Middle-East Journal of Scientific Research 12 (11): 1454-1458, 2012
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2012.12.11.3

Nonlinear Chemometrics Model for Prediction Retention Behavior of Nanoparticles

Hadi Noorizadeh and Abbas Farmany

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

Abstract: Nanoparticles have attracted extensive attention in various fields of chemistry, physics and medicine due to their novel and fascinating properties. Genetic algorithm and kernel partial least square (GA-KPLS) technique was used to investigate the correlation between capacity factor and descriptors for 43 nanoparticle compounds. The KPLS was utilized to construct the nonlinear quantitative structure- retention relationship (QSRR) model. The model was validated using Leave-Group-Out cross validation (LGO-CV) and external test set.

Key words: Nanoparticle compounds · QSRR · Chemometrics

INTRODUCTION

Nano has been and continues to be one of the most hyped areas in science and technology today [1, 2]. It is a collective realization that interesting chemistry and physics occurs in the previously unexplored hinterland between the truly molecular (the traditional realm of chemistry and even physics) and bulk matter (the traditional realm of engineers). To illustrate, when gold, silver, as well as other metals and even semiconductors are made small enough, they no longer behave in ways that we are accustomed to seeing.

Nanoparticles constitute a crucial and technology intensive area of research and development in the burgeoning field of nanotechnology. The attraction of nanoparticles lies in the myriad attractive characteristics which can be achieved by reducing suitable materials from the bulk to the nanometer size, these characteristics ranging from increased surface/volume ratio to novel quantum confinement effects. Examples of property enhancements include magnetic, optical, biosensing, thermoelectric, semiconducting, catalytic, energy storage and thermal properties. Industries that are interested in novel nanoparticles include data storage, plasmonic, photonic, microelectronic, energy, pharmaceutical, biomedical and cosmetics. An interesting aspect of nanoparticles is the wide range of materials classes in which nanoparticles are useful including semiconductor, dielectric, metallic, ceramic, composite and polymer nanoparticles. The unconventional and highly interesting topic dealing with nanoparticles in cosmetics such as whitening agents, moisturizers and antiaging products, toxicology issues is also discussed. The book is organized according to the type of material including magnetic nanoparticles, followed by nanoparticles for biomedicine, oxide nanoparticles and particles for actuation, sensing and cosmetic. Gold nanoparticles, also known as colloidal gold, are small spheres of gold and can be obtained in sizes between 3 and 150 nm in diameter. In this reaction, sodium citrate reduces the gold cations in hydrogen tetrachloroaurate. As the gold metal forms, anions coat the outside of the particles preventing them from forming larger particles. The small gold nanoparticles that are formed stay in solution because their coatings are negatively charged and repel each other, preventing aggregation. Nano-sized materials often have different properties than the bulk materials; for example, gold nanoparticles appear red.

Particle size and size distribution are the most important characteristics of nanoparticle systems. They determine the *in vivo* distribution, biological fate, toxicity and the targeting ability of nanoparticle systems. In addition, they can also influence the drug loading, drug release and stability of nanoparticles.

When nanoparticles are administered intravenously, they are easily recognized by the body immune systems and are then cleared by phagocytes from the circulation. Apart from the size of nanoparticles, their surface hydrophobicity determines the amount of adsorbed blood components, mainly proteins (opsonins). This in

Corresponding Author: Hadi Noorizadeh, Faculty of Science, Islamic Azad University, Ilam Branch, Ilam, Iran.
Tel: + 989183452507.

turn influences the *in vivo* fate of nanoparticles. Binding of these opsonins onto the surface of nanoparticles called opsonization acts as a bridge between nanoparticles and phagocytes. The association of a drug to conventional carrier's leads tomodification of the drug biodistribution profile, as it is mainly delivered to the mononuclear phagocytes system (MPS) such as liver, spleen, lungs and bone marrow. Indeed, once in the blood stream, surface non-modified nanoparticles (conventional nanoparticles) are rapidly opsonized and massively cleared by the macrophages of MPS rich organs.

Ideally, a successful nanoparticulate system should have a high drug-loading capacity thereby reduce the quantity of matrix materials for administration. Drug loading can be done by two methods: Incorporating at the time of nanoparticles production (incorporation method) and absorbing the drug after formation of nanoparticles by incubating the carrier with a concentrated drug solution (adsorption /absorption technique).

Mathematical modeling of interactions in chromatography helps chemists to find a model that can be used to obtain a deep understanding about the mechanism of interaction and to predict the capacity factor (k') of new or even unsynthesized compounds. Building retention prediction models may initiate such theoretical approach and several possibilities for retention prediction in GC. Among all methods, quantitative structure-retention relationships (QSRR) are the most popular. In QSRR, the retention of given chromatographic system is modeled as a function of solute (molecular) descriptors. A number of reports, dealing with QSRR calculation of several compounds, have been published in the literature [6-8].

There is a trend to develop QSRR from a variety of methods. In particular, genetic algorithm (GA) is frequently used as search algorithms for variable selection in chemometrics and QSRR. GA is a stochastic method to solve the optimization problems defined by fitness criteria, applying the evolution hypothesis of Darwin and different genetic functions, i.e. crossover and mutation [9, 10]. Kernel partial least square (KPLS) is the most commonly used multivariate calibration method [11, 12]. In the present study, GA-KPLS was employed to generate QSRR models that correlate the structure of petroleum hydrocarbons; with observed capacity factor.

MATERIALS AND METHODS

Data Set: Capacity factor (k') of the nanoparticles which contains 43 compounds was taken from literature [13] are presented in Table 1. All chromatograms were obtained with an injection source and FID temperature of 250°C. The inlet pressure was maintained at 48,000 Pa with a variable split as stated, while the auxiliary pressure (column pressure) was varied independently as dictated by the experimental method being employed. The oven temperature was constant at 50°C unless otherwise noted.

Table 1: The data set and corresponding observed k' values

No	Name	k'
	Training Set	
1	Cyclohexane	0.461
2	Ethyl formate	0.472
3	Ethyl acetate	0.483
4	1,1,1-Trichloroethane	0.492
5	n-Butylamine	0.522
6	Benzene	0.525
7	Hexyne	0.551
8	1-Heptyne	0.551
9	2-Pentanone	0.556
10	1-Chlorobutane	0.567
11	Nitroethane	0.576
12	Methylcyclohexane	0.583
13	2-Butanol	0.59
14	Hexanal	0.759
15	Pyridine	0.76
16	Butyl acetate	0.773
17	2-Hexanone	0.779
18	Triethylamine	0.78
19	Toluene	0.793
20	1-Nitrobutane	0.846
21	Cycloheptane	0.89
22	1-Bromopentane	0.906
23	1-Chlorohexane	0.995
24	cis-1,2 Dimethylcyclohexane	1
25	2-Pentanol	1.12
26	Heptanal	1.239
27	Ethylbenzene	1.269
28	Chlorobenzene	1.274
29	1-Nonene	1.315
30	Cyclohexylamine	1.488
31	1-Nonyne	1.491
32	Bromohexane	1.84
33	Methyl phenyl ether (anisole)	1.859
34	Cyclooctane	1.869
35	Bromobenzene	2.519
36	Octanal	2.572
	Test Set	
1	Hexane	0.44
2	Heptene	0.535
3	Butyl formate	0.596
4	1,1,2-Trichloroethane	0.83
5	1-Butanol	1.151
6	p-Xylene	1.444
7	1-Pentanol	2.054

For the GC × GC experiments, either a 4 or 15m poly (ethyleneglycol) column with a 250 µm i.d. and 0.2 µm film thickness (IMMOWax, Agilent Technologies, Palo Alto, CA, USA) was used as the first column of the GC × GC system and a dodecanethiol MPN column as the second column. In order to evaluate the generated models, we used leave-group-out cross validation (LGO-CV). This methodology systematically removed one group data at a time from the data set. A QSRR model was then constructed on the basis of this reduced data set and subsequently used to predict the removed data set. This procedure was repeated until a complete set of predicted was obtained.

Descriptor Calculation: All structures were drawn with the HyperChem software (version 6). Optimization of molecular structures was carried out by semi-empirical AM1 method using the Fletcher- Reeves algorithm until the root mean square gradient of 0.01 was obtained. Since the calculated values of the electronic features of molecules will be influenced by related conformation. In the current research an attempt was made to use the most stable conformations. Some electronic descriptors such as polarizability, dipole moment and orbital energies of LUMO and HOMO were calculated by the Hyper Chem software. Also optimized structures were used to calculate 1497 descriptors by DRAGON software Version 3.

One of the challenging parts in developing models is choosing suitable parameters encoding different aspects of the molecular structure. A large number of structural descriptors can be calculated using existing software's such as Dragon. However, nowadays the main problem is choosing the most adequate and interpretable parameters needed for developing the models among a large number of them. To reduce the original pool of descriptors to an appropriate size, 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 that of other descriptors present in the pool. Any descriptor that had identical or zero values for greater than 90% of the compounds was eliminated.

Genetic Algorithm: Genetic algorithm has been proposed by John Holland in the early 1970s but it was possible to apply them with reasonable computing times only since 1990s, when computers became much faster. GA is a stochastic method to solve the optimization problems, defined by fitness criteria applying to the evolution hypothesis of Darwin and different genetic functions, i.e., crossover and mutation. Compared to the traditional search and optimization procedures, GA is robust, global and generally more straightforward to apply to situations where there is little or no a priori knowledge about the process to be controlled. Since GA does not require derivative information or a formal initial estimate of the solution region and because of the stochastic nature of the search mechanism, it is capable to search the entire solution space with a greater probability of finding the global optimum. In GA, each individual of the population, defined by a chromosome of binary values as the coding technique, represented a subset of descriptors. The number of the genes at each chromosome was equal to the number of the descriptors. The population of the first generation was selected randomly. A gene was given the value of one, if its corresponding descriptor was included in the subset; otherwise, it was given the value of zero.

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 Hyper Chem (Version 7.0 Hypercube, Inc.), Dragon software was used to calculate of RI. MLR analysis was performed by the SPSS Software (version 13, SPSS, Inc.) by using enter method for model building. MINITAB software (version 14, MINITAB) was used for the simple PLS analysis. Cross validation, GA-PLS, GA-MLR, L-M ANN and other calculation were performed in the MATLAB (Version 7, Mathworks, Inc.) environment.

RESULTS AND DISCUSSION

Ga-kpls Analysis: With the aim of improving the predictive performance of nonlinear QSRR model, GA-KPLS modeling was performed. The leave-group-out cross validation 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 set to be 1), m is the dimension of the input space and σ^2 is the variance of the data [14]. It means that the value of c depends on the system under the study. The best model is selected on the basis of the root mean square error (RMSE) of prediction and simplicity of the model. The best GA-KPLS model contains 24 selected descriptors in 16 latent variables space. The RMSE for training and test sets was (0.035, 0.112). The predicted values of k' are plotted against the experimental values for training and test set in Fig. 1. Obviously, there is a close agreement between the

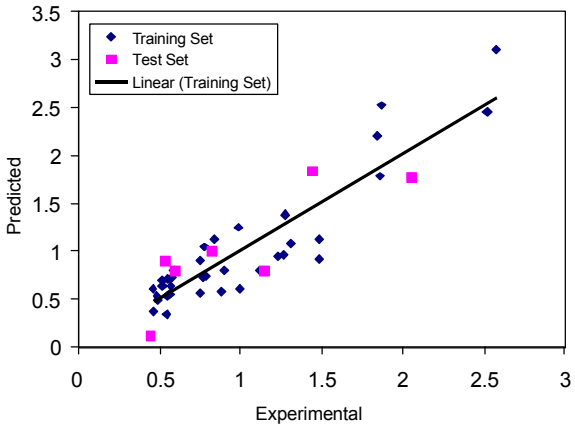

Fig. 1: Predicted vs. experimental k' by GA-KPLS

experimental and predicted k' and the data represent a very low scattering around a straight line with respective slope and intercept close to one and zero. The result indicates that the GA-KPLS model have good statistical quality with low prediction error. The Q^2, which is a measure of the model fit to the cross validation set, can be calculated as:

$$R_{cv}^2 \equiv Q^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (1)$$

where y_i, $\hat{y_i}$ and $\bar{y}$ were respectively the experimental, predicted and mean k' 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 as proof of the high predictive ability of the model [15]. However, several authors suggest that a high value of Q^2 appears to be a necessary but not sufficient condition for a model to have a high predictive power and consider that the predictive ability of a model can only be estimated using a sufficiently large collection of compounds that was not used for building the model [16].

We believe that applying only LGO-CV is not sufficient to evaluate the predictive ability of a model. Thus we employed a two-step validation protocol which contains internal (LGO-CV) and external (test set) validation methods. The data set was randomly divided into training (calibration and prediction sets) and test sets after sorting based on the k' values. The training set consisted of 36 molecules and the test set, consisted of 7 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. Inspection of the results reveals a lower RMSE for GA-KPLS model for the training and test sets. This clearly shows the strength of GA-KPLS as a nonlinear feature selection method. Result indicates that the k' of petroleum hydrocarbons possesses some nonlinear characteristics.

CONCLUSION

In this study, an accurate QSRR model for estimating the k' of nanoparticles was developed by employing the one nonlinear model (GA-KPLS). A model has good predictive capacity and excellent statistical parameters. It is easy to notice that there was a good prospect for the GA-KPLS application in the QSRR modeling. This indicates that k' of these compounds possesses some nonlinear characteristics. It can also be used successfully to estimate the k' for new compounds or for other compounds whose experimental values are unknown.

REFERENCES

1. Pavel Řezanka, Hana Řezanková, Pavel Matějka and Vladimír Král, 2010. The chemometric analysis of UV–visible spectra as a new approach to the study of the NaCl influence on aggregation of cysteine-capped gold nanoparticles, Colloids and Surfaces A: Physicochemical and Engineering Aspects, 364: 94-98.
2. Aurora Mocanu, Ileana Cernica, Gheorghe Tomoaia, Liviu-Dorel Bobos and Ossi Horovitz, 2009. Maria Tomoaia-Cotisel, Self-assembly characteristics of gold nanoparticles in the presence of cysteine, Colloids and Surfaces A: Physico chemical and Engineering Aspects, 338: 93-101.
3. Kenji Hamaguchi, Hideya Kawasaki and Ryuichi Arakawa, 2010. Photochemical synthesis of glycine-stabilized gold nanoparticles and its heavy-metal-induced aggregation behavior, Colloids and Surfaces A: Physicochemical and Engineering Aspects, 367: 167-173.
4. Tatsuo Taniguchi, Takumi Inada, Takuya Kashiwakura, Fumiyasu Murakami, Michinari Kohri and Takayuki Nakahira, 2011. Preparation of polymer core-shell particles supporting gold nanoparticles, Colloids and Surfaces A: Physicochemical and Engineering Aspects, 377: 63-69.

5. Guojun Weng, Jianjun Li, Jian Zhu and Junwu Zhao, 2010. Decreased resonance light scattering of citrate-stabilized gold nanoparticles by chemisorption of mercap toacetic acid, Colloids and Surfaces A: Physico Chemical and Engineering Aspects, 369: 253-259.
6. D¹browska, M., M. Starek and J. Skuciñski, 2011. Lipophilicity study of some non-steroidal anti-inflammatory agents and cephalosporin antibiotics: A review, Talanta, 86: 35-51.
7. D'Archivio, A.A., A. Incani and F. Ruggieri, 2011. Cross-column prediction of gas-chromatographic retention of polychlorinated biphenyls by artificial neural networks, J. Chromatogr A, 12(18): 8679-8690.
8. Noorizadeh, H. and A. Farmany, 2010. QSRR Models to Predict Retention Indices of Cyclic Compounds of Essential Oils, Chromatographia, 72: 563-569.
9. Noorizadeh, H. and M. Noorizadeh, 2011. QSRR-based estimation of the retention time of opiate and sedative drugs by comprehensive two-dimensional gas chromatography, Med Chem Res, in Press.
10. Van Dijck, G. and M.M. Van Hulle, 2011. Genetic algorithm for informative basis function selection from the wavelet packet decomposition with application to corrosion identification using acoustic emission, Chemom. Intell. Lab. Syst., 107: 318-332.
11. Noorizadeh, H. and A. Farmany, 2011. Quantitative structure-retention relationship for retention behavior of organic pollutants in textile wastewaters and landfill leachate in LC-APCI-MS, Environ Sci Pollut Res, in Press.
12. Ribeiro, R.J., F. Augusto, T.J.S. Salva, R.A. Thomaziello and M.C. Ferreira, 2009. Prediction of sensory properties of Brazilian Arabica roasted coffees by headspace solid phase microextraction-gas chromatography and partial least squares, Anal. Chim. Acta, 634: 172-179.
13. Hamid Karimi, Abbas Farmany and Hadi Noorizadeh, 2011. Prediction of Linear Retention Index of Teucrium chamaedrys Volatiles in GC×GC-TOF/MS by Linear Model, World Applied Sciences Journal, 15(8): 1086-1088.
14. Noorizadeh, H., A. Farmanya, H. Narimani and M. Noorizadeh, 2011. QSRR using evolved artificial neural network for 52 common pharmaceuticals and drugs of abuse in hair from UPLC-TOF-MS, Drug Test. Anal, in Press.
15. Tan, A., I.A. Lévesque, I.M. Lévesque, F. Viel and N. Boudreau, 2011. Analyte and internal standard cross signal contributions and their impact on quantitation in LC-MS based bioanalysis, J. Chromatogr. B., 879: 1954-1960.
16. Hadi Noorizadeh and Abbas Farmany, 2011. QSPR Studies of Artemisia Essential Oils by the Combination of Genetic Algorithms and PLS Analysis, World Applied Sciences Journal, 14(4): 603-606.