Environmental Sciences Vol.11, No.4 , Winter 2014

105-110

QSRR study of Gas Chromatographic Behavior of Selected Pesticides in Green Tea Samples

Amir Salemi*

Assistant Professor, Departmetn of , Environmental Sciences Research Institute, Shahid Beheshti University,Tehran, Iran.

Received: 2014/0/0 Accepted: 2014/0/0

Abstract

Different classes of pesticides are used during production of various tea products. Beside their health impact on the consumers, these pesticides are likely to be released into the different natural media and affect the exposed environment. To help predict chromatographic behavior of these pesticides, the gas chromatographic retention behavior of 57 of these pesticides were modeled using quantitative structure-retention relationship approach. The most important descriptors encoding structural and topological properties of the studied compounds were showed to be in a linear relation with their retention time. A stepwise variable selection strategy in MLR modeling resulted in models with acceptable R2 values of which a model based on five molecular descriptors was selected to compromise between high R2 values and low variable numbers. The selected descriptors were VRD2, X1sol, Rww, MLOGP and More09e. The model was tested for its prediction capability, by examining a prediction set of randomly selected compounds (10 pesticides) and the average of the prediction error was used as the criteria. The model, regarding its simplicity, was successful in predicting the retention times of the proposed set of compounds.

Key words: Gas chromatography; Pesticides; Quantitative structure-retention relationship; Tea leave.

مطالعه ارتباط کمی ساختار - بازداری آفت کش‌های برگزیده در برگ سبز چای

امیر سالمی

استادیار گروه آلاینده های محیط زیست، پژوهشکده علوم محیطی، دانشگاه شهید بهشتی، تهران، ایران

چکیده

دسته های گوناگونی از آفت کش ها طی مراحل مختلف تولید فراورده های چای به کار می روند. این ترکیبات، علاوه بر تاثیر منفی بر سلامت مصرف کنندگان، امکان ورود به محیط های طبیعی و ایجاد تاثیرات سوء در آن ها را نیز دارند. به منظور کمک به درک رفتار کروماتوگرافی این آفت کش ها حین انجام آنالیز، زمان بازداری ۵۷ مورد از آن ها با استفاده از روش ارتباط کمی ساختار- بازداری مطالعه شد. نشان داده شد که مهم ترین توصیف گرهایی که اطلاعات ساختاری و توپولوژیک ترکیبات را بیان می کنند، رابطه ای خطی با زمان بازداری دارند. استفاده از روش مرحله ای در رگرسیون خطی چندگانه مدل هایی ایجاد کرد که مقدار R2 قابل قبولی داشتند و از میان این مدل ها، مدلی با پنج توصیف گر انتخاب شد تا شرایط بهینه بین حداقل تعداد توصیف گرها و حداکثر مقدار R2 برقرار شود. توصیف گرهای منتخب عبارت بودند از VRD2، X1Sol، Rww، MLOGP و Mor09e. توانایی مدل در پیش بینی زمان بازداری با استفاده از یک مجموعه پیش بینی که به صورت تصادفی از بین ترکیبات انتخاب شده بودند (۱۰ آفت کش) ارزیابی شد و مینگین خطای پیش بینی به عنوان معیار در نظر گرفته شد. مدل نهایی، با وجود سادگی، در پیش بینی زمان بازداری آفت کش های مورد بررسی کاملاً موفق بود.

کلمات کلیدی: کروماتوگرافی گازی، آفت کش، ارتباط کمی ساختاربازداری، برگ چای.

* Corresponding author. *Email Address*: a_salemi@sbu.ac.ir

1. Introduction

The leaves of the plant Camillia sinensis is used to produce tea which depending on its processing procedure would result in one of the green, black, red, or other types of tea [1]. A wide range of pesticides in different steps of tea production is applied and so the determination of all of the possible pesticides in tea leaves would be a rather difficult multi-residue analysis, even when a high resolution gas chromatograph coupled with tandem mass spectrometry is considered as the analytical instrument [2]. The accurate and precise analysis of these pesticides is not restricted to the different tea products, and in fact, the environmental samples taken from the nearby area are also likely to be contaminated by the same classes of pollutants.

Quantitative structure-activity relationship (QSAR) is a powerful technique that enables scientists to relate different properties of the compounds of interest to their structural properties. This way, not only the fundamental description of the studied phenomenon becomes possible, but also the calculated and validated model would provide a mean to predict the behavior of the similar compounds. Regarding the wide variety of applications and capabilities of the gas and liquid chromatographic techniques in quantitative and qualitative analyses, a considerable portion of QSAR studies have focused on the investigation of chromatographic behavior of the chemical compounds and for the interpretation and also the prediction of their chromatographic properties [3]. The relative retention times, the Kovats indices, retention volume and finally, the retention times of the analytes have been the most commonly selected representatives of the chromatographic behavior of the compounds of interest [4]. So, this field of the scientific research is routinely called quantitative structure-retention relationships (QSRR) to emphasize on its specific application.

The first step in QSRR is to generate or find a relatively large set of retention data obtained under similar experimental conditions; the same instrument, analytical column, detector, etc [5]. Then, all possible structural properties of the analytes (compounds) are converted to numerical values, called descriptors. Today, thousands of descriptors can be calculated in different categories, such as geometrical descriptors, quantum descriptors, three dimensional descriptors and many more. The last and usually the most important step in QSRR is to find a mathematical relationship (model) between the selected descriptors and the chromatographic retention values [6-8].

Multiple linear regression has been one of the most frequently used technique to find the above mentioned relationship and build the model. Although MLR, in its original approach, is only applicable for the linear responses related to a single dependent variable, however its simplicity and straightforwardness keep it useful and worth testing. More complicated models could be build using artificial neural network, AAN [9], partial least square, PLS [10] and principal component analysis, PCA [11] which are capable of dealing with non-linear phenomena and reducing the dimensions of the original dataset.

It is worth noting that success of any QSRR model is partly determined by the degree of similarity between the studied analytes. In other words, building a single model to quantitatively describe the chromatographic behavior of dissimilar classes of compounds, might be a challenging task. In the present work, regarding the multi-class selection of pesticides for protection of tea leaves, which produces a heterologous data set, applicability of the QSRR approach will be investigated for interpreting and predicting the gas chromatographic retention of the mentioned analytes. Another important fact that must be pointed is that the complex set of the selected pesticides may be of great importance in the environmental resources affected by the transported contaminants from tea farms and related industries.

2. Experimental

Data set

The chromatographic retention data was obtained from a study on multi-residue analysis of pesticides in

tea leaves [12]. Three different extraction techniques has been compared regarding their efficiency and recovery values, while similar analytical instrumentation was used for the determination of the analytes. A capillary gas chromatography coupled with tandem mass spectrometry was optimized to obtain the best resolution and peak separation. The studied pesticides were as follows; Alachlor, Ethoprophos, Mevinphos, Benalaxyl, Etofenprox, Oxadixyl, Bifenthrin, Fenarimol, Parathion ethyl, Biphenyl, Fenitrotion, Parathion methyl, Boscalid, Fenpropathrin, Penconazole, Chinomethionat, Fenvalerate RR, Permethrin, Chlorpyrifos, Fenvalerate SS, Phorate, Chlorpyrifos methyl, Fenvalerate RS, Phosmet, Cypermethrin, Fenvalerate SR, Propargite, Cyproconazole, Flucythrinate, Propyzamide, Cyprodinil, Fludioxonil, Pyridaben, Diazinon, Fonofos, Pyrimethanil, Dichlofluanid, Iprodione, Pyriproxyfen, Dichlorvos, Isocarbophos, Quinalphos, Dicloran, Isofenfos methyl, Terbufos, Endosulphan I, Kresoxim methyl, Tetrachlorvinphos, Endosulphan II, Lindane, Triazophos, Endosulphan sulphate, Malathion, Trifluralin, Ethion, Methidathion and Vinclozolin.

Generation of descriptors

First, the chemical structure of the analyzed pesticides were drawn and saved using ChemDraw from ChemOffice 12 molecular modeling package (CambridgeSoft, Whitehouse, USA). Then the files were imported to Hyperchem (ver. 8) to build their three dimensional structure and to optimize using semiempirical, quantum-chemical methods of AM1 Hamiltonian. Dragon software was employed to generate descriptors. This software has been widely used for calculating chemical descriptors in many similar researches.

Preliminary selection of descriptors

The following is the procedure to minimize the total number of the calculated descriptors.

- All the descriptors with zero/constant values were deleted.
- A binary correlation matrix was generated, including the retention time data. Then all the descriptors having correlation coefficient between -0.3 and 0.3 with retention time, were removed from the data set.
- For each cluster of descriptors having correlation coefficients more than 0.9 with each other, one which had the highest correlation with retention time was kept and other descriptors were deleted.

The final data set, containing the remained descriptors as well as the retention time values of the analytes, were exported to SPSS software for the rest of the study.

MLR model

Ten randomly selected analytes were separated from the dataset to form a prediction set. This way, the training set as a matrix of 47 rows (compounds/ retention times) and 154 columns (153 descriptors) were subjected to the MLR studies. Among various approaches in multiple linear regression, stepwise variable selection strategy was selected to build the model for relating the dependent variable (retention time) to the independent variables (descriptors). The resulting model was tested using the prediction set and the error in the prediction of the experimental retention time as well as R^2 values were used to verify the efficiency of the model.

3. Results and discussion

Generating the QSRR model

Model selection based on the MLR results. The resulting models after implying MLR on the training set, is summarized in Table 1. The selected descriptors (named "predictors" by the SPSS software) for each model are also can be observed. To demonstrate the change in the model R^2 value versus the change in the number of descriptors, the data was pictured in Figure 1. Here both the R^2 values as well as their changes due to the increasing the model variables can be observed more clearly. The trend lines in the plot refer to the application of the

break-point technique, by which the point that the slope of the curve changes is determined. The similar result might be concluded from the model summary (Table 1), where the change in R^2 becomes less than 0.02. In this work, both approaches showed that the model did not improved significantly after the model number 5 (consisting of 1 constant and 5 descriptors).

Table1. Summary of the models obtained by MLR and corresponding descriptors

Model	*R*	*Adjusted R Square*	*R Square Change*	*Model descriptors*
1	0.903	0.811	.000	(Constant), VRD2
2	0.922	0.844	.034	(Constant), VRD2, X1sol
3	0.945	0.886	.043	(Constant), VRD2, X1sol, Rww
4	0.958	0.912	.026	(Constant), VRD2, X1sol, Rww, MLOGP
5	0.965	0.924	.013	(Constant), VRD2, X1sol, Rww, MLOGP, Mor09e
6	0.970	0.933	.010	(Constant), VRD2, X1sol, Rww, MLOGP, Mor09e, MATS1v
7	0.975	0.943	.010	(Constant), VRD2, X1sol, Rww, MLOGP, Mor09e, MATS1v, GGI7
8	0.979	0.951	.008	(Constant), VRD2, X1sol, Rww, MLOGP, Mor09e, MATS1v, GGI7, RDF060v
9	0.981	0.955	.004	(Constant), VRD2, X1sol, Rww, MLOGP, Mor09e, MATS1v, GGI7, RDF060v, RDF065v
10	0.984	0.962	.006	(Constant), VRD2, X1sol, Rww, MLOGP, Mor09e, MATS1v, GGI7, RDF060v, RDF065v, RDF030v
11	0.984	0.961	-.001	(Constant), X1sol, Rww, MLOGP, Mor09e, MATS1v, GGI7, RDF060v, RDF065v, RDF030v
12	0.983	0.960	-.002	(Constant), X1sol, Rww, MLOGP, Mor09e, MATS1v, GGI7, RDF065v, RDF030v
13	0.985	0.971	.005	(Constant), X1sol, Rww, MLOGP, Mor09e, MATS1v, GGI7, RDF065v, RDF030v, RDF090v

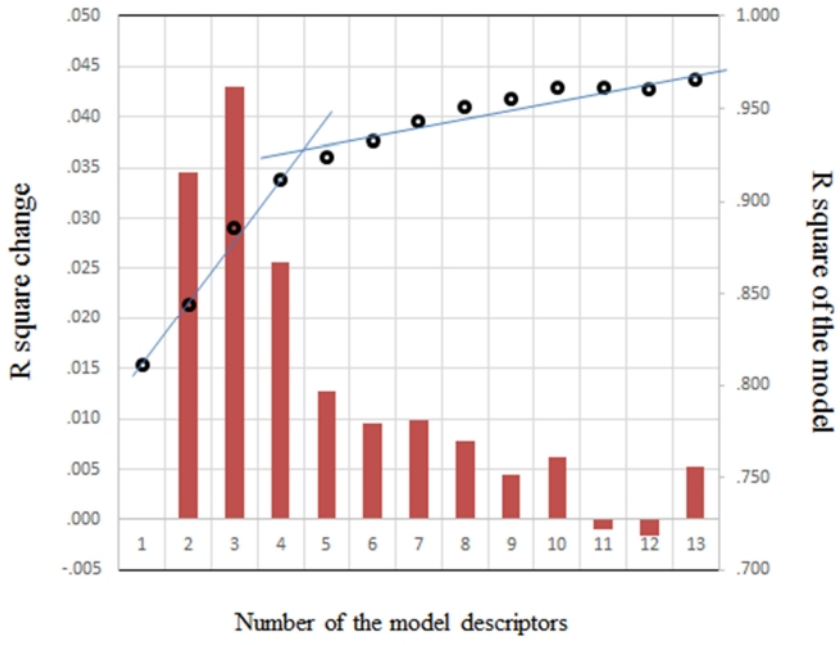

Figure 1. Demonstration of R square values of the model (circles) as well as its change (bars) versus the number of model descriptors. The plot was used in determining the break-point of the model.

Model selection based on the prediction accuracy. Beside the calculated statistical results of the models, their capability in predicting the retention time of the analytes in the prediction set, can be considered as an important criteria for evaluating the model performance and variable selection. So, the retention times of the compounds in prediction set were calculated using models 1 to 7. Figure 2 shows the average error of prediction for the mentioned model and it is clear that this error decreased as result of increasing the number of the descriptors. It is worth noting that the main goal of this study was to find the simplest model to describe the retention behavior of the studied compounds. So, again the model number 5 could be selected as the optimum model which could compromise between the lowest possible number of variables and the best practical model accuracy.

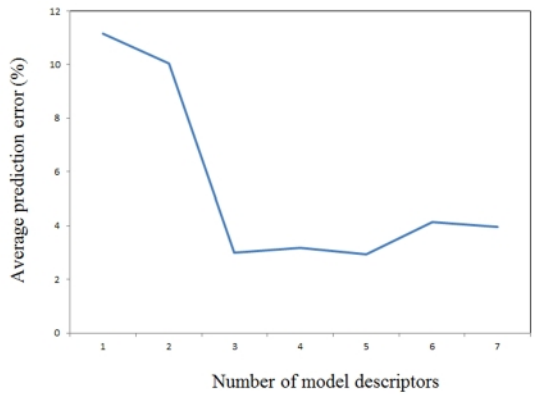

Figure 2. Plot of average prediction error for the ten pesticide of the prediction set versus the number of the descriptors used in MLR models.

The final model and contributing descriptors

The results of MLR lead to a quantitative relationship between gas chromatographic retention time and 5 descriptors as the equation t_R = -4.85 + (1.585VRD2) + (0.724X1sol) + (-0.542Rww) + (-0.818MLOGP) + (-0.187Mor09e). The following is a brief description of these descriptors.

VRD2. This is an eigenvalue-based descriptor known as one of the Randic-like indices based on the coefficients ℓ_{iA} of the eigenvector associated with the largest negative eigenvalue.

X1sol. Solvation connectivity index chi-1 or X1sol is a topological descriptor and was proposed by Zefirov and Palyulin in order to treat the enthalpies of non-specific solvation.

Rww. This descriptor a reciprocal hyper-detour index and is categorized as a topological descriptor.

MLOPG. Moriguchi model based on structural parameters, is a model described by a regression equation based on 13 structural parameters. The heteroatoms and functional groups have significant effect on this descriptor.

Mor09e. Abbreviated from 3D-MOlecule Representation of Structures based on Electron, is three dimensional descriptor with mathematical representation as a vector.

4. Conclusion

The gas chromatographic retention behavior of 57 pesticides which are frequently used in production of tea leaves were modeled. The most important descriptors encoding structural and topological properties of the studied compounds were showed to be in a linear relation with their retention time. A stepwise variable selection strategy in MLR modeling resulted in models with acceptable R^2 values of which a model based on five molecular descriptors were selected to compromise between higher R^2 values and lower variables.

Acknowledgments

Research Council of Shahid Beheshti University, G.C. is appreciated for supporting the presented research.

References

[1] Zhang X, Mobley N, Zhang J, Zheng X, Lu L, Ragin O, Smith C J. Analysis of agricultural residues on tea using d-SPE sample preparation with GC-NCI-MS and UHPLC-MS/MS. Journal of Agricultural and Food Chemistry; 2010; 58(22):11553–11560.

[2] Xu X-M, Yu C, Han J-L, Li J-P, El-Sepai F, Zhu

Y, Huang B-F, Cai Z-X, Wu H-W, Ren Y-P. Multi-residue analysis of pesticides in tea by online SEC-GC/MS. Journal of Separation Science; 2011; 34(2):210-216.

[3] Heberger K. Quantitative structure–(chromatographic) retention relationships. Journal of Chromatography A; 2007; 1158:273–305.

[4] Liu F, Liang Y, Cao C, Zhou N. QSPR study of GC retention indices for saturated esters on seven stationary phases based on novel topological indices. Talanta; 2007; 72:1307–1315.

[5] Put R, Vander Heyden Y. Review on modelling aspects in reversed-phase liquid chromatographic quantitative structure–retention relationships. Analytica Chimica Acta; 2007; 602:164–172.

[6] Rena Y, Liua H, Yaoa X, Liu M. Three-dimensional topographic index applied to the prediction of acyclic C5–C8 alkenes Kováts retention indices on polydimethylsiloxane and squalane columns. Journal of Chromatography A; 2007; 1155:105–111.

[7] Czek T B, Kaliszan R. Combination of linear solvent strength model and quantitative structure–retention relationships as a comprehensive procedure of approximate prediction of retention in gradient liquid chromatography. Journal of Chromatography A; 2002; 962: 41–55.

[8] Lu C, Guo W, Yin C. Quantitative structure-retention relationship study of the gas chromatographic retention indices of saturated esters on different stationary phases using novel topological indices. Analytica Chimica Acta; 2006; 561:96–102.

[9] Jevrić L R, Podunavac-Kuzmanović S O, Švarc-Gajić J V, Tepić A N, Kovačević S Z, Kalajdžija N D. QSRR modeling of retention behavior of some s-triazine derivatives. Acta Chimica Slovenica; 2013; 60(4): 732-42.

[10] Akbar J, Iqbal S, Batool F, Karim A, Chan K W. Predicting Retention Times of Naturally Occurring Phenolic Compounds in Reversed-Phase Liquid Chromatography: A Quantitative Structure-Retention Relationship (QSRR) Approach. International Journal of Molecular Sciences; 2012; 13(11):15387-15400.

[11] Santos C B, Vieira J B, Lobato C C, Hage-Melim L I, Souto R N, Lima C S, Costa E V, Brasil D S, Macêdo W J, Carvalho J C. A SAR and QSAR study of new artemisinin compounds with antimalarial activity. Molecules; 2013; 19(1):367-399.

[12] Rajski L, Lozano A, Belmonte-Valles N, Ucles A, Ucles S, Mezcuaa M, Fernandez-Alba A R. Comparison of three multiresidue methods to analyse pesticides in green tea with liquid and gas chromatography/ tandem mass spectrometry. Analyst; 2013; 138:921-931.