Liu. Preparation of mixed ligand immobilized artificial membranes for predicting drug binding to membranes. *Anal. Chem.* **1994**, *66*, 2701.

[11] C. Pidgeon, S. Ong. Predicting drug-membrane interactions. *Chemtech* **1995**, *25*, 38.

[12] S. Ong, H. Liu, X. Qiu, C. Pidgeon. Membrane partition coefficients chromatographically measured using immobilized artificial membrane surfaces. *Anal. Chem.* **1995**, *67*, 755.

[13] C. Y. Yang, S. J. Cai, H. Liu, C. Pidgeon. Immobilized Artificial Membranes — screens for drug membrane interactions. *Adv. Drug Deliv. Rev.* **1996**, *23*, 229.

[14] P. Artursson, J. Karlsson. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem. Bioph. Res. Co.* **1991**, *175*, 880.

[15] C. Altomare, R. Tsai, N. E. Tayar, B. Testa, A. Carotti, S. Cellamare, *et al*. Determination of lipophilicity and hydrogen-bond donor acidity of bioactive sulphonyl-containing compounds by reversed-phase HPLC and centrifugal partition chromatography and their application to structure-activity relations. *J. Pharm. Pharmacol.* **1991**, *43*, 191.

[16] P. Artursson. Epithelial transport of drugs in cell culture. I: A model for studying the passive diffusion of drugs over intestinal absorbtive (Caco-2) cells. *J. Pharm. Sci.* **1990**, *79*, 476.

[17] F. Beigi, I. Gottschalk, C. Lagerquist Hägglund, L. Haneskog, E. Brekkan, Y. Zhang, *et al*. Immobilized liposome and biomembrane partitioning chromatography of drugs for prediction of drug transport. *J. Pharm.* **1998**, *164*, 129.

[18] C. Lagerquist, F. Beigi, A. Karlén, H. Lennernäs, P. Lundahl. Effects of cholesterol and model transmembrane proteins on drug partitioning into lipid bilayers as analysed by immobilized-liposome chromatography. *J. Pharm. Pharmacol.* **2001**, *53*, 1477.

[19] E. Boija, A. Lundquist, J. J. Martínez Pla, C. Engvall, P. Lundahl. Effects of ions and detergents in drug partition chromatography on liposomes. *J. Chromatogr. A* **2004**, *1030*, 273.

[20] A. Avdeef. pH-Metric log P. Part 1. Difference Plots for Determining Ion-Pair Octanol-Water Partition Coefficients of Multiprotic Substances. *Quant. Struct-Act. Rel.* **1992**, *11*, 510.

[21] A. Avdeef, J. E. A. Comer, S. Thomson. pH-Metric log P. 3. Glass electrode calibration in methanol-water, applied to pKa determination of water-insoluble substances. *J. Anal. Chem.* **1993**, *65*, 42.

[22] E. Ornskov, A. Linusson, S. Folestad. Determination of dissociation constants of labile drug compounds by capillary electrophoresis. *J. Pharmaceut. Biomed. Anal.* **2003**, *33*, 379.

[23] M. C. Levesque, D. K. Ghosh, B. E. Beasley, Y. Chen, A. D. Volkheimer, C. W. O'Loughlin, *et al*. P102. Induction of chronic lymphocytic leukemia (CLL) apoptosis by nitric oxide synthase (NOS) inhibitors: Drug efficacy correlates with lipid solubility and NOS1 dissociation constant. *Nitric Oxide* **2006**, *14*, 49.

[24] G. Klopman, H. Zhu. Recent methodologies for the estimation of n-octanol/water partition coefficients and their use in the prediction of membrane transport properties of drugs. *Mini Rev. Med. Chem.* **2005**, *5*, 127.

[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, *et al*. 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 pKa. *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. Control. 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. *Microchem. J.* **2010**, *95*, 140.

[32] H. Noorizadeh, A. Farmany, A. Khosravi. Investigation of retention behaviors of essential oils by using QSRR. *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. *Chemometr. Intell. Lab.* **2008**, *94*, 60.

[34] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd edn. Prentice-Hall: Upper Saddle River, NJ, USA, **1999**.

[35] T. Osterberg, M. Svensson, P. Lundahl. Chromatographic retention of drug molecules on immobilised liposomes prepared from egg phospholipids and from chemically pure phospholipids. *Eur. J. Pharm. Sci.* **2001**, *12*, 427.

[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, USA, **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. Exploration of Linear and Nonlinear Modeling Techniques to Predict of Retention Index of Essential Oils. *J. Chin. Chem. Soc.* **2010**, *57*, 1.

[40] U. Depczynski, V. J. Frost, K. Molt. Genetic algorithms applied to the selection of factors in principal component regression. *Anal. Chim. Acta* **2000**, *420*, 217.

[41] J. Acevedo-Martınez, J. C. Escalona-Arranz, A. Villar-Rojas, F. Tellez-Palmero, R. Perez-Roses, L. Gonzalez, *et al*. Quantitative study of the structure-retention index relationship in the imine family. *J. Chromatogr. A* **2006**, *1102*, 238.

[42] A. Af1ntitis, G. Melagraki, H. Sarimveis, P. A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou. A novel QSAR model for predicting induction of apoptosis by 4-aryl-4*H*-chromenes. *Bioorg. Med. Chem.* **2006**, *14*, 6686.

[43] A. Golbraikh, A. Tropsha. Beware of q2! *J. Mol. Graph. Model.* **2002**, *20*, 269.

[44] K. Kim, J. M. Lee, I. B. Lee. A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction. *Chemometr. Intell. Lab.* **2005**, *79*, 22.

[45] 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.

[46] 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.

[47] J. P. Barbara. Factors affecting drug absorption and distribution. *Anaesth. Intens. Car. Med.* **2005**, *6*, 135.

[48] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH: Weinheim, Germany, **2000**.