We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

2D- and 3D-QSRR Studies of Linear Retention Indices for Volatile Alkylated Phenols

Assia Belhassan, Samir Chtita, Tahar Lakhlifi and Mohammed Bouachrine

Abstract

In this study, 29 volatile alkylated phenols were subjected to a quantitative structure retention relationships (QSRR) studies; we have developed two- and three-dimensional quantitative structure retention relationships (2D- and 3D-QSRR) for this series; and these molecules were subjected to a 2D-QSRR analysis for their retention property using stepwise multiple linear regression (MLR) and 3D-QSRR analysis using partial least squares (PLS). The 28 descriptors are calculated for the 29 molecules using the ChemOffice and ChemSketch software to construct 2D-QSRR model. The 3D-QSRR models were constructed using comparative molecular field analysis (CoMFA) method. The models were used to predict the linear retention indices of the test set compounds, and agreement between the experimental and predicted values was verified. The statistical results indicate that the predicted values are in good agreement with the experimental results ($r^2 = 0.980$; $r^2_{CV} = 0.977$ and $r^2 = 0.998$; $r^2_{CV} = 0.959$ for MLR and CoMFA methods, respectively). To validate the predictive power of the resulting models, external validation multiple correlation coefficient was calculated; in addition to a performance prediction power, this coefficient has a favorable estimation of stability for the two methods ($r_{test} = 0.938$ and $r_{test} = 0.955$ for MLR and CoMFA methods, respectively).

Keywords: quantitative structure retention relationship, linear retention indices, multiple linear regression, molecular field analysis, external validation, alkylated phenols

1. Introduction

Phenols are widely present in the environment as building blocks for plants [1]. They are formed naturally from decomposition of leaves and wood as well as through human activity like water purification processes [2]. Alkylphenols are a family of organic compounds obtained by the alkylation of phenols. The term is usually reserved for major industrial compounds such as propylphenol, amylphenol, heptylphenol, octylphenol, nonylphenol, dodecylphenol, and other long-chain carbon compounds. Methylphenols and ethylpenols are also alkylphenols, but are more often referred to by their specific names, cresols and xylenols, respectively. The alkylated phenols have a good ability to be adsorbed on

solid materials and some are toxic to fish and other forms of aquatic environment. Very low concentrations of these molecules have unfavorable effects on the taste and odor of water and fish [3].

All phenolic compounds can be considered as important parameters of the organoleptic (color, flavor, and aroma) and nutritional qualities of food products. The phenolic compounds which participate in the vegetable aroma are relatively simple volatile compounds whose odors can be pleasant or unpleasant. Vanilla, for example, is the most popular aroma in the world, and its production is estimated at 1500 tons per year [4]. Approximately 250 compounds are responsible for vanilla aroma and among these are about 20 phenolic compounds, the most abundant of which are vanillin, p-hydroxybenzaldehyde, and vanillic acid [5]. The spices we use to enhance taste and flavor of food contain volatile compounds characterized by the presence of a methoxyl group. 4-vinyl guaiacol is responsible for the pleasant odors that occur during the manufacture and storage of citrus juices (orange and grapefruit in particular). This compound is formed from the degradation of ferulic acid, and the quality of the orange juice aroma is directly related to changes in free ferulic acid and 4-vinyl guaiacol contents [6]. These two compounds are also produced during the thermal degradation of lignin. With their derivatives (4-methyl guaiacol, 4-ethyl guaiacol, vanillin, vanillic acid, etc.), they are at the origin of the aroma developed by the smoking techniques used in meat and fish conservation [7].

Some alkylated phenols represent another group of compounds with a constantly weak odor. In addition, some individual odorants in this group have been described in several studies as having various sensory properties. Because of their obviously high odor potency, the odor thresholds of the alkylated phenols have been extensively evaluated.

The multidimensional quantitative structure-activity/property relationship (multidimensional-QSAR/QSPR) analysis is a computational method used to predict biological activities or chemical properties of existing or supposed chemical compounds. With incessant development, the multidimensional-QSAR/QSPR analyses have made notable achievement in diverse fields, such as toxicology and medicinal chemistry [8, 9]. Through the fast progress of computer science and theoretical study, it can quickly and accurately find molecular information (chemical descriptors) of compounds by computation. These chemical descriptors used in the construction of the QSAR/QSPR models can increase the interpretability and can predict the activity/property of new molecules [10].

The release of odorant molecules from a solid or liquid medium and their passage in the vapor phase is the first step before a possible perception due to the activation of the olfactory receptors present in the nasal cavity followed by a series of complex neurophysiological reactions, in order to code a particular smell, that's why in this study, a series of 29 volatile alkylated phenols, including monoalkylated phenols and di- and trimethylphenols, were subjected to a quantitative structure retention relationships (QSRR) studies, we have developed two- and three-dimensional quantitative structure retention relationships (2D- and 3D-QSRR) for a series of 29 molecules odorants based on phenol. We construct 2D-QSRR model using 28 descriptors. The 3D-QSAR/QSPR models were constructed using the comparative molecular field analysis (CoMFA) [11] tools that collect and interpret complex data from series of bioactive molecules to construct computational models that correlate chemical properties with biological activity/propriety [12]. Through this approach, molecular features responsible for the retention property of the investigated compounds (alkylated phenols) were identified using the CoMFA contour plots. Furthermore, the statistical consistency of the developed models was evaluated on the basis of their correlation ability for the training set, as well as their

predictive power for an external test set. We accordingly propose quantitative models, using stepwise multiple linear regression (MLR) for 2D-QSRR analysis and the partial least squares (PLS) for 3D-QSRR model, and we try to interpret the retention property of the compounds relying on the multidimensional-QSRR analyses [13].

2. Material and methods

2.1 2D-QSRR study

2.1.1 Data set

The reliability of the 2D-QSRR analysis is depending on the available data set, and the method of analysis and the validations. In the present analysis, a series of 29 selected alkylated phenols that have been evaluated for their linear retention indices was taken from literature, and as reported in the literature [14], high-resolution GC/O (HRGC/O) analyses were performed with a type 5160 gas chromatograph (Carlo Erba), and the analyses were accomplished using DB-1701, as demonstrated by Czerny et al. [14]. We considered to carry out the 2D-QSRR analysis: 24 molecules are selected to propose the quantitative model (training set) and 5 compounds that have been selected randomly and were not used in training set have served to test the performance of the proposed model (test set). **Table 1** shows the studied compounds and the experimental linear retention indices values (LRI).

No.	Compound	LRI	Log(LRI)	No.	Compound	LRI	Log(LRI)
1[a]	Phenol	1167	3.067	16	4-n-hexylphenol	1784	3.251
Monoalkylated phenols				17	4-n-heptylphenol	1886	3.276
2	2-Methylphenol	1244	3.095	18	4-n-octylphenol	1994	3.300
3	3-Methylphenol	1270	3.104	**Dimethylated phenols**			
4	4-Methylphenol	1269	3.103	19	2,3-Dimethylphenol	1387	3.142
5	2-Ethylphenol	1330	3.124	20	2,4-Dimethylphenol	1344	3.128
6	3-Ethylphenol	1371	3.137	21	2,5-Dimethylphenol	1342	3.128
7	4-Ethylphenol	1369	3.136	22	2,6-Dimethylphenol	1300	3.114
8	2-n-propylphenol	1415	3.151	23	3,4-Dimethylphenol	1406	3.148
9[a]	3-n-propylphenol	1463	3.165	24	3,5-Dimethylphenol	1369	3.136
10	4-n-propylphenol	1463	3.165	**Trimethylated phenols**			
11[a]	2-Isopropylphenol	1414	3.150	25	2,3,5-Trimethylphenol	1483	3.171
12	3-Isopropylphenol	1418	3.152	26	2,3,6-Trimethylphenol	1440	3.158
13	4-Isopropylphenol	1419	3.152	27	2,4,5-Trimethylphenol	1472	3.168
14	4-n-Butylphenol	1571	3.196	28[a]	2,4,6-Trimethylphenol	1400	3.146
15	4-n-Pentylphenol	1678	3.225	29[a]	3,4,5-Trimethylphenol	1551	3.191

LRI: linear retention indices.
[a]Test set.

Table 1.
Alkylated phenols used in this study and their experimental linear retention indices.

2.1.2 Molecular descriptors generation

Twenty-eight molecular descriptors were calculated using ACD/ChemSketch and ChemOffice programs [15, 16] to predict the correlation between these descriptors and the retention property of studied compounds and to develop a linear model [17]. The descriptors used in this study are displayed in **Table 2**.

2.1.3 Statistical analysis

To explain the structure-property relationship, 28 descriptors are calculated for the 29 molecules using the ChemOffice and ChemSketch software, and they were subjected to a stepwise multiple linear regression (MLR) available in the SPSS software [18]. The stepwise MLR was generated to predict retention property values Log(LRI). Equation was justified by the correlation coefficient (r), the root mean square of the errors (RMSE), the Fishers F-statistic (F), and the significance level (*P*-value) [19].

The final stage of this 2D-QSRR analysis consists of statistical validation in order to assess the significance of the model and hence its ability to predict property of other compounds. In this chapter, the model was validated internally by the cross-validation test. The cross validations are statistical techniques in which different proportions of chemicals are iteratively held out from the training set used for model development. In this chapter, the leave-one-out procedure is used; this process sequentially removes one compound from the training set containing 24 compounds. A 2D-QSRR model is created on a "23" set of molecules, and the molecule removed is predicted by the constructed model. This process is repeated "24" times in order to predict the retention property of all compounds [20].

2.2 3D-QSRR study

2.2.1 Minimization and alignment

Chemical structures of studied compounds were sketched with sketch module in SYBYL [21] and minimized using Tripos force field [22] with the Gasteiger-Hückel charges [23] and conjugated gradient method, and gradient convergence criteria of 0.01 kcal/mol. Simulated annealing on the energy minimized structures was performed with 20 cycles.

Molecular alignment is one of the most sensitive parameters in 3D-QSRR methods. In this work, all studied compounds were aligned on the common core (compound no. 1), using the simple alignment method in Sybyl [24]. Compound

Software	Descriptors
ChemOffice	Melting point T (Kelvin); molecular weight MW (g/mol); critical temperature CT (Kelvin); heat of formation H° (kJ mol^{-1}); boiling point TB (Kelvin); Gibbs free energy G (kJ mol^{-1}); critical pressure CP (Bar); Connolly solvent-excluded volume V (A°)^3; shape coefficient I; total connectivity TC; Log P; number of rotatable bonds NRB; winner index (W); number of H-bond acceptors (NHA); molecular topological index MTI; number of H-bond donors (NHD); partition coefficient PC; Balaban index (J); Henry's law constant KH; polar surface area PSA (A°)^2; total valence connectivity TVC; sum of valence degrees SVD
ChemSketch	Percent ratios of nitrogen, hydrogen, oxygen, and carbon atoms (H$_\%$; O$_\%$; C$_\%$); surface tension γ (dyne/cm); index of refraction (n); density (d)

Table 2.
Descriptors selected and software packages used in the calculation of descriptors.
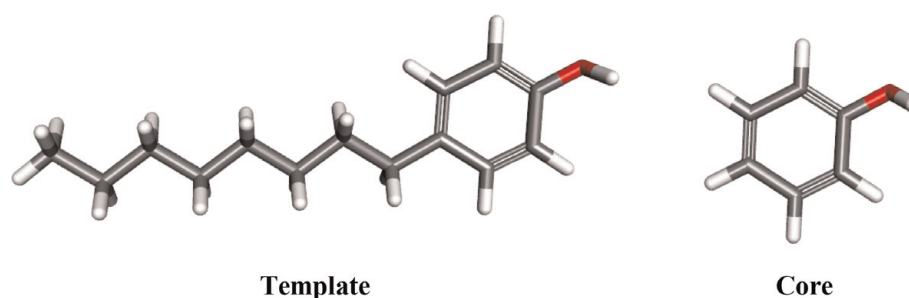
Template   **Core**

Figure 1.
3D structure of the core (molecule no. 1) and the template (molecule no. 18).

no. 18, which was the most active compound (with highest Log(LRI)), was used as template (**Figure 1**).

2.2.2 CoMFA studies

Based on the molecular alignment, CoMFA studies were performed to analyze the specific contributions of steric and electrostatic effects. These interactions were calculated using the Tripos force field with a distance-dependent dielectric constant at all interactions in a regularly spaced (2 Å) grid taking a sp^3 carbon atom as steric probe and a +1 charge as electrostatic probe. The cutoff was set to 30 kcal/mol [25]. With standard options for scaling of variables, the regression analysis was carried out using the fully cross-validated partial least squares (PLS) method (leave one out) [26]. The final model that is non–cross-validated conventional analysis was developed with the optimum number of components to yield a non–cross-validated r^2 value.

2.2.3 Partial least squares analysis (PLS) and validation

The 3D-QSRR models were generated using a training set of 24 molecules. Predictive power of the resulting models was evaluated using a test set of five molecules (**Table 1**). The test compounds have been selected randomly. PLS analysis used to construct the 3D-QSRR models is an extension of multiple regression analysis in which the initial variables are replaced by optimum number of components of their linear combinations. PLS statistical method with leave-one-out (LOO) cross-validation procedure was used in this work to determine the optimal numbers of components considering cross-validated coefficient r_{CV} for the training set of 24 molecules. The external validation of created models was determined using five compounds (test set). The final analysis (non-cross-validated analysis) was carried out using the optimum number of components obtained from the cross-validation analysis to get correlation coefficient r^2 [27, 28].

3. Results and discussions

3.1 2D-QSRR study

3.1.1 Data set for analysis

A 2D-QSRR study was carried out for a series of 29 alkylated phenols, as indicated above, to determine a quantitative relationship between the structure and the

retention property. The values of the 28 descriptors are shown in Table S1 (in Supplementary Material).

3.1.2 Stepwise multiple linear regression MLR

The stepwise multiple linear regression (MLR) procedure based on the forward selection and backward elimination method (including the critical probability: P-value <0.05 for all descriptors and for the model complete) was employed to determine the best regression model.

The 2D-QSRR model built using stepwise MLR is represented by the following equation:

$$Log(LRI) = 2.935 + 1.682 \times 10^{-3} \times V \qquad (1)$$

N = 24; r = 0.990; r^2 = 0.980; RMSE = 0.008; F = 1085.981; $P < 0.0001$.

In this equation, V is the Connolly solvent-excluded volume, N is the number of compounds, r is the correlation coefficient, r^2 is the coefficient of determination, RMSE is the root mean square of the errors, F is the Fisher's criterion, and P is the significance level.

It is observed that the coefficient of correlation r is high, and RMSE is low, which makes it possible to indicate that the model is reliable. A P value much smaller than 0.05 indicates that the regression equation is statistically significant; thus, we can conclude, with confidence, that the model provides a significant amount of information [29, 30].

The predicted Log(LRI) values calculated from equation are given in **Table 4** in comparison to the observed values. The correlation between the predicted and observed Log(LRI) is shown in **Figure 2**.

3.1.3 Internal validation (cross-validation)

The 2D-QSRR model expressed by the equation of stepwise MLR method is validated by its appreciable value of r^2_{CV} obtained using the leave-one-out (LOO) procedure. The value of r^2_{CV} greater than 0.5 is the basic condition for qualifying a
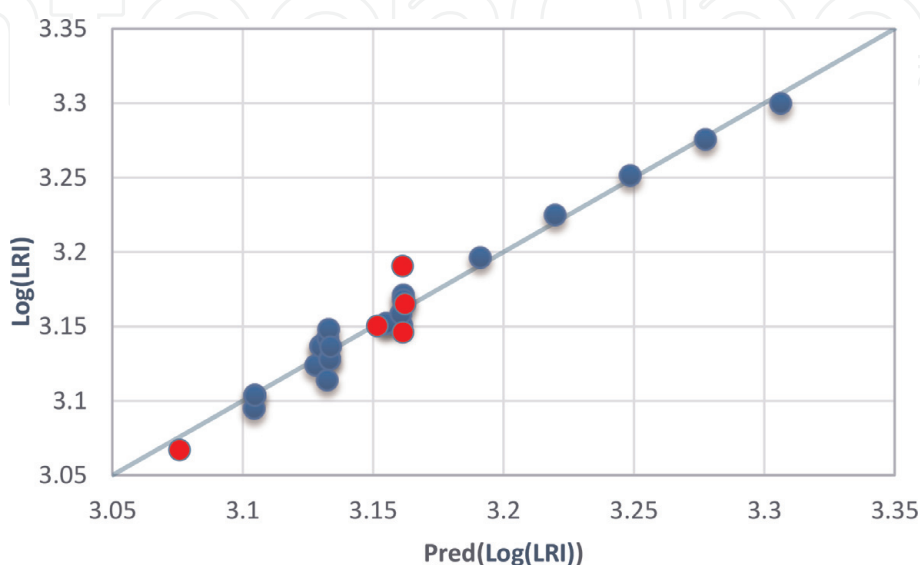
Figure 2.
Correlations of observed and predicted Log(LRI) with MLR stepwise (training set in blue; test set in red).

2D-QSRR model as valid. The model's performance was good and was characterized by r^2_{CV} value of 0.977 with the descriptor (V) proposed by the stepwise MLR.

3.1.4 External validation

The model created in the calculation process using the alkylated phenols is used to predict the retention property values (Log(LRI)) of the remaining (five molecules). The results obtained by stepwise MLR model are very sufficient to conclude the performance of models; it is confirmed by the test done with the five compounds (r_{test} = 0.938; r^2_{test} = 0.880).
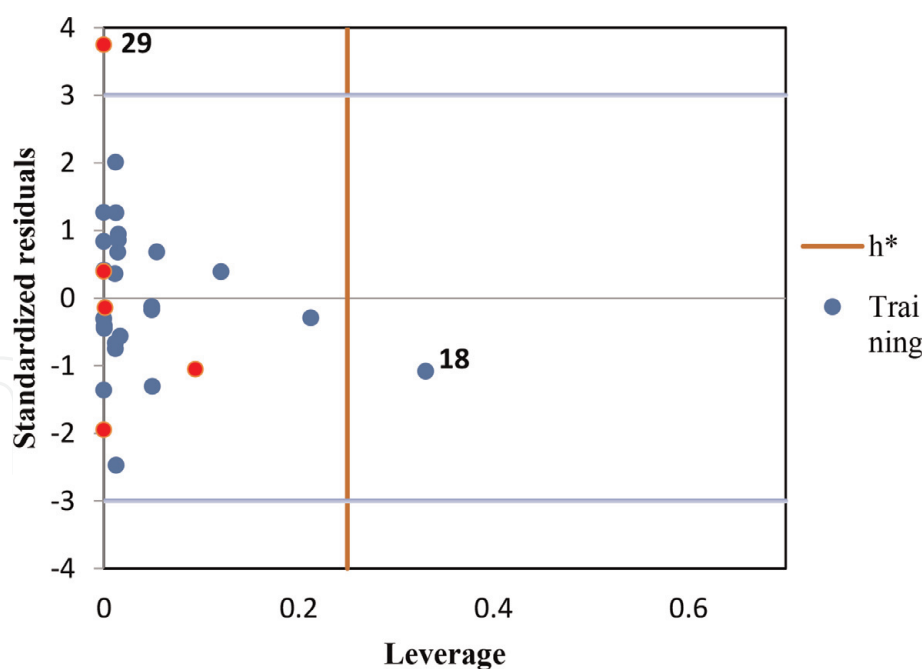
3.1.5 Domain of applicability

Evaluation of the applicability domain of the 2D-QSRR model is considered as an important step to establish that the model is reliable to make predictions within the chemical space for which it was developed [31]. In this chapter, we used leverage approach [20]. Leverage of a given chemical compound h_i is defined as follows:

$$h_i = x_i^T(X^TX)^{-1}x_i \, (i = 1 \ldots n)$$ (2)

where x_i is the descriptor row of the query compound and X is the descriptor matrix of the training set compounds used to develop the model. As a prediction tool, the warning leverage h^* is defined as follows:

$$h^* = 3(P + 1)/n$$ (3)

Figure 3.
Williams plot to evaluate the applicability domain of stepwise MLR model.

	r^2_{CV}	r^2	SE	F-t	N	r^2_{test}
CoMFA	0.959	0.998	0.003	2937.327	7	0.913

Table 3.
Statistical parameters of CoMFA model.

where n is the number of training compounds and *P* is the number of descriptors in the model.

From the Williams plot (**Figure 3**), it is obvious that all the compounds in the data set are within the applicability domain of the model (the warning leverage limit is 0.250) except one training compound (no. 18); these compounds have their

No	Log(LRI) (obs.)	Log(LRI) (calc.)			
		2D-QSRR		3D-QSRR	
		MLR	Residual	CoMFA	Residual
1[a]	3.067	3.076	−0.009	3.062	0.005
2	3.095	3.104	−0.010	3.094	0.001
3	3.104	3.105	−0.001	3.103	0.001
4	3.103	3.105	−0.001	3.100	0.003
5	3.124	3.128	−0.004	3.124	0.000
6	3.137	3.130	0.007	3.141	−0.004
7	3.136	3.130	0.006	3.137	−0.001
8	3.151	3.161	−0.010	3.149	0.002
9[a]	3.165	3.162	0.003	3.150	0.015
10	3.165	3.162	0.003	3.163	0.002
11[a]	3.150	3.152	−0.001	3.150	0.000
12	3.152	3.155	−0.003	3.147	0.005
13	3.152	3.155	−0.003	3.150	0.002
14	3.196	3.191	0.005	3.196	0.000
15	3.225	3.220	0.005	3.226	−0.001
16	3.251	3.249	0.003	3.250	0.001
17	3.276	3.277	−0.002	3.280	−0.004
18	3.300	3.306	−0.007	3.298	0.002
19	3.142	3.133	0.009	3.137	0.005
20	3.128	3.133	−0.005	3.132	−0.004
21	3.128	3.133	−0.006	3.127	0.001
22	3.114	3.132	−0.018	3.128	−0.014
23	3.148	3.133	0.015	3.144	0.004
24	3.136	3.134	0.003	3.136	0.000
25	3.171	3.162	0.010	3.168	0.003
26	3.158	3.161	−0.002	3.170	−0.012
27	3.168	3.162	0.006	3.167	0.001
28[a]	3.146	3.161	−0.015	3.164	−0.018
29[a]	3.191	3.161	0.029	3.175	0.016

LRI: linear retention indices.
[a]Test set.

Table 4.
Actual and predicted Log(LRI) along with residual of training and test sets using stepwise MLR and CoMFA models.

leverage values greater than the warning h* value and could be high leverage compound influencing the performance of the model. However, their standard residual values are very low and within the established limit [32]. As a result, this compound could be considered as influential in fitting the model performance but not necessarily outliers to be deleted from the training dataset, and thus, the model can be applied with confidence within the defined applicability domain.

For all the compounds in the training and test sets, their standardized residuals are smaller than three standard deviation units $(3 \pm \delta)$ except one test compound (No 29). Thus, compound no. 29 can be as outlier. Because this compound is one of the test set compounds, there is no need to remove this compound from the data set.

Therefore, the predicted of linear retention indices values (Log(LRI)) by the developed stepwise MLR model is reliable.

3.2 3D-QSRR study

3.2.1 Molecular alignment

All other compounds were aligned on the basis of the common structure (compound no. 1). Alignment of training and test set compounds using distill module is shown in **Figure 4**.

3.2.2 CoMFA result

The 3D-QSRR models were obtained from the CoMFA analysis, and its statistical parameters are listed in **Table 3**. The values of predicted Log(LRI) are calculated by CoMFA model, and the observed values are given in **Table 4**. The correlations of predicted and observed Log(LRI) values are illustrated in **Figure 5**.

We use cross-validation as an internal test of the quality of the PLS models. And to evaluate the predictive power of a QSRR model (external test), the Log(LRI) of the remained set of five molecules (test set) are deduced from the constructed model with the 24 compounds (training set) by CoMFA model (**Table 3**).

where r^2_{CV} is the square of the LOO cross-validation (CV) coefficient; r^2 is the square of the non-CV coefficient; SE is the standard error of prediction; F is the F-test value; N is the optimum number of components; and r^2_{test} is the external validation correlation coefficient for test set compounds.

The 3D-QSRR models gave good statistical results in terms of r^2 value ($r^2 = 0.998$) for the CoMFA model. This approach has good predictive capability
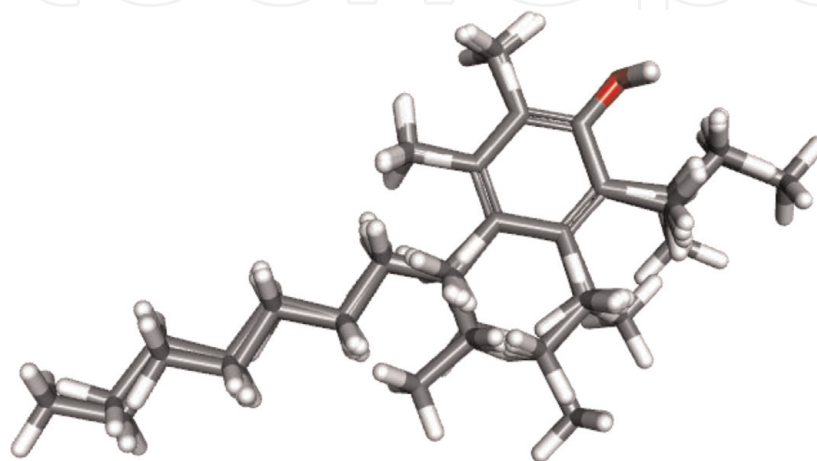
Figure 4.
3D-QSRR structure superposition and alignment of training set using molecule no. 18 as a template.
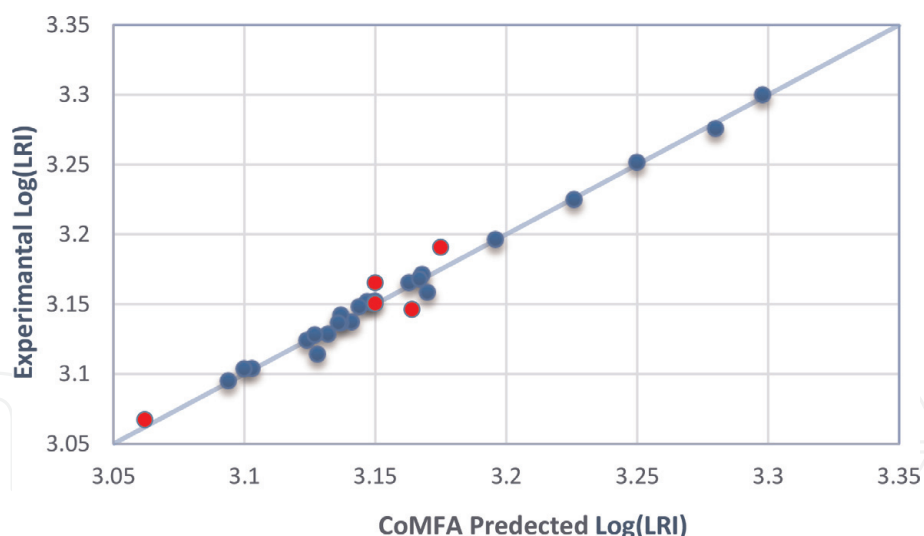
Figure 5.
Correlations of observed and predicted Log(LRI) derived from CoMFA model (training set in blue; test set in red).

2D-QSRR	3D-QSRR
MLR	PLS
Stepwise	CoMFA
$r = 0.990$; $r^2 = 0.980$	$r = 0.999$; $r^2 = 0.998$
$r_{CV} = 0.988$; $r^2_{CV} = 0.977$	$r_{CV} = 0.979$; $r^2_{CV} = 0.959$
$r_{test} = 0.938$; $r^2_{test} = 0.880$	$r_{test} = 0.955$; $r^2_{test} = 0.913$

Table 5.
Statistical parameters for the stepwise MLR and PLS models using multidimensional-QSRR analyses.

gives good results ($r^2_{CV} = 0.956$). The model was able to establish a satisfactory relationship between the molecular descriptors and the linear retention indices of the studied compounds. The results obtained by CoMFA analysis are sufficient to conclude the performance of the model; it is confirmed by the test done with the five compounds (**Table 3**).

3.3 Comparison between 2D- and 3D-QSRR results and design of novel alkylated phenols

Aiming to provide a comparison among the stepwise MLR and CoMFA models, **Table 5** lists the main statistical indicators for 2D- and 3D-QSRR models.

A comparison of the quality of stepwise MLR and CoMFA model (**Table 5**) shows that the two approaches stepwise MLR and CoMFA have better predictive capability gives better results. Stepwise MLR and CoMFA models were able to found a suitable relationship between the chemical descriptors and the linear retention indices of the studied molecules.

Multidimensional-QSRR correlates retention property with the physicochemical and structural descriptors of a series of molecules. It has been habitually used to predict retention of new molecules and to propose molecules with preferred properties. The constructed models can be used for the designing of new alkylated phenols with higher or lower property values (Log(LRI)).

In this way, we can design new compounds by adding suitable substituents and calculate their property using stepwise MLR equation. The stepwise MLR equation indicated the positive correlation of the Connolly solvent-excluded volume (V).

The obtained results show that, to increase propriety of alkylated phenols, we will increase Connolly solvent-excluded volume (V) value of these molecules. Moreover, to decrease property, we will decrease the descriptor (V) value, by adding suitable substituents, and calculate their property using the regression equation. This study consists of the first step explored to code a particular odor of this group of molecules, followed by docking molecular study that allows understand the mechanism of activation of olfactory receptor present in the nasal cavity by this type of molecules.

We can also use the results of 3D-QSRR to design new alkylated phenols with higher or lower retention property values (Log(LRI)). The CoMFA contour plots were able to identify that molecular fragments, functional groups, and physicochemical properties strongly correlated with the linear retention indices of this series. CoMFA steric and electrostatic contours are shown in **Figure 6**.

The steric interaction is represented by green and yellow contours, while electrostatic interaction is denoted by red and blue contours. The green region around the 2, 3, 4, 5, and 6 positions (the carbon to which the initial ▬OH is bonded is counted as the first position) (**Figure 6a**) indicates that bulky groups are favored and they might increase the property. That can explain very well why the property of the alkylated phenols with a group bigger than Et group (case of compounds 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, and 18) is higher than those of other compounds. We can also explain, that for the alkylated phenols with a group smaller than Pr group, on the one hand, the property of dimethylated phenols is higher than those of monoalkylated phenols and, on the other hand, the property of the trimethylated phenols is higher than those of dimethylated phenols and monoalkylated phenols. The bigger green region is observed around the four positions in comparison with the other positions, suggesting that groups with steric tolerance are required at this position to reach the green area, which means to increase the property, this fact can
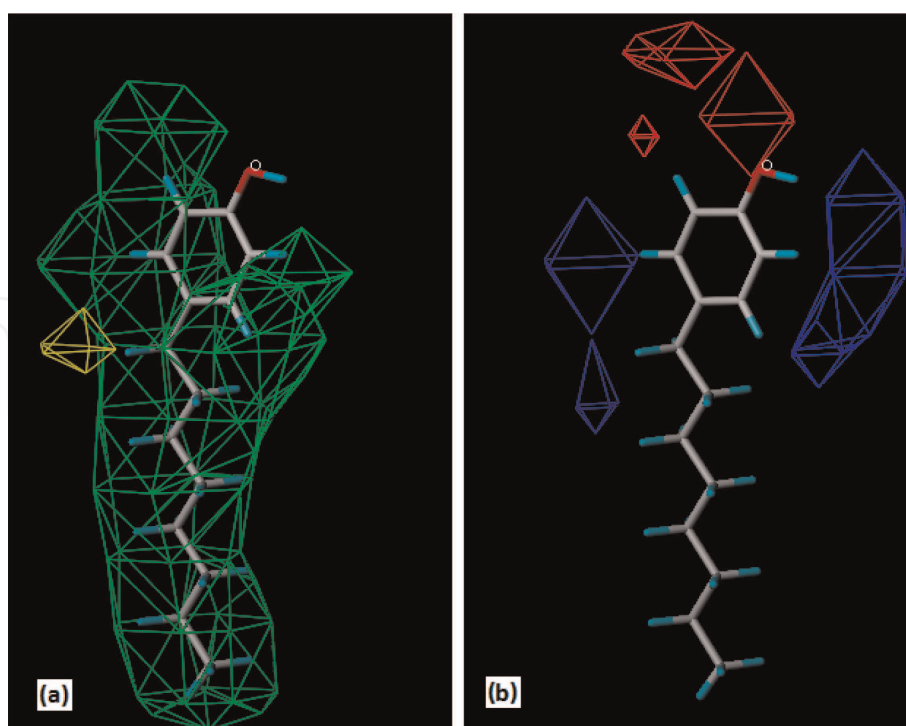
Figure 6.
(a) Std* coeff. contour maps of CoMFA analysis with 2 Å grid spacing in combination with compound 18. Steric fields: green contours (80% contribution) indicate regions where bulky groups increase property, while yellow contours (20% contribution) indicate regions where bulky groups decrease property. (b) Std* coeff. contour maps of CoMFA analysis with 2 Å grid spacing in combination with compound 18. Electrostatic fields: blue contours (80% contribution) indicate regions where groups with positive charges increase property, while red contours (20% contribution) indicate regions where groups with negative charges increase property.

be used to further explain why compounds 14, 15, 16, 17, and 18 have highest property than those of all other compounds.

The CoMFA electrostatic contour plot is displayed in **Figure 6b**. A blue contour indicates that substituents should be electron deficient, and red color indicates that substituents should be electron rich. The blue contour near the 2, 3, 4, and 5 positions (**Figure 6b**) indicates that electron-donating substituents (such Alkyl group) are beneficial for propriety in this area. The electrostatic contour map displays a region of red contours neighbor to the 1 and 6 positions indicating that groups with negative charges may increase the property.

All these findings may be used to design improved compounds with higher or lower retention property, as observed in the CoMFA maps, by adding suitable substituents.

4. Conclusion

In this study, 2D- and 3D-QSRR analyses were used to predict the linear retention indices of a set of alkylated phenols. The multidimensional-QSRR models gave good statistical results in terms of r_{CV} and r values. The stepwise MLR and CoMFA models showed high internal and external consistency; this is verified using different validation methods to evaluate their statistical quality. External validation using a test series verified the capacity of these models to estimate with appropriate precision the linear retention indices of alkylated phenols. In addition, the stepwise MLR equation and CoMFA contour plots can identify that physicochemical properties, organic functional groups, and chemical molecular fragments strongly correlated with the linear retention indices of this studied compounds. The highlighted features are important information for delineating the chemical space, which can be used to design new volatile alkylated phenols. This study consists of the first step explored to code a particular odor of this group of molecules, followed by docking molecular study that allows understand the mechanism of activation of olfactory receptor present in the nasal cavity by this kind of chemical compounds.

Acknowledgements

We are grateful to the "Association Marocaine des Chimistes Théoriciens" (AMCT) for its pertinent help concerning the programs.

Conflict of interest

"The authors declare that they have no competing interests."

Supplementary material

Table S1. Molecular descriptors computed by ChemOffice and ChemSketch software.

No	MW	H°	G	V	T	CT	TB	CP	KH	TVC	PC	MTI	NRB	I	SVD	TC	H%	O%	C%	n	γ	D	log P	W	NHA	NHD	J	PSA
1[a]	94,113	−96.48	−23.94	83.387	282.50	677.45	439.20	59.26	4.64	0.014	1.475	360	0	0	24	0.102	0.064	0.170	0.766	1.553	40.9	1.071	1.64	42	1	1	2050	20.23
2	108,140	−128.59	−34.15	100.414	306.29	692.39	467.06	50.30	4.60	0.012	1.924	506	0	0	26	0.083	0.075	0.148	0.777	1.545	38.8	1.038	2.13	60	1	1	3818	20.23
3	108,140	−128.59	−34.15	100.612	306.29	692.39	467.06	50.30	4.60	0.012	1.974	512	0	0	26	0.083	0.075	0.148	0.777	1.545	38.8	1.038	2.13	61	1	1	3882	20.23
4	108,140	−128.59	−34.15	100.611	306.29	692.39	467.06	50.30	4.60	0.012	1.974	518	0	0	26	0.083	0.075	0.148	0.777	1.545	38.8	1.038	2.13	62	1	1	3942	20.23
5	122,167	−149.23	−25.73	114.476	317.56	707.14	489.94	44.09	4.47	0.009	2.453	714	1	0	28	0.059	0.083	0.131	0.787	1.536	37.6	1.012	2.55	86	1	1	6909	20.23
6	122,167	−149.23	−25.73	115.630	317.56	707.14	489.94	44.09	4.47	0.009	2.503	726	1	0	28	0.059	0.083	0.131	0.787	1.536	37.6	1.012	2.55	88	1	1	7069	20.23
7	122,167	−149.23	−25.73	115.634	317.56	707.14	489.94	44.09	4.47	0.009	2.503	738	1	1	28	0.059	0.083	0.131	0.787	1.536	37.6	1.012	2.55	90	1	1	7222	20.23
8	136,194	−169.87	−17.31	134.066	328.83	720.74	512.82	38.97	4.35	0.006	2.982	992	2	1	30	0.042	0.089	0.117	0.794	1.529	37.1	0.992	2.97	121	1	1	11,979	20.23
9[a]	136,194	−169.87	−17.31	134.747	328.83	720.74	512.82	38.97	4.35	0.006	3.032	1010	2	1	30	0.042	0.089	0.117	0.794	1.529	37.1	0.992	2.97	124	1	1	12,277	20.23
10	136,194	−169.87	−17.31	134.751	328.83	720.74	512.82	38.97	4.35	0.006	3.032	1028	2	0	30	0.042	0.089	0.117	0.794	1.529	37.1	0.992	2.97	127	1	1	12,562	20.23
11[a]	136,194	−175.15	−19.75	128.458	313.83	709.69	512.38	39.36	4.35	0.007	2.702	938	1	0	30	0.048	0.089	0.117	0.794	1.525	35.7	0.987	2.88	114	1	1	11,308	20.23
12	136,194	−175.15	−19.75	130.527	313.83	709.69	512.38	39.36	4.35	0.007	2.902	956	1	0	30	0.048	0.089	0.117	0.794	1.525	35.7	0.987	2.88	117	1	1	11,607	20.23
13	136,194	−175.15	−19.75	130.529	313.83	709.69	512.38	39.36	4.35	0.007	2.902	974	1	1	30	0.048	0.089	0.117	0.794	1.525	35.7	0.987	2.88	120	1	1	11,891	20.23
14	150,221	−190.51	−8.89	151.962	340.10	733.30	535.70	34.68	4.23	0.004	3.561	1396	3	1	32	0.029	0.094	0.107	0.800	1.522	36.7	0.977	3.38	174	1	1	20,809	20.23
15	164,248	−211.15	−0.47	169.037	351.37	744.89	558.58	31.07	4.10	0.003	4.090	1850	4	0	34	0.021	0.098	0.097	0.804	1.518	36.4	0.965	3.8	232	1	1	33,013	20.23
16	178,275	−231.79	7.95	186.175	362.64	755.62	581.46	27.99	3.98	0.002	4.619	2398	5	1	36	0.015	0.102	0.090	0.808	1.514	36.2	0.954	4.22	302	1	1	50,439	20.23
17	192,302	−252.43	16.37	203.315	373.91	765.54	604.34	25.35	3.86	0.002	5.148	3048	6	0	38	0.010	0.105	0.083	0.812	1.510	36.0	0.946	4.64	385	1	1	74,587	20.23
18	206,329	−273.07	24.79	220.457	385.18	774.71	627.22	23.07	3.74	0.001	5.677	3808	7	1	40	0.007	0.107	0.078	0.815	1.507	35.8	0.939	5.05	482	1	1	107,219	20.23
19	122,167	−160.70	−35.36	117.217	330.08	706.16	494.92	43.23	4.55	0.011	2.373	682	0	0	28	0.068	0.083	0.131	0.787	1.540	37.2	1.014	2.62	82	1	1	6599	20.23
20	122,167	−160.70	−35.36	117.634	330.08	706.16	494.92	43.23	4.55	0.011	2.423	696	0	0	28	0.068	0.083	0.131	0.787	1.540	37.2	1.014	2.62	84	1	1	6756	20.23
21	122,167	−160.70	−35.36	117.636	330.08	706.16	494.92	43.23	4.55	0.011	2.423	698	0	0	28	0.068	0.083	0.131	0.787	1.540	37.2	1.014	2.62	84	1	1	6756	20.23
22	122,167	−160.70	−35.36	117.077	330.08	706.16	494.92	43.23	4.55	0.011	2.373	684	0	0	28	0.068	0.083	0.131	0.787	1.540	37.2	1.014	2.62	82	1	1	6599	20.23
23	122,167	−160.70	−35.36	117.406	330.08	706.16	494.92	43.23	4.55	0.011	2.423	694	0	0	28	0.068	0.083	0.131	0.787	1.540	37.2	1.014	2.62	84	1	1	6756	20.23
24	122,167	−160.70	−35.36	117.845	330.08	706.16	494.92	43.23	4.55	0.011	2.473	696	0	0	28	0.068	0.083	0.131	0.787	1.540	37.2	1.014	2.62	84	1	1	6759	20.23
25	136,194	−192.81	−36.57	134.437	353.87	718.88	522.78	37.55	4.51	0.009	2.872	906	0	0	30	0.056	0.089	0.117	0.794	1.535	36.1	0.996	3.11	110	1	1	10,919	20.23

No	MW	H°	G	V	T	CT	TB	CP	KH	TVC	PC	MTI	NRB	I	SVD	TC	H%	O%	C%	n	γ	D	log P	W	NHA	NHD	J	PSA
26	136,194	−192.81	−36.57	133.873	353.87	718.88	522.78	37.55	4.51	0.009	2.822	900	0	0	30	0.056	0.089	0.117	0.794	1.535	36.1	0.996	3.11	109	1	1	10,821	20.23
27	136,194	−192.81	−36.57	134.435	353.87	718.88	522.78	37.55	4.51	0.009	2.872	912	0	0	30	0.056	0.089	0.117	0.794	1.535	36.1	0.996	3.11	111	1	1	11,015	20.23
28[a]	136,194	−192.81	−36.57	134.301	353.87	718.88	522.78	37.55	4.51	0.009	2.872	906	0	0	30	0.056	0.089	0.117	0.794	1.535	36.1	0.996	3.11	110	1	1	10,919	20.23
29[a]	136,194	−192.81	−36.57	134.214	353.87	718.88	522.78	37.55	4.51	0.009	2.872	902	0	0	30	0.056	0.089	0.117	0.794	1.535	36.1	0.996	3.11	110	1	1	10,919	20.23

[a]*Test set.*

Author details

Assia Belhassan[1,2], Samir Chtita[3], Tahar Lakhlifi[1] and Mohammed Bouachrine[1,2]*

1 MCNS Laboratory, Faculty of Science, Moulay Ismail University, Meknes, Morocco

2 Materials, Environment and Modeling Laboratory, ESTM, Moulay Ismail University, Meknes, Morocco

3 Laboratory Physical Chemistry of Materials, Faculty of sciences Ben M'Sik, Hassan II University of Casablanca, Casablanca, Morocco

*Address all correspondence to: m.bouachrine@est-umi.ac.ma

IntechOpen

References

[1] Baltussen E, David F, Sandra P, Janssen H, Cramers C. Automated sorptive extraction-thermal desorption-gas chromatography-mass spectrometry analysis: Determination of phenols in water samples. Journal of Microcolumn Separations. 1999;**11**:471-474. DOI: 10.1002/(SICI)1520-667X(1999)11:6<471::AID-MCS9>3.0.CO;2-9

[2] Mahugo Santana C, Sosa Ferrera Z, Esther Torres Padrón M, Juan Santana Rodríguez J. Methodologies for the extraction of phenolic compounds from environmental samples: New approaches. Molecules. 2009;**14**:298-320. DOI: 10.3390/molecules14010298

[3] Czaplicka M. Determination of phenols and chlorophenols in bottom sediments. Chromatographia. 2001;**53**:S470-S473. DOI: 10.1007/BF02490380

[4] Fleuriet A, Uhel C, Dédaldéchamp F. Les composés phénoliques et la qualité des produits d'origine végétale consommés par l'homme. Acta Botanica Gallica. 1996;**143**:493-500. DOI: 10.1080/12538078.1996.10515346

[5] Hartman TG, Karmas K, Chen J, Shevade A, Deagro M, Hwang H-I. Determination of vanillin, other phenolic compounds, and flavors in vanilla beans: Direct thermal desorption—gas chromatography and—gas chromatography—mass spectrometry analysis. ACS Publications. 1992;**506**:60-76. DOI: 10.1021/bk-1992-0506.ch004

[6] Naim M, Zehavi U, Nagy S, Rouseff RL. Hydroxycinnamic acids as off-flavor precursors in citrus fruits and their products. ACS Publications. 1992;**506**:180-191. DOI: 10.1021/bk-1992-0506.ch014

[7] Maga JA. Contribution of phenolic compounds to smoke flavor. ACS Publications. 1992;**506**:170-179. DOI: 10.1021/bk-1992-0506.ch013

[8] Wang T, Yuan X, Wu M-B, Lin J-P, Yang L-R. The advancement of multidimensional QSAR for novel drug discovery-where are we headed? Expert Opinion on Drug Discovery. 2017;**12**:769-784. DOI: 10.1080/17460441.2017. doi: 10.1080/17460441.2017.1336157

[9] De Souza AS, Ferreira LG, Andricopulo AD. 2D and 3D QSAR studies on a series of antichagasic fenarimol derivatives. International Journal of Quantitative Structure-Property Relationships. 2017;**2**:44-63. DOI: 10.4018/IJQSPR.2017010104

[10] Belhassan A, Chtita S, Lakhlifi T, Bouachrine M. QSRR study of linear retention indices for volatile compounds using statistical methods. Chemical Science Transactions. 2018;**7**:558-575. DOI: 10.7598/cst2018.1501

[11] Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. Journal of the American Chemical Society. 1988;**110**:5959-5967. DOI: 10.1021/ja00226a005

[12] Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. Journal of Medicinal Chemistry. 1994;**37**:4130-4146. DOI: 10.1021/jm00050a010

[13] Vaidya A, Jain S, Jain S, Jain AK, Agrawal RK. Quantitative structure-activity relationships: A novel approach of drug design and discovery. Journal of Pharmaceutical Sciences and Pharmacology. 2014;**1**:219-232. DOI: 10.1166/jpsp.2014.1024

[14] Czerny M, Brueckner R, Kirchhoff E, Schmitt R, Buettner A. The influence of molecular structure on odor qualities and odor detection thresholds

of volatile alkylated phenols. Chemical Senses. 2011;**36**:539-553. DOI: 10.1093/chemse/bjr009

[15] ACD/LABS. 10 Advanced Chemistry Development, Inc. Toronto, ON, Canada; 2015. Available from: www.acdlabs.com

[16] ChemBioOffice. PerkinElmer Informatics. 2010. Available from: http://www.cambridgesoft.com

[17] Belhassan A, Chtita S, Lakhlifi T, Bouachrine M. QSPR study of the retention/release property of odorant molecules in water, dairy and pectin gels. Materials Today: Proceedings. 2019;**13**:621-629. DOI: 10.1016/j.matpr.2019.04.021

[18] SPSS 19.0. Available from: http://www.ibm.com/analytics/fr/fr/technology/spss/

[19] Chtita S, Ghamali M, Ousaa A, Aouidate A, Belhassan A, Taourati AI, et al. QSAR study of anti-human African trypanosomiasis activity for 2-phenylimidazopyridines derivatives using DFT and Lipinski's descriptors. Heliyon. 2019;**5**:e01304. DOI: 10.1016/j.heliyon.2019.e01304

[20] Gramatica P. Principles of QSAR models validation: Internal and external. QSAR and Combinatorial Science. 2007;**26**:694-701. DOI: 10.1002/QSPR.200610151

[21] Zaki H, Belhassan A, Aouidate A, Lakhlifi T, Benlyas M, Bouachrine M. Antibacterial study of 3-(2-amino-6-phenylpyrimidin-4-yl)-N-cyclopropyl-1-methyl-1H-indole-2-carboxamide derivatives: CoMFA, CoMSIA analyses, molecular docking and ADMET properties prediction. Journal of Molecular Structure. 2019;**1177**:275-285. DOI: 10.1016/j.molstruc.2018.09.073

[22] Clark M, Cramer IIIRD, Van Opdenbosch N. Validation of the general purpose tripos 5.2 force field. Journal of Computational Chemistry. 1989;**10**:982-1012

[23] Purcell WP, Singer JA. A brief review and table of semiempirical parameters used in the Hueckel molecular orbital method. Journal of Chemical & Engineering Data. 1967;**12**:235-246

[24] AbdulHameed MDM, Hamza A, Liu J, Zhan C-G. Combined 3D-QSAR modeling and molecular docking study on indolinone derivatives as inhibitors of 3-phosphoinositide-dependent protein kinase-1. Journal of Chemical Information and Modeling. 2008;**48**:1760-1772. DOI: 10.1021/ci800147v

[25] Ståhle L, Wold S. 6 multivariate data analysis and experimental design in biomedical research. Progress in Medicinal Chemistry. 1988;**25**:291-338

[26] Bush BL, Nachbar RB. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. Journal of Computer-Aided Molecular Design. 1993;**7**:587-619

[27] Belhassan A. 3D-QSAR study of biodegradability in water for aromatic compounds. RHAZES: Green and Applied Chemistry. 2018;**1**:21-30

[28] Srivastava V, Kumar A, Mishra BN, Siddiqi MI. CoMFA and CoMSIA 3D-QSAR analysis of DMDP derivatives as anti-cancer agents. Bioinformation. 2008;**2**:384

[29] Belhassan A, Chtita S, Lakhlifi T, Bouachrine M. QSPR study of the retention/release property of odorant molecules in water using statistical methods. Orbital: The Electronic Journal of Chemistry. 2017;**9**:234-247. DOI: 10.17807/orbital.v9i4.978

[30] Chtita S, Ghamali M, Hmamouchi R, Elidrissi B, Bourass M,

Larif M, et al. Investigation of antileishmanial activities of acridines derivatives against promastigotes and amastigotes form of parasites using quantitative structure activity relationship analysis. Advances in Physical Chemistry. 2016;**2016**:16. DOI: 10.1155/2016/5137289

[31] Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. Environmental Health Perspectives. 2003;**111**:1361-1375

[32] Garg R, Smith CJ. Predicting the bioconcentration factor of highly hydrophobic organic chemicals. Food and Chemical Toxicology. 2014;**69**: 252-259. DOI: 10.1016/j.fct.2014.03.035