

RESEARCH PAPER

Novel consensus quantitative structure-retention relationship method in prediction of pesticides retention time in nano-LC

Zahra Pahlavan Yali*, Mohammad Hossein Fatemi

Chemometrics Laboratory, Faculty of Chemistry, University of Mazandaran, Babolsar, Iran

ARTICLE INFO

Article History:

Received 10 March 2018

Accepted 29 May 2018

Published 1 August 2018

Keywords:

Average Consensus Model

Genetic Algorithm-Multiple

Linear Regression

Nano-LC Retention Time

Pesticides

Quantitative Structure-

Retention Relationship

ABSTRACT

In this study, quantitative structure-retention relationship (QSRR) methodology was employed for modeling the retention times of 16 banned pesticides in nano-liquid chromatography (nano-LC) column. Genetic algorithm-multiple linear regression (GA-MLR) method was employed for developing global and consensus QSRR models. The best global GA-MLR model was established by adjusting GA parameters. Three descriptors of *SpMax2_Bhp*, *Mor31u* and, *MATS6c* appeared in this model. Consensus QSRR models were developed as an average consensus model (ACM) and weighted consensus model (WCM) by the combination of a subset of the GA-MLR models. Comparison of statistical parameters of the developed models indicated that an ACM, which is a combination of the best global QSRR model with four-descriptor sub-model, can be selected as the best consensus QSRR model. *CrippenLogP*, *RDF070m*, *Lop*, and *HASA1* descriptors appeared in a four-descriptor sub-model. In ACM, the square of correlation coefficients (R^2) was 0.973 and 0.939, and the SE was 0.49 and 0.40, for the training and test sets, respectively. The ACM was assessed by leave one out cross-validation ($Q^2_{cv} = 0.935$) as well as internal validation. Descriptors appeared in this model suggest that electrostatic, steric and hydrophobic interactions play the main role in the chromatographic retention of the studied pesticides in nano-LC conditions.

How to cite this article

Pahlavan Yali Z, Fatemi MH. Novel consensus quantitative structure-retention relationship method in prediction of pesticides retention time in nano-LC. *Nanochem Res*, 2018; 3(2):205-211. DOI: [10.22036/ncr.2018.02.010](https://doi.org/10.22036/ncr.2018.02.010)

INTRODUCTION

Pesticides are considered a vital component of modern farming playing a major role in food security and public health [1-3]. Food security is defined as access of all people to adequate and safe food [4]. Pesticides enhanced the quantity and quality of products that respectively lead to a high agricultural productivity and improving the product health [3]. Also, these chemicals directly improve the public health by mitigating some pathogens such as malaria [3, 5], lice [6], flea and tick [7], and Zika virus [3]. Despite their benefits, these compounds could pose unintended risks to the environment and human health [1]. The presence of pesticide residues in water, herbal plants, and grocery has raised serious concerns [8-10], for example, a significant increase in their

concentration in the food sample is associated with cancer [11]. Thus, identification and quantification of pesticide residues in food chain are substantial for inspection of the potential health risk. Nowadays, gas chromatography-mass spectrometry (GC-MS) with electron impact (EI) ionization and liquid chromatography-tandem mass spectrometers (LC-MS / MS) combined with electrospray ionization (ESI) technique are frequently applied as multi-residue methods for pesticides analysis [12, 13]. Certainly, for analysis of banned pesticides especially in a susceptible instance, a more sensitive technique is requisite.

Nano-liquid chromatography (nano-LC) is a miniaturized high-performance liquid chromatography (HPLC) technique introduced as the heart of this gain in sensitivity [14]. Miniaturization is one

* Corresponding Author Email: z.pahlavanyali@stu.umz.ac.ir



of the actual trends in science and technology that appears in the separation techniques as microchip devices, nano-capillary electrophoresis and nano-LC [15]. The first tendency for miniaturization of chromatography systems was revealed as nanometer-sized silica spheres stationary phase in GC [16]. Gradually, demands of life science and rapid developments of LC case to a considerable attention paid to the expansion of miniaturized LC [15]. Nano-LC column was applied for the analysis of thyroid hormones [17], modified ribonucleosides [18], biomarkers [19], drugs [20], proteins [21, 22], doping control [23] and environmental pollutants [24, 25].

Quantitative structure-retention relationship (QSRR) method is a modeling approach which quantitatively relates the chromatographic retention parameters of chemicals to their structural features [26]. Whenever the retention parameters of compounds were measured on a nano-column, extended QSRR model can explain the basic intermolecular interactions that determine the behavior of chemicals in the down-scaling column. QSRR model was employed for prediction of the retention time of some pesticides in cereals and oilseeds by multiple linear regression (MLR) and partial least square regression (PLS) methods [27]. Also, QSRR modeling was utilized for comparing the behavior of peptides in four common reversed-phase liquid chromatography (RPLC) columns. Less complexity of stepwise multiple linear regression (SW-MLR) models led to their superiority to un-informative variable elimination partial least squares (UVE-PLS) models, despite the good prediction of both equations [28].

Consensus quantitative structure-property / activity relationship (consensus QSAR / QSPR) is a novel modeling approach which is used the average of multiple predictions from individual QSAR / QSPR models [29]. It is a method for collaborative decision-making that reduces the uncertainty of the selected model used for classification [30-32] and regression [33-35] analysis. Different consensus approaches enhanced performance over the global and local models. A global QSAR / QSPR model uses a large and diverse training set covering a wide range of chemical space while a local model focuses on analog subsets [36]. The consensus models developed by combining the individual models (sub-models) differ in training subset [37], molecular descriptor [33, 34, 36], modeling method [30] or only vary in some statistical parameter

with considering their admission threshold [29]. A consensus regression model includes average consensus model (ACM) and weighted consensus model (WCM) which respectively desired property was predicted by simply or weighted averaging of sub-models [35, 36]. Genetic algorithm-multiple linear regression (GA-MLR) modeling method defines an explicit functional relationship between independent variables and desired property by optimizing forms and coefficients of equation simultaneously [29]. Consensus GA-MLR model was developed for prediction of OH tropospheric degradation of volatile organic compounds (VOCs) [33] and soil sorption partition coefficient ($\log K_{oc}$) of organic non-ionic compounds [34].

The main goal of this work is to develop a QSRR model to predict retention time of 16 banned pesticides in food sample on nano-LC column. GA-MLR method was employed for developing the global and consensus models, and their corresponding accuracy and prediction power were compared.

MATERIALS AND METHODS

Data Set

The retention times of 16 banded pesticides in food sample (organic apples and organic apple puree baby food) that their ultra-sensitive analysis was done on nano-LC-DBDI system was taken from [38]. The nano-LC experiments were performed in an Eksper nano-LC 400 (AB Sciex, Darmstadt, Germany) column packed with reverse phase C18 (3 μm \times 75 μm \times 15 cm). The names of these compounds and their experimental nano-LC retention times are indicated in Table 1. As can be seen in this table the retention value is varied in the range of 8.01 min (*Simazine*) to 17.52 min (*Phoxim*). Compounds in the data set randomly divided into the training set (12 compounds) and test set (4 compounds) that are respectively used for developing the models and evaluating their predictability.

Descriptors Calculation and Selection

The purpose of the QSRR model is to quantitatively correlate the structural features of studied chemicals to their nano-LC retention parameters by using theoretical molecular descriptors. In the first step of QSRR modeling, the structures of all studied chemicals were drawn and optimized by the semi-empirical AM1 method using Hyperchem program (version 7).

Table 1. Data set and experimental and predicted values of retention time by ACM.

NO.	Compound name	Retention time Exp. (min)	Retention time ACM (min)	Residual
1	Simazine	8.01	7.72	-0.29
2	Metolcarb	8.41	8.81	0.40
3	Dichlorvos	8.58	8.28	-0.30
4 ^a	Propoxur	8.75	8.86	0.11
5	Carbofuran	9.28	9.78	0.50
6	Bendiocarb	9.53	9.26	-0.27
7 ^a	Atrazine	9.59	9.28	-0.31
8	Metaxyl	10.00	10.55	0.55
9	Isoproturon	10.37	11.22	0.85
10 ^a	Ethiofencarb	10.45	10.12	-0.33
11	Isoprocarb	10.95	10.43	-0.52
12 ^a	Propham	11.47	11.87	0.40
13	Terbuthylazine	12.33	12.38	0.05
14	Metolachlor	14.47	13.66	-0.81
15	Malathion	14.75	14.65	-0.10
16	Phoxim	17.52	17.45	-0.07

^a test compounds in data set.

Afterward, molecular descriptors were calculated for 16 pesticides by Padel [39], Dragon [40], and Codessa [41] softwares. Then, a pre-processing step involved exclusion constant, near constant, and collinear descriptor ($r > 0.9$) was employed in QSARINS (QSAR-INSUBRIA) software [29].

RESULTS AND DISCUSSION

Models Development and Validation

The present study investigates the use of global and consensus GA-MLR models for developing a QSRR model to predict the nano-LC retention times of some banned pesticides. Also, the developed model can create an insight into the molecular interactions between bonded phases and solute molecules on the nano-scale column. The following steps of QSRR modeling were implemented in QSARINS software [29] as an new software for the development, analysis, and validation of global and consensus GA-MLR models. In this implementation, form and coefficients of GA-MLR models are simultaneously optimized by adjusting GA parameters (maximum number of descriptors = 4, generation per size = 1000, population size = 100, and mutation rate = 0.5). In each implementation, several GA-MLR models are generated varied in the number and type of descriptors or statistical parameters. A multi-criteria decision making (MCDM) was employed for the selection of best global model and sub-models of the consensus model. In MCDM, a large subset of fitness parameters including fitting

criteria, internal validation criteria, and model applicability domain (AD) simultaneously were computed and considered as features of GA-MLR models for ranking. The equation of best global QSRR model based on MCDM is shown in Table 2. Three descriptors of *MATS6c*, *SpMax2_Bhp*, and *Mor31u* appeared in this model with good statistical parameters ($R^2_{\text{training}} = 0.929$, $R^2_{\text{test}} = 0.807$ and $SE_{\text{training}} = 0.81$, $SE_{\text{test}} = 0.78$). The value of variation in inflation variance (VIF < 10) of selected descriptors in Table 2 indicates that there is no multicollinearity among them. This value is calculated from Equation (1).

$$VIF = \frac{1}{1 - R^2} \quad (1)$$

where R is correlation coefficient of multiple regression between each descriptor of the model with others [42].

Any individual QSRR model might be over-emphasized some aspects and underestimate or ignore other important features. It seems reasonable that a consensus QSRR model, which can be derived by calculating an average of sub-models, predicts the data better than the majority of individual models [34]. Several consensus models were developed by combining various sub-models selected by MCDM. Both ACM and WCM were applied for developing consensus QSRR models. Finally, an ACM which is combination of two sub-models was obtained as the best consensus QSRR model that one of them (three-descriptor

Table 2. Names, definitions, and coefficients of descriptors appearing in ACM.

	Descriptor name	Type	Notation	Coefficient	SE	VIF	ME*100
Three descriptor sub-model (global GA-MLR)							
-	-	-	Constant	5.21	-	-	-
1	Largest absolute eigenvalue of Barysz matrix - n 2 / weighted by relative polarizabilities	Barysz matrix	SpMax2_Bhp	4.92	±0.42	1.85	2.28
2	3D-MoRSE of lag 31 / un-weighted	3D-MoRSE	Mor31u	4.74	±0.51	1.56	2.19
3	Moran autocorrelation of lag 6 / weighted by gasteiger charge	2D-autocorrelation	MATS6c	3.22	±0.31	1.24	1.49
Four descriptor sub-model							
-	-	-	Constant	6.88	-	-	-
1	Crippen's LogP	Crippen logP and MR	CrippenLogP	6.10	±0.62	1.57	-3.05
2	Radial distribution function -70 / weighted by atomic mass	RDF	RDF070m	4.82	±0.51	1.79	-2.41
3	Lopping centric index	Centric index	Lop	3.65	±0.39	1.69	-1.83
4	Hydrogen bonding acceptor ability	CPSA	HASA1	-2.21	±0.31	1.73	1.11

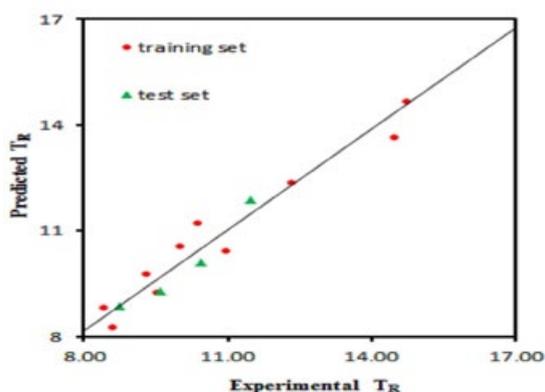


Fig. 1. Plot ACM predicted vs. experimental values of retention time.

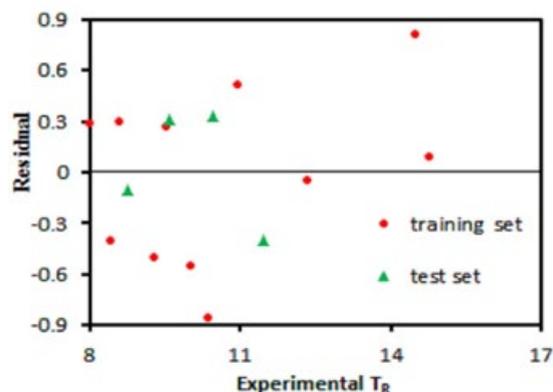


Fig. 2. Residuals of ACM predicted vs. experimental values of retention time.

sub-model) is the same best global GA-MLR model (Table 2). The predicted values of retention time with ACM are listed in Table 1.

Comparison of statistical parameters of developed models indicated that an ACM which is the combination of the best global QSRR model with a four-descriptor sub-model can be selected as the best consensus QSRR model. The statistical parameters of sub-models, WCM and ACM, are shown in Table 3. Comparison between these parameters confirmed the superiority of ACM over other models. The plot of the ACM predicted values versus experimental values of retention time are shown in Fig. 1 which indicates a good correlation among them ($R^2_{\text{training}} = 0.973$, $R^2_{\text{test}} = 0.939$ and $SE_{\text{training}} = 0.49$, $SE_{\text{test}} = 0.40$). Moreover, the residuals of predicted values were plotted against the experimental values, as shown in Fig. 2. The random distribution of residuals around the zero line indicates that there is no any systematic error

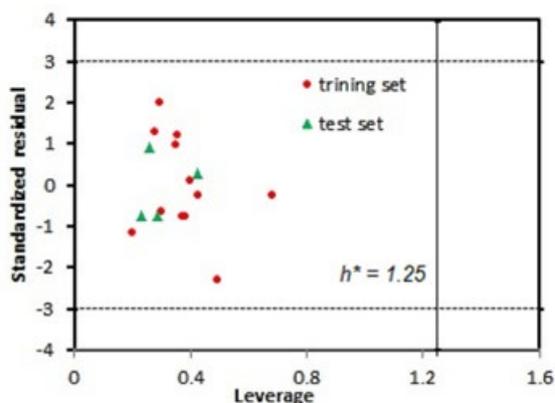
in the best-developed consensus QSRR model.

Model Validation

The fitting criteria say nothing about the stability of the developed QSRR model. Therefore, leave-one-out (LOO) cross-validation procedure is used to check the robustness of both sub-models, WCM and ACM. In LOO method the retention time of each pesticide in the data set is removed and the model is expanded on the remained chemicals, then the resulting model is employed to predict the retention time of the removed compound. This procedure is repeated for all chemicals in data set and then the cross-validated correlation coefficient (Q^2_{cv}) is calculated by Equation (2):

$$Q^2_{cv} = 1 - \frac{\sum (y_i - y_{0i})^2}{\sum (y_{0i} - y_{\text{mean}})^2} \quad (2)$$

In this equation, y_i , y_{0i} and y_{mean} are the predicted, experimental and mean values of the experimental

Fig. 3. William plot for ACM model ($h^* = 1.25$).

retention time for i th pesticide, respectively [43]. The calculated values of Q^2_{cv} for leave-one-out of selected sub-models and consensus QSRR models, shown in Table 3, are higher than 0.5 that indicated the robustness of all models.

Applicability Domain Analysis

The applicability domain of ACM models is evaluated by leverage analysis expressed by William plot. AD considers the competition features of GA-MLR models for selection as the best global equation or sub-model of consensus models [44]. It is proposed to prefer a model with the lowest number of bad prediction (Y-outlier) and chemicals far from the training structural domain (X-outlier). In the first step of obtaining William plot, the leverage (hat) value is calculated for each compound as follows:

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (3)$$

where h_i is the leverage of i th compound in the descriptor space, x_i is the descriptor raw vector, X is the matrix of the descriptor, and superscript T refers to the transpose of the vector or matrix [44]. Then, the standardized residuals of predicted Y-values were plotted against their calculated hat values. In this graph, the Y-outlier indicates the compounds with standardized residuals higher than ± 3 , while

X-outlier concerns to those with a leverage value higher than warning hat (h^*). The value of h^* in global models is constant at $\frac{3(p+1)}{n}$ that n is the number of compounds in training set and p is the number of descriptors in the model, while in consensus models, p is the maximum number of descriptors that is defined in GA parameters setting step for each implementation. William plot for ACM is shown in Fig. 3 that defines an applicability domain with all studied chemicals.

Descriptors

Theoretical molecular descriptors are distinguished by physicochemical properties as well as a mathematical tool or algorithm employed for their calculation. Three descriptors of *SpMax2_Bhp*, *Mor31u* and, *MATS6c* appeared in a three-descriptor sub-model while *CrippenLogP*, *RDF070m*, *Lop*, and *HASA1* descriptors were emerged in a four-descriptor sub-model. The values of mean effect of each descriptor in the sub-models are calculated and represented in the last column of Table 2. These values indicate the important order of descriptors in these models. *SpMax2_Bhp* is a Barysz matrix type descriptor in which the maximum absolute eigenvalue of *Barysz* matrix for $n = 2$ was weighted by polarizability [39]. The *Barysz* distance matrix is defined as a weighted distance matrix that simultaneously accounts the presence of multiple bonds and heteroatoms in the chemicals [39, 45]. In *Barysz* distance matrix diagonal elements correspond to related values of the atomic charge, polarizability and H-bond abilities which, respectively, are represented as the number 1, 2 and 3 in symbolic presentation of descriptor (e.g., polarizability in *SpMax2_Bhp*). *Mor31u* descriptor belongs to 3D-MoRSE (“molecule representation of structures based on electron diffraction”) type descriptor. [40]. In *Mor31u* a spectral representation of chemicals along a topological distance between two atoms (lag) 31 is un-weighted. These atoms are the vertices of the graph that a topological distance between them defined as a lag. *MATS6c* is a Moran autocorrelation topological structure (MATS) type

Table 3. Statistical parameters and validation of sub-models, WCM, and ACM.

Model	R^2_{training}	R^2_{test}	SE _{training}	SE _{test}	Q^2_{LOO}
Three descriptor sub-model (global GA-MLR)	0.929	0.807	0.81	0.78	0.840
Four descriptor sub-model	0.979	0.570	0.45	1.36	0.932
WCM	0.966	0.980	0.55	0.25	0.946
ACM	0.973	0.939	0.49	0.40	0.935

descriptor containing a Moran type coefficient along the lag 6 weighted by Gasteiger charge [40]. *CrippenLogP* is a Crippen's log P that is related to hydrophobicity [39]. *RDF070m* descriptor interpreted the probability distribution of atoms in a spherical volume with a radius of 7 Å as a function of atomic masses. A radial distribution function (RDF) descriptor is calculated for all atoms of a molecule to consider its interatomic distances [40, 45]. The lopping centric index (*Lop*) descriptor is a topological type descriptor defined as the mean information content which is derived from the pruning partition of acyclic graphs [40, 45]. The hydrogen bonding acceptor ability (*HASA1*) descriptor belonging to the charged partial surface area (CPSA) type descriptor reflects the solvent accessible surface area of H-bonding acceptor atoms [41].

Overall, the important structural connectivity information of the studied chemicals on the nano-LC retention time appeared alone or weighed by physicochemical properties of polarizability, Gasteiger charge, and atomic mass. Also, the significant surface properties emerged as *Crippen Log P* and *HASA1* descriptors. *Crippen Log P* is related to hydrophobicity while *HASA1* describes the charge related surface properties responsible for H-bonding formation. In RP-HPLC a solute that its atoms are incapable of formation of H-bonds is unwilling to present in water solvent (hydrophobic) [46]. Hence, steric, electrostatic, and hydrophobic parameters of the studied pesticides are the predominant interactions that are responsible for their retention time on the miniaturized RP-HPLC column.

CONCLUSION

The present study investigates the use of global and consensus GA-MLR models for developing a QSRR model to predict the nano-LC retention time of some banned pesticides. The best global QSRR model was established by adjusting GA parameters. Three descriptors of *SpMax2_Bhp*, *Mor31u*, and *MATS6c* appeared in global GA-MLR model. Consensus QSRR models were developed as ACM and WCM by the combination of a subset of the GA-MLR models. The statistical parameters confirmed the superiority of ACM over other QSRR models. *CrippenLogP*, *RDF070m*, *Lop*, and *HASA1* descriptors appeared in four-descriptor sub-model. Considering descriptors that appeared in ACM model, it was concluded that electrostatic, steric and hydrophobic interactions play the main role

in the chromatographic retention of the studied pesticides in nano-LC conditions.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests regarding the publication of this paper.

REFERENCES

- Zhang M, Zeiss MR, Geng S. Agricultural pesticide use and food safety: California's model. *Journal of Integrative Agriculture*. 2015;14(11):2340-57.
- Popp J, Pető K, Nagy J. Pesticide productivity and food security. A review. *Agronomy for Sustainable Development*. 2012;33(1):243-55.
- Bonner MR, Alavanja MCR. Pesticides, human health, and food security. *Food and Energy Security*. 2017;6(3):89-93.
- Pinstrup-Andersen P. Food security: definition and measurement. *Food Security*. 2009;1(1):5-7.
- Rehewagen C. WHO recommends DDT to control malaria. *BMJ*. 2006;333(7569):622.3.
- Franklin LU, Cunnington GD, Young DE. Terpene based pesticide treatments for killing terrestrial arthropods including, amongst others, lice, lice eggs, mites and ants. Google Patents; 2000.
- Dashtbozorgi Z, Golmohammadi H, Konoz E. Support vector regression based QSPR for the prediction of retention time of pesticide residues in gas chromatography-mass spectroscopy. *Microchemical Journal*. 2013;106:51-60.
- Madureira FD, da Silva Oliveira FA, de Souza WR, Pontelo AP, de Oliveira MLG, Silva G. A multi-residue method for the determination of 90 pesticides in matrices with a high water content by LC-MS/MS without clean-up. *Food Additives & Contaminants: Part A*. 2012;29(4):665-78.
- Gómez-Ramos MM, Ferrer C, Malato O, Agüera A, Fernández-Alba AR. Liquid chromatography-high-resolution mass spectrometry for pesticide residue analysis in fruit and vegetables: Screening and quantitative studies. *Journal of Chromatography A*. 2013;1287:24-37.
- Gómez-Pérez ML, Plaza-Bolaños P, Romero-González R, Martínez-Vidal JL, Garrido-Frenich A. Comprehensive qualitative and quantitative determination of pesticides and veterinary drugs in honey using liquid chromatography-Orbitrap high resolution mass spectrometry. *Journal of Chromatography A*. 2012;1248:130-8.
- Wang Q-L, Zhang Y-J, Zhou C-X, Zhang JIE, Dou YE, Li Q-Q. Risk assessment of mouse gastric tissue cancer induced by dichlorvos and dimethoate. *Oncology Letters*. 2013;5(4):1385-9.
- Alder L, Greulich K, Kempe G, Vieth B. Residue analysis of 500 high priority pesticides: Better by GC-MS or LC-MS/MS? *Mass Spectrometry Reviews*. 2006;25(6):838-65.
- Thurman EM, Ferrer I, Pozo OJ, Sancho JV, Hernandez F. The even-electron rule in electrospray mass spectra of pesticides. *Rapid Communications in Mass Spectrometry*. 2007;21(23):3855-68.
- Remco LRE-JS. Nano LC: Principles, Evolution, and State-of-the-Art of the Technique. 2011.
- Gama MR, Collins CH, Bottoli CBG. Nano-Liquid Chromatography in Pharmaceutical and Biomedical Research. *Journal of Chromatographic Science*. 2013;51(7):694-703.

16. Kirkland JJ. Porous Thin-Layer Modified Glass Bead Supports for Gas Liquid Chromatography. *Analytical Chemistry*. 1965;37(12):1458-61.
17. Ruuskanen S, Hsu B-Y, Heinonen A, Vainio M, Darras VM, Sarraude T, et al. A new method for measuring thyroid hormones using nano-LC-MS/MS. *Journal of Chromatography B*. 2018;1093-1094:24-30.
18. Sarin LP, Kienast SD, Leufken J, Ross RL, Dziergowska A, Debiec K, et al. Nano LC-MS using capillary columns enables accurate quantification of modified ribonucleosides at low femtomol levels. *RNA*. 2018;24(10):1403-17.
19. Bai H-Y, Lin S-L, Chung Y-T, Liu T-Y, Chan S-A, Fuh M-R. Quantitative determination of 8-isoprostaglandin F2 α in human urine using microfluidic chip-based nano-liquid chromatography with on-chip sample enrichment and tandem mass spectrometry. *Journal of Chromatography A*. 2011;1218(15):2085-90.
20. Fanali S, Aturki Z, D'Orazio G, Rocco A. Separation of basic compounds of pharmaceutical interest by using nano-liquid chromatography coupled with mass spectrometry. *Journal of Chromatography A*. 2007;1150(1-2):252-8.
21. Dionisio G, Uddin M, Vincze E. Enrichment and Identification of the Most Abundant Zinc Binding Proteins in Developing Barley Grains by Zinc-IMAC Capture and Nano LC-MS/MS. *Proteomes*. 2018;6(1):3.
22. Li B, Takahashi D, Kawamura Y, Uemura M. Plasma Membrane Proteomics of Arabidopsis Suspension-Cultured Cells Associated with Growth Phase Using Nano-LC-MS/MS. In: Mock H-P, Matros A, Witzel K, editors. *Plant Membrane Proteomics: Methods and Protocols*. New York, NY: Springer New York; 2018. p. 185-94.
23. Zhu KY, Leung KW, Ting AKL, Wong ZCF, Ng WYY, Choi RCY, et al. Microfluidic chip based nano liquid chromatography coupled to tandem mass spectrometry for the determination of abused drugs and metabolites in human hair. *Analytical and Bioanalytical Chemistry*. 2012;402(9):2805-15.
24. Buonasera K, D'Orazio G, Fanali S, Dugo P, Mondello L. Separation of organophosphorus pesticides by using nano-liquid chromatography. *Journal of Chromatography A*. 2009;1216(18):3970-6.
25. Asensio-Ramos M, D'Orazio G, Hernandez-Borges J, Rocco A, Fanali S. Multi-walled carbon nanotubes-dispersive solid-phase extraction combined with nano-liquid chromatography for the analysis of pesticides in water samples. *Analytical and Bioanalytical Chemistry*. 2011;400(4):1113-23.
26. Mardia K. V., Kent J. T., Bibby J. M. *Multivariate analysis*: Academic Press, London.; 1979.
27. LI M-p, ZHANG S-w. Quantitative Structure-Retention Relationship of Gas Chromatography for Organophosphorus Pesticide Residues in Cereals and Oilseeds [J]. *Journal of Shanxi University (Natural Science Edition)*. 2011; 3: 026
28. Bodzioch K, Durand A, Kaliszan R, Bączek T, Vander Heyden Y. Advanced QSRR modeling of peptides behavior in RPLC. *Talanta*. 2010;81(4-5):1711-8.
29. Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *Journal of Computational Chemistry*. 2013;34(24):2121-32.
30. Ganguly M, Brown N, Schuffenhauer A, Ertl P, Gillet VJ, Greenidge PA. Introducing the Consensus Modeling Concept in Genetic Algorithms: Application to Interpretable Discriminant Analysis. *Journal of Chemical Information and Modeling*. 2006;46(5):2110-24.
31. Afantitis A, Leonis G, Gambari R, Melagraki G. Consensus Predictive Model for Human K562 Cell Growth Inhibition through Enalos Cloud Platform. *ChemMedChem*. 2018;13(6):555-63.
32. Lei T, Li Y, Song Y, Li D, Sun H, Hou T. ADMET evaluation in drug discovery: 15. Accurate prediction of rat oral acute toxicity using relevance vector machine and consensus modeling. *Journal of Cheminformatics*. 2016;8(1).
33. Gramatica P, Pilutti P, Papa E. Validated QSAR Prediction of OH Tropospheric Degradation of VOCs: Splitting into Training-Test Sets and Consensus Modeling. *Journal of Chemical Information and Computer Sciences*. 2004;44(5):1794-802.
34. Gramatica P, Giani E, Papa E. Statistical external validation and consensus modeling: A QSPR case study for Koc prediction. *Journal of Molecular Graphics and Modelling*. 2007;25(6):755-66.
35. Li J, Lei B, Liu H, Li S, Yao X, Liu M, et al. QSAR study of malonyl-CoA decarboxylase inhibitors using GA-MLR and a new strategy of consensus modeling. *Journal of Computational Chemistry*. 2008;29(16):2636-47.
36. Lei B, Xi L, Li J, Liu H, Yao X. Global, local and novel consensus quantitative structure-activity relationship studies of 4-(Phenylaminomethylene) isoquinoline-1, 3 (2H, 4H)-diones as potent inhibitors of the cyclin-dependent kinase 4. *Analytica Chimica Acta*. 2009;644(1-2):17-24.
37. Khairullina VR, Gerchikov AY, Lagunin AA, Zarudii FS. QSAR Modelling of Thymidylate Synthase Inhibitors in a Series of Quinazoline Derivatives. *Pharmaceutical Chemistry Journal*. 2018;51(10):884-8.
38. Mirabelli MF, Wolf J-C, Zenobi R. Pesticide analysis at ppt concentration levels: coupling nano-liquid chromatography with dielectric barrier discharge ionization-mass spectrometry. *Analytical and Bioanalytical Chemistry*. 2016;408(13):3425-34.
39. Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*. 2010;32(7):1466-74.
40. Todeschini R, al. e. DRAGON Web version 3.0: Milano, Italy; 2003.
41. Katritzky A, Lobanov V, Karelson M. CODESSA: Training Manual University of Florida Gainesville. FL Google Scholar. 1995.
42. Ren Y, Zhang Y, Yao X. QSPRs for estimating nematic transition temperatures of pyridine-containing liquid crystalline compounds. *Liquid Crystals*. 2017;45(2):238-49.
43. Wold S, L. Eriksson, S. Clementi. Statistical Validation of QSAR Results. *Chemometric Methods in Molecular Design*1995. p. 309-38.
44. Atkinson AC. Plots, transformations and regression; an introduction to graphical methods of diagnostic regression analysis. 1985.
45. Todeschini R, Consonni V. *Handbook of molecular descriptors*: John Wiley & Sons; 2008.
46. Abraham MH, Ross M, Poole CF, Poole SK. HYDROGEN BONDING. 42. Characterization of Reversed-Phase High-Performance Liquid Chromatographic C18 Stationary Phases. *Journal of Physical Organic Chemistry*. 1997;10(5):358-68.