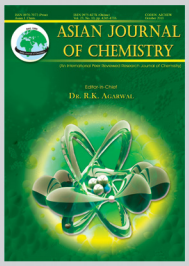
Asian Journal of Chemistry; Vol. 24, No. 1 (2012), 203-208

ASIAN JOURNAL OF CHEMISTRY

www.asianjournalofchemistry.co.in



Study of the Retention Time of Nanoparticle Compounds by Quantitative Structure Retention Relationship

HADI NOORIZADEH* and ABBAS FARMANI

Department of Chemistry, Faculty of Sciences, Arak Branch, Islamic Azad University, Arak, Iran

*Corresponding author: Fax: +98 841 2227051; Tel: +98 841 2227051; E-mail: hadinoorizadeh@yahoo.com; a.farmany@usa.com

(*Received*: 6 December 2010; *Accepted*: 15 September 2011) AJC-10400

The quantitative structure-retention relationship (QSRR) of nanoparticle compounds against the comprehensive 2-D gas chromatography system (GC × GC) retention time (RT) was studied. Application of the dodecanethiol monolayer-protected gold nanoparticle (MPN) column was for a high-speed separation as the second column of GC × GC. A suitable set of molecular descriptors was calculated and the genetic algorithm (GA) was employed to select those descriptors that resulted in the best-fit models. The partial least squares (PLS), the kernel PLS (KPLS) and Levenberg-Marquardt artificial neural network (L-M ANN) were utilized to construct the linear and nonlinear QSRR models. The models were validated using leave-Group-out cross validation LGO-CV. The results indicate that L-M ANN can be used as an alternative modeling tool.

Key Words: Nanoparticle compounds, QSRR.

INTRODUCTION

For decades, chromatographers have utilized nanometersized materials in the development of highly efficient chromatographic stationary phases. These materials offer a variety of advantages from improved mass transfer characteristics to greater stability of traditional polymer phases by incorporating nanoparticle additives. Nanoparticles on the order of 10 nm in diameter have also been used to stabilize polymer stationary phases for gas chromatography, similar in concept to that for support coated open tubular columns[1]. The unaccompanied use of nanoparticles as chromatographic stationary phases has been put forth. Both silica and polymer nanoparticles have been applied in electrophoretic chromatography as a stationary phase[2,3]. In high-speed separations, the need for more selective and efficient stationary phases increases as scientists push the envelope of chromatography and hence the need to focus on novel stationary phase development. The call for new stationary phases for gas chromatography is not limited to high-speed gas chromatography (HS-GC) but also complementary separation techniques such as comprehensive 2-D gas chromatography (GC × GC)[4-6] as well as miniaturized chromatographic systems[7].

The GC × GC system relies upon having a highly efficient stationary phase within a short, smaller dimensional capillary as the second column and a more traditionally sized capillary for the first column. The two columns should provide complementary chemical selectivity in order to more fully utilize the peak capacity of the 2-D separation. Often, a non-polar stationary phase is used for the first column, but since the square capillary monolayer-protected gold nanoparticle (MPN) column is more appropriately configured for the second column, a polar poly(ethylene glycol) column was used for the first column. This reversed stationary phase pairing provided an excellent comprehensive 2-D separation for the 23-component mixture in the present investigation[8].

The dodecanethiol monolayer-protected gold nanoparticles (MPNs) were used stationary phase for open tubular gas chromatography. The nanoparticles we have been studying are known as gold-centered monolayer protected nanoparticles (MPNs) that consist of a gold core with a thiol-linked monolayer of organic molecules on the surface of the gold core[9]. Monolayer-protected gold nanoparticles are of particular interest because their properties can be influenced by the structure of the monolayer forming molecules and the surface monolayer stabilizes them relative to aggregation as compared to bare gold nanoparticles[10]. The selectivity achieved for chemical sensing with a gold-centered MPN is dominated by the chemical structure and functionality selected for the organic surface layer[11]. Potential areas of application for MPNs are demonstrated including complementary separations such as

GC × GC, exploration of a model system for use in microfabricated gas chromatography systems, as well as efficient single dimension, high-speed GC separations. Microfabricated GC systems result in angular or square cornered channels in contrast to the traditional round capillary used for bench-top open tubular GC columns[12].

Quantitative structure-retention relationship studies have received much attention in chemometrics, biological chemistry, medicinal chemistry and many other fields. Quantitative structure-retention relationship models are mathematical equations relating chemical structure to their property. A number of reports, deals with QSRR retention time calculation of several compounds have been published in the literature[13,14]. The QSRR models apply to partial least squares method often combined with genetic algorithms for feature selection[15,16].

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 is nonparametric nonlinear modeling technique that has attracted increasing interest. In recent years, nonlinear kernel-based algorithms as kernel partial least squares (KPLS) have been proposed[17, 18]. The kernel partial least squares can efficiently compute latent variables in the feature space by means of integral operators and nonlinear kernel functions[19]. 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 non-linearities. In the present study, GA-PLS, GA-KPLS and L-M ANN were employed to generate QSRR models that correlate the structure of nanoparticles with observed retention time. The present study is a first approach on QSRR of the nanoparticle compounds against the retention time, using GA-PLS, GA-KPLS and L-M ANN.

EXPERIMENTAL

Data set: Retention time of 23 nanoparticle compounds were taken from the literature[8] are presented in Table-1. Sample components are separated, identified and measured by the GC × GC using the dodecanethiol MPN stationary phase within a square capillary system as the second dimension separation. A square deactivated silica capillary, 100 µm width, was used for the production of the open tubular MPN column (Polymicro Technologies, Phoenix, AZ, USA). All chromatograms were obtained with an Agilent 6890 gas chromatograph using a standard commercial FID system and injector with chem station computer control (Agilent Technologies, Palo Alto, CA, USA). The GC × GC separation was obtained using a 4 m poly(ethylene glycol) column (200 µm i.d., 0.2 µm film) as the first column at 34,000 Pa (40 cm/s) with 0.9 m of the dodecanethiol MPN 100 µm square capillary column as the second column operated at 210,000 Pa (235 cm/s). The oven was held constant at 60 ºC with the FID and inlet temperatures at 250 ºC. A 0.5 µL injection was introduced with a 150:1 split on the inlet. The valve injection onto the second column had a 15 ms wide injection pulse width, a 1.3 µL loop with a 1s modulation period. In this study, retention data was resulted by non polar capillary [poly(ethylene glycol) column] used for QSRR research. The retention time of these compounds was increased in the range of 12.5 and 98.5 for both hexane and 1,2,4-trimethylbenzene, respectively.

TABLE-1
DATA SET AND THE CORRESPONDING OBSERVED AND PREDICTED RT VALUES BY L-M ANN

Name	RT_{Exp}	RT_{Cal}	RE	SE
Hexane	12.5	11.85	5.20	1.49
Octane	15	15.46	3.07	1.33
Chlorobutane	16.1	16	0.62	1.31
Nonane	18	19.37	7.61	1.16
1-Heptyne	19.7	17.38	11.78	1.25
Benzene	20.2	20.67	2.33	1.11
2-Pentanone	22	22.01	0.05	1.05
Decane	25	22.47	10.12	1.03
2-Butanol	27	28.61	5.96	0.76
Toluene	28.5	28.94	1.54	0.75
1-Propanol	29.1	29.19	0.31	0.74
2-Hexanone	33.2	30.14	9.22	0.70
Ethylbenzene	42.2	45.34	7.44	0.03
p-Xylene	44.1	42.51	3.61	0.16
1-Butanol	45.3	45.01	0.64	0.05
Chlorobenzene	64.2	58.34	9.13	0.53
4-Ethyltoluene	69	71.24	3.25	1.09
Mesitylene	75.5	82.14	8.79	1.57
1-Pentanol	84.5	80.24	5.04	1.48
1,1,2-Trichloroethane	88.5	88.42	0.09	1.84
3-Octanone	90.5	91.67	1.29	1.98
Bromoheptane	92.3	93.16	0.93	2.05
1,2,4-Trimethylbenzene	98.5	95.07	3.48	2.13

RT = Retention time

Computer hardware and software: All calculations were run on a HP laptop computer with AMD turion64X2 processor with windows XP operating system. The optimizations of molecular structures were done by the HyperChem 7.0 (AM1 method) and descriptors were calculated by Dragon version 3.0 software's. MINITAB software version 14 was used for the simple PLS analysis. Cross validation, GA-PLS, GA-KPLS and other calculation were performed in the MATLAB (version 7, mathworks, Inc.) environment.

Genetic algorithm: A detailed description of the genetic algorithm can be found in the literature[20-22]. Genetic algorithm is simulated methods based on ideas from Darwin's theory of natural selection and evolution (the struggle for life). In genetic algorithm a chromosome (or an individual) can be defined as an enciphered entity of a candidate solution, which is expressed as a set of variables. Genetic algorithm 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.*[23]. The parameter algorithm reported in Table-2.

Non-linear models

Kernel partial least squares: The KPLS method is based on the mapping of the original input data into a high dimensional

feature space $\Im$ where a linear PLS model is created. By nonlinear mapping Φ: $x \in \Re^n \rightarrow \Phi(x) \in \Im$, a KPLS algorithm can be derived from a sequence of NIPALS steps and has the following formulation[24]: (1) Initialize score vector w as equal to any column of Y; (2) Calculate scores $u = \Phi\Phi^T w$ and normalize u to $\|u\| = 1$, where Φ is a matrix of regressors; (3) Regress columns of Y on u: $c = Y^T u$, where c is a weight vector; (4) Calculate a new score vector w for Y: $w = Yc$ and then normalize w to $\|w\|=1$; (5) Repeat steps 2-4 until convergence of w; 6. Deflate $\Phi\Phi^T$ and Y matrices:

$$\Phi\Phi^T = (\Phi - uu^T\Phi)(\Phi - uu^T\Phi)^T \quad (1)$$

$$Y = Y - uu^T Y \quad (2)$$

(7) Go to step 1 to calculate the next latent variable.

Without explicitly mapping into the high-dimensional feature space, a kernel function can be used to compute the dot products as follows:

$$k(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) \quad (3)$$

$\Phi\Phi^T$ represents the $(n \times n)$ kernel Gram matrix K of the cross dot products between all mapped input data points $\Phi(x_i)$, i=1,....,n. The deflation of the $\Phi\Phi^T = K$ matrix after extraction of the u components is given by:

$$K = (I - uuT)K(I - uu^T) \quad (4)$$

where I is an m-dimensional identity matrix. Taking into account the normalized scores u of the prediction of KPLS model on training data $\hat{Y}$ is defined as:

$$\hat{Y} = KW(U^TKW)^{-1}U^TY = UU^TY \quad (5)$$

For predictions on new observation data $\hat{Y}$, the regression can be written as:

$$\hat{Y}_t = K_t W(U^TKW)^{-1}U^TY \quad (6)$$

where, K_t is the test matrix whose elements are $K_{ij} = K(x_i, x_j)$ where x_i and x_j present the test and training data points, respectively.

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-one-out
Number subset	24
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

Artificial neural network: An artificial neural network (ANN) with a layered structure is a mathematical system that stimulates the biological neural network consist of computing units named neurons and connections between neurons named synapses[25-27]. Input or independent variables are considered as neurons of input layer, while dependent or output variables are considered as output neurons. Synapses connect input neurons to hidden neurons and hidden neurons to output neurons. The strength of the synapse from neuron i to neuron j is determined by mean of a weight, W_{ij}. In addition, each neuron j from the hidden layer and eventually the output neuron, are associated with a real value b_j, named the neuron's bias and with a nonlinear function, named the transfer or activation function. Because the artificial neural networks (ANNs) are not restricted to linear correlations, they can be used for nonlinear phenomena or curved manifolds[25]. Back propagation neural networks (BNNs) are most often used in analytical applications[26]. The back propagation network receives a set of inputs, which is multiplied by each node and then a nonlinear transfer function is applied. The goal of training the network is to change the weight between the layers in a direction to minimize the output errors. The changes in values of weights can be obtained using eqn. (7):

$$\Delta W_{ij,n} = F_n + \alpha \Delta W_{ij,n-1} \quad (7)$$

where ΔW_{ij} is the change in the weight factor for each network node, α is the momentum factor and F is a weight update function, which indicates how weights are changed during the learning process. There is no single best weight update function which can be applied to all nonlinear optimizations. One need to choose a weight update function based on the characteristics of the problem and the data set of interest. Various types of algorithms have been found to be effective for most practical purposes such as Levenberg-Marquardt (L-M) algorithm.

Levenberg-Marquardt algorithm: While basic back propagation is the steepest descent algorithm, the Levenberg-Marquardt algorithm[28] is an alternative to the conjugate methods for second derivative optimization. In this algorithm, the update function, F_n, can be calculated using eqn. (8) and (9):

$$F_0 = -g_0 \quad (8)$$

$$F_n = -[J^T \times J + \mu I]^{-1} \times J^T \times e \quad (9)$$

where, J is the Jacobian matrix, μ is a constant, I is an identity matrix and e is an error function[29].

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, 1104 descriptors were remained. These descriptors were employed to generate the models with the GA-PLS and GA-KPLS program. The best model is selected on the basis of the highest multiple correlation coefficient leave-group-out cross validation (LGO-CV) (Q^2), the least root mean squares error (RMSE), standard error (SE), absolute error (AbsE) 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 contains 5 selected descriptors in 2 latent variables space. These descriptors were obtained constitutional descriptors (sum of atomic van der Waals volumes [scaled on carbon atom) (Sv)], functional group (number of donor atoms for H-bonds (N and O) (nHDon), 3D-MoRSE descriptors (signal 1/weighted by atomic van der Waals volumes (Mor1v), atom-centred fragments [H attached to $C_1(sp^3)/C_0(sp^2)$ (H-047) and quantum chemical

descriptors (dipole moment (μ)]. For this in general, the number of components (latent variables) is less than number of independent variables in PLS analysis. The obtained statistic parameters of the GA-PLS model were shown in Table-3. 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
STATISTICAL PARAMETERS OF DIFFERENT CONSTRUCTED QSRR MODELS

Model	R^2	Q^2	RE	AbsE	RMSE	SE	N
GA-PLS	0.864	0.864	10.48	3.94	5.23	3.82	23
GA-KPLS	0.885	0.881	9.13	3.55	4.67	3.20	23
L-M ANN	0.984	0.980	4.41	1.85	2.59	1.11	23

Non-linear models

Results of the GA-KPLS model: 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 set to be 1), m is the dimension of the input space and σ^2 is the variance of the data[30]. It means that the value of c depends on the system under the study. The 4 descriptors in 1 latent variables space chosen by GA-KPLS feature selection methods were contained. These descriptors were obtained constitutional descriptors [number of carbon atoms (nC)], WHIM descriptors (A total size index / weighted by atomic van der Waals volumes) (Av), atom-centred fragments (H attached to C_1 (sp^3)/C_0 (sp^2) (H-047)) and quantum chemical descriptors [high occupied molecular orbital (HOMO)]. Table-3 shows the statistical parameters of the results, attained by these models studies for the same set of nanoparticle compounds. The RMSE values of the GA-KPLS model was much lower than GA-PLS model. From this table, it can be noticed that the GA-KPLS model gives the highest square correlation coefficient (R^2) values, so this model provides the most satisfactory results, compared with the results obtained from the GA-PLS model. The data presented in Table-3 indicate that the GA-PLS linear model have 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 retention time of the nannoparticle 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. This result indicates that the retention time of nanoparticle compounds possesses some non-linear characteristics.

Results of the L-M ANN model: With the aim of improving the predictive performance of nonlinear QSRR model, L-M ANN modeling was performed. Descriptors of GA-KPLS model were selected as inputs in L-M ANN model. The network architecture consisted of four neurons in the input layer corresponding to the four mentioned descriptors. The output layer had one neuron that predicts the retention time. 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 was calculated for all of the possible combination of values for the mentioned variables in leave-group-out cross validation (LGO-CV). It was realized that the RMSE is minimum when one neuron were selected in the hidden layer and the learning rate and the momentum values were 0.4 and 0.2, respectively. Finally, the number of iterations was optimized with the optimum values for the variables. It was realized that after 15 iterations, the RMSE was minimum. The values of experimental, calculated, per cent relative error and absolute error are shown in Table-1. For the constructed model, six general statistical parameters were selected to evaluate the prediction ability of the model for the retention time. Table-3 shows the statistical parameters for the compounds obtained by applying models. The R^2, Q^2, RMSE, RE, SE, AbsE and RE were obtained for proposed models. Each of the statistical parameters mentioned above were used for assessing the statistical significance of the QSRR model. The statistical parameters obtained by LGO-CV for L-M ANN, GA-KPLS and the linear QSRR model are compared in Table-3. Inspection of the results of the table reveals a higher R^2 and Q^2 values and lower RMSE and RE for L-M ANN model compared with their counterparts for GA-KPLS and linear models. Plots of predicted retention time *versus* experimental retention time values by L-M ANN are shown in Fig. 1. Obviously, there is a close agreement between the experimental and predicted retention time 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.

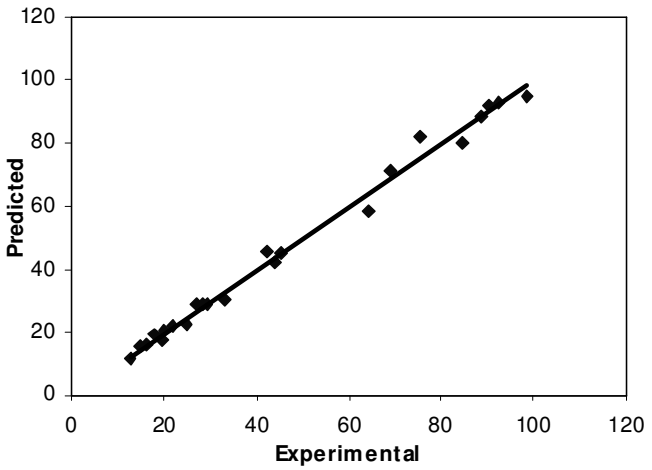


Fig. 1. Plot of predicted retention time obtained by L-M ANN against the experimental values

Model validation: Validation is a crucial aspect of any QSPR/QSRR modeling[31-33]. The accuracy of proposed models was illustrated using the evaluation techniques such as leave-group-out cross validation (LGO-CV) and leave-One-out cross

validation (LOO-CV). In this study, we have used LGO-CV for GA-PLS and GA-KPS models and LOO-CV for L-M ANN model.

Cross validation technique: Cross validation is a popular technique used to explore the reliability of statistical models. Based on this technique, a number of modified data sets are created by deleting in each case one or a small group (leave-some-out) of objects. For each data set, an input-output model is developed, based on the utilized modeling technique. Each model is evaluated, by measuring its accuracy in predicting the responses of the remaining data (the ones or group data that have not been utilized in the development of the model) [34,35]. A QSRR model was then constructed on the basis of this reduced data set and subsequently used to predict the removed data. This procedure was repeated until a complete set of predicted was obtained. 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:

$$R^2_{cv} \equiv Q^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - y^-\right)^2} \quad (10)$$

where y_i, $\hat{y}_i$ and y^- were the experimental, predicted and mean retention time values of the samples, respectively. 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 as proof of the high predictive ability of the model[36-38]. The result clearly displays a significant improvement of the QSRR model consequent to non-linear statistical treatment and a substantial independence of model prediction from the structure of the test molecule. In the above analysis, the descriptive power of a given model has been measured by its ability to predict retention of unknown nanoparticles. For instance, as to prediction ability, it can be observed in Fig. 1 that scattering of data points from the ideal trend in molecule set is poor.

Brief description of the selected molecular descriptors: In GC × GC, the entire sample is submitted to two online GC separations involving different properties of analytes, *i.e.*, the volatility related to the carbon atom number and the polarity related to the chemical group. In this research, retention data was resulted non polar capillary [capillary (poly(ethylene glycol) column)] used to QSRR models. In the chromatographic retention of compounds in the non polar or low polarity stationary phases two important types of interactions contribute to the chromatographic retention of the compounds *i.e.*, the induction and dispersion forces. The dispersion forces are related to interaction of several intermolecular forces such as dispersion (or London forces), orientation (dipole-dipole or keesom forces) while the induced forces are related to the dipolar moment, which should stimulate dipole-induced dipole or debye forces interactions.

Constitutional descriptors are most simple and 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. Number of C atoms, the average bond order of a C atom and the minimum atomic state energy for a C atom quantify the bond strength between the C atoms. A molecule locked in a rigid conformation due to strong intramolecular interactions is in fact less free to move and is expected to have a higher retention.

The hydrogen bonding a measure of the tendency of a molecule to form hydrogen bonds. This is related to number of hydrogen atoms. Hydrogen-bonding may be divided into an electrostatic term and a polarization/charge transfer term. A particularly strong type of polar interaction occurs in molecules where a hydrogen atom is attached to an extremely electron-attracting atom such as oxygen, nitrogen or fluorine. Understandably, hydrogen bonding plays a significant role. Hydrogen bonding not a true bond, but a very strong form of dipole-dipole attraction.

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.

Different hydrogen bond donors and acceptors are two important parameters introduced to describe molecular properties important for a nanoparticles pharmacokinetics. The availability to form H bonds is an important parameter to define the physico-chemical properties of a naoparticle compounds.

The WHIM descriptors are built in such a way as to capture the relevant molecular 3D information regarding the molecular size, shape, symmetry and atom distribution with respect to some invariant reference frame. WHIM descriptors are quickly computed from the atomic positions of the molecule atoms (hydrogens included). WHIM descriptors are based on principal component analysis of the weighted covariance matrix obtained from the atomic cartesian coordinates. Although functional group, constitutional and WHIM descriptors are often successful in rationalizing retention time of nanoparticles, 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 QSRR.

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 inter-molecular interactions. They include thermodynamic properties (system energies) and electronic properties (HOMO 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.

As expected, the model included HOMO energy to quantify electronic effects of compounds. HOMO energy is a useful descriptor that presents information on the distribution of π electron and explains π–π charge transfer interactions of unsaturated compounds. HOMO energy plays a important role in nucleophilic behaviour and it represents molecular reactivity as a nucleophile. Good nucleophiles are those in which electrons reside in high lying orbitals. Electron affinity was also shown to greatly influence the chemical behaviour of compounds, as demonstrated by its inclusion in the QSRR. The eigenvalue of HOMO reflect the chemical activity of the molecule. HOMO as an electron donor represents the ability to donate an electron.

Polar functional groups account for many of the dipole-dipole, dipole-induced dipole and hydrogen bond interactions. Dipole moment is the measure of polarity of the molecule. Dipole moment describes the intramolecular electronic effect, which may be related to molecular reactivity. The activity of a molecule increases as the dipole moment is increases[39-41]. From the above discussion, it can be seen that the particle size, hydrogen bonding and electrostatic interactions are the likely three factors controlling the retention of these nanoparticles. All the descriptors involved in the model, which have explicit physical meaning, may account for the structure responsible for the retention time of these compounds.

Conclusion

In this research, an accurate QSRR model for estimating the retention time of nanoparticle compounds was developed by employing the one linear model (GA-PLS) and two non-linear models (GA-KPLS and L-M ANN). The most important molecular descriptors selected represent the constitutional, functional group and quantum descriptors that are known to be important in the retention of nanoparticles. Three models have good predictive capacity and excellent statistical parameters. A comparison between these models revealed the superiority of the GA-KPLS and L-M ANN to linear model. It is easy to notice that there was a good prospect for the GA-KPLS and L-M ANN application in the QSRR modeling. This indicates that retention time of nanoparticles possesses some nonlinear characteristics. In comparison with two non-linear models, 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 retention time for new compounds or for other compounds whose experimental values are unknown.

REFERENCES

1. G. Tolnai, G. Alexander, Z. Horvolgyi, Z. Juvancs and A. Dallos, *Chromatographia*, **53**, 69 (2001).
2. P. Viberg, M. Jornten-Karlsson, P. Petersson, P. Spegel and S. Nilsson, *Anal. Chem.*, **74**, 4595 (2002).
3. K.J. Reynold and L.A. Colon, *J. Chromatogr. Rel. Technol.*, **23**, 161 (2000).
4. F. Adam, F. Bertoncini, N. Brodusch, E. Durand, D. Thiebaut, D. Espinat and M.C. Hennion, *J. Chromatogr. A*., **1148**, 55 (2007).
5. M. Deursen, J. Beens, J. Reijenga, P. Lipman, C. Cramers and J. Blomberg, *J. High Resolut. Chromatogr*., **23**, 507 (2000).
6. J. Dalluge, J. Beens and U.A.Th. Brinkman, *J. Chromatogr. A*, **1000**, 69 (2003).
7. G.A. Eiceman, J. Gardea-Torresdey, E. Overton, K. Carney and F. Dorman, *Anal. Chem.*, **74**, 2771 (2002).
8. G.M. Gross, J.W. Grate and R.E. Synovec, *J. Chromatogr. A*, **1029**, 185 (2004).
9. M.C. Daniel and D. Astruc, *Chem. Rev.*, **104**, 293 (2004).
10. A.C. Templeton, M.P. Wuelfing and R.W. Murray, *Acc. Chem. Res*., **33**, 27 (2000).
11. J.W. Grate, D.A. Nelson and R. Skaggs, *Anal. Chem.*, **75**, 1868 (2003).
12. A.C. Lewis, J.F. Hamilton, C.N. Rhodes, J. Halliday, K.D. Bartle, P. Homewood, R.J.P. Grenfell, B. Goody, A.M. Harling, P. Brewer, G.Y. Vargha and M.J.T. Milton, *J. Chromatogr. A*, **1217**, 768 (2010).
13. J. Chen, T. Yang and S.M. Cramer, *J. Chromatogr. A,* **1177**, 207 (2008).
14. X. Hui-Ying, Z. Jian-Wei, H. Gui-Xiang and W. Wei, *Chemosphere*, **80**, 665 (2010).
15. A. Gajewicz, M. Haranczyk and T. Puzyn, *Atmos. Environ*., **44**, 1428 (2010).
16. H. Noorizadeh and A. Farmany, *Chromatographia*, **72**, 563 (2010).
17. A. Niazi, S. Jameh-Bozorghi and D. Nori-Shargh *J. Hazard. Mater*., **151**, 603 (2008).
18. S.H. Woo, C.H.O. Jeon, Y.S. Yun, H. Choi, Ch. Lee and D.S. Lee, *J. Hazard. Mater.*, **161**, 538 (2009).
19. H. Noorizadeh and A. Farmany, *J. Chin. Chem. Soc.*, **57**, 1 (2010).
20. D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley-Longman, Reading, MA, USA, 2000.
21. S. Riahi, E. Pourbasheer, R. Dinarvand, M.R. Ganjali and P. Norouzi, *Chem. Biol. Drug Des*., **72**, 205 (2008).
22. J.A.D. Sousa, M.C. Hemmer and A. Casteiger, *J. Anal. Chem*., **74**, 80 (2002).
23. U. Depczynski, V.J. Frost and K. Molt, *Anal. Chim. Acta,* **420**, 217 (2000).
24. R. Rosipal and L.J. Trejo, *J. Mach. Learning Res*., **2**, 97 (2001).
25. J. Zupan and J. Gasteiger, Neural Network in Chemistry and Drug Design, Wiley-VCH, Weinheim (1999).
26. S. Kara, A.S. Güven, M. Okandan and F. Dirgenali, *Comput. Biol. Med*., **36**, 473 (2006).
27. A. Yasri and D.J. Hartsough, *Chem. Inf. Comput. Sci*., **41**, 1218 (2001).
28. H. Noorizadeh and A. Farmany, *Drug Test Anal.*, (2011), in press.
29. M. Salvi, D. Dazzi, I. Pelistri, F. Neri and J.R. Wall, *Ophthalmology*, **109**, 1703 (2002).
30. K. Kim, J.M. Lee and I.B. Lee, *Chemom. Intell. Lab. Syst*., **79**, 22 (2005).
31. J. Acevedo-Martinez, J.C. Escalona-Arranz, A. Villar-Rojas, F. Tellez-Palmero, R. Perez-Roses, L. Gonzalez and R. Carrasco-Velar, *J. Chromatogr. A,* **102**, 238 (2006).
32. X. Du and Y. Chen, *Shipin Kexue*, **30**, 61 (2009).
33. H. Noorizadeh, A. Farmany and M. Noorizadeh, *Quim. Nova*, **34**, 242 (2011).
34. A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos and O. Igglessi-Markopoulou, *Bioorg. Med. Chem*., **14**, 6686 (2006).
35. J. Wu, M. Li, C. Zhang, L. Wei and S. Zhang, *Jingxi Huagong*, **26**, 968 (2009).
36. A. Golbraikh and A. Tropsha, *J. Mol. Graphics Modell*., **20**, 269 (2002).
37. S. Riahi, E. Pourbasheer, M.R. Ganjali, P. Norouzi and A.Z. Moghaddam, *J. Chin. Chem. Soc*., **55**, 1086 (2008).
38. A. Beheshti, A. Mohammadi, M.R. Ganjali and P. Norouzi, *J. Chin. Chem. Soc*., **55**, 345 (2008).
39. R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, Germany, 2000.
40. H. Khajehsharifi and E. Pourbasheer, *J. Chin. Chem. Soc*., **55**, 63 (2008).
41. M.R. Majidi and K. Asadpour-Zeynali, *J. Chin. Chem. Soc*., **52**, 21 (2005).