Journal of the Argentine Chemical Society

QUANTITATIVE STRUCTURE–PROPERTY RELATIONSHIPS FOR PREDICTING THE RETENTION INDICES OF FRAGRANCES ON STATIONARY PHASES OF DIFFERENT POLARITY

Cristian Rojas[a,b,*], Pablo R. Duchowicz[a,*], Piercosimo Tripaldi[b], and Reinaldo Pis Diez[c]

[a] *Inst. de Invest. Fisicoquímicas Teóricas y Aplicadas (INIFTA), CONICET, UNLP, Diag. 113 y 64, C.C. 16, Suc. 4, 1900 La Plata, Argentina*
[b] *Vicerrectorado de Investigaciones, U. del Azuay, Av. 24 de Mayo 7–77 y Hernán Malo, Cuenca, Ecuador*
[c] *CEQUINOR, Centro de Química Inorgánica (CONICET, UNLP), Dto. de Química, F. de Cs. Exactas, UNLP, C.C. 962, 1900 La Plata, Argentina*
[*] *Corresponding authors: crojasvilla@gmail.com ; pabloducho@gmail.com*

Received Feb. 28th, 2017. Accepted Jul. 6th, 2017.

Resumen

El objetivo de este trabajo fue el desarrollo de relaciones cuantitativas estructura–propiedad predictivas para el modelado de índices de retención (*I*) de fragancias, medidas en tres fases estacionarias de diferente polaridad: DB–225MS, HP5–MS y HP–1. Se ha prestado particular atención al curado de los datos experimentales. Posteriormente, se usó el método de subconjuntos balanceados (BSM) para dividir cada base de datos en grupos de calibración, validación y predicción. Los modelos se construyeron a partir de 1819 descriptores moleculares independientes de la conformación, los cuales fueron analizados mediante el método de reemplazo (RM) para la selección de los mismos, con la finalidad de obtener los mejores modelos. Para la fase estacionaria DB–225MS se obtuvo un modelo basado en cuatro descriptores, mientras que para las columnas HP5–MS y HP–1 se propusieron modelos con tres descriptores. Los modelos fueron validados mediante validación cruzada de dejar–uno–fuera y dejar–varios–fuera, así como otros criterios de validación. Adicionalmente, con la finalidad de cumplir los principios propuestos por la Organization for Economic Co–operation and Development (OECD), la capacidad predictiva de los modelos se evaluó mediante la predicción de los índices de retención del grupo externo de predicción, el dominio de aplicabilidad fue apropiadamente definido y se realizó una interpretación de cada descriptor molecular involucrado.

Abstract

The purpose of this work was to develop predictive quantitative structure–property relationships for modeling the retention indices (*I*) of fragrances measured in three stationary phases of different polarities: DB–225MS, HP5–MS and HP–1. Attention was paid to the curation of the experimental data. Subsequently, the Balanced Subsets method (BSM) was used to split each dataset into training, validation and test sets. Models were established by using 1819 conformation–independent molecular descriptors which were analyzed by the replacement method (RM) variable subset selection in order to obtain the optimal models. A four–descriptor model was obtained for the DB–225MS stationary phase while a three–parametric model was proposed for both the HP5–MS and HP–1 columns. Models were validated by means of the leave–one–out and leave–many–out cross–validation procedures, as well as other validation criteria. Moreover, in order to accomplish the principles proposed by the Organization for Economic Co–operation and Development (OECD), the model's predictive ability was measured by predicting retention indices of the external test set. The applicability domain was properly defined and the interpretation of each of the molecular descriptors used in this study was provided.

***Palabras clave**:* aromas, índice de retención, descriptores Dragon, método de subconjuntos balanceados, método de reemplazo

***Keywords**:* fragrance; retention index, Dragon descriptors; balanced subsets method; replacement method

1. Introduction

Fragrances are usually defined as organic molecules that impart a pleasant odor as their main characteristic. Odors that are unpleasant are usually given different defining words such as noxious, stench, reek, stink, foul or a smelly odor. Fragrances contain chemical messengers or structures that are detected by receptors on olfactory cells located in the nose. For this reason, fragrances have an important role in several industries, such as food, wines and spirits, perfume, hygiene, medicine, and tobacco [1] to name a few. Since fragrances are volatile organic compounds (VOCs), their aromatic profile is easily assessed by gas chromatography–mass spectrometry (GC–MS), due to its high separation performance and accurate identification capability [2]. The chromatographic retention index (*I*) is a useful parameter for compound identification and for the investigation of the mechanisms of chromatographic retention behavior by means of quantitative structure–property relationships (QSPRs). Since the pioneering studies of the applications of QSPR theory to retention indices carried out in 1977 by Kaliskan and Foks [3, 4] and Michotte and Massart [5], there is a high interest of this approach that avoids the use

of time–consuming and expensive procedures, making the accurate prediction of retention indices of new molecules possible [6, 7].

A large number of QSPR studies have been published in the past years for modeling and predicting the *I* parameter on the weakly polar HP5–MS stationary phase. In 2003, Eckel and Kind [8] used the normal boiling point of 56 compounds to perform a simple linear regression (SLR) against the Lee retention index on the DB–5 GC column (equivalent to the HP5–MS), achieving a good coefficient of determination ($R^2 = 0.97$). In the same year, Wang et al. [9] developed a QSPR model for 18 chlorinated polycyclic aromatic hydrocarbons (Cl–PAHs) using eight descriptors. They obtained a MLR model exhibiting a coefficient of determination of $R^2 = 0.99$.

Later, in 2009, Qin et al. [10] performed three QSPR models for the retention times of 96 essential oils. The 72 training molecules were used for the variable selection and modeling based on prediction (VSMP). They developed a six parametric multiple linear regression (MLR) model ($R^2_{train} = 0.98$, $RMSD_{train} = 1.64$, $R^2_{loo} = 0.97$, $RMSD_{loo} = 1.84$, $R^2_{test} = 0.94$ and $RMSD_{test} = 2.65$), a principal component regression (PCR) model with seven molecular descriptors and four components ($R^2_{train} = 0.97$, $RMSD_{train} = 1.92$, $R^2_{loo} = 0.96$, $RMSD_{loo} = 2.06$, $R^2_{test} = 0.95$ and $RMSD_{test} = 2.44$), as well as a partial least squares (PLS) model using seven descriptors in 5 latent variables (LVs) ($R^2_{train} = 0.97$, $RMSD_{train} = 1.81$, $R^2_{loo} = 0.96$, $RMSD_{loo} = 2.01$, $R^2_{test} = 0.96$ and $RMSD_{test} = 2.19$). The same year, Riahi et al. [11] worked on a dataset of retention indices of 100 essential oil compounds from the *Citrus sudachi*. Molecular descriptors for the 80 training molecules were analyzed by means of the Genetic Algorithms (GAs) in order to calibrate a MLR model ($R^2_{train} = 0.95$, $RMSD_{train} = 48.3$, $R^2_{loo} = 0.94$, $R^2_{test} = 0.93$ and $RMSD_{test} = 60.4$) and a PLS model with six descriptors in three latent variables (LVs) ($R^2_{train} = 0.94$, $RMSD_{train} = 50.3$, $R^2_{loo} = 0.92$, $R^2_{test} = 0.91$ and $RMSD_{test} = 67.0$). In addition, they developed two nonlinear models: the polynomial PLS (poly–PLS) model with six descriptors in three latent variables (LVs) ($R^2_{train} = 0.95$, $RMSD_{train} = 46.0$, $R^2_{loo} = 0.94$, $R^2_{test} = 0.93$ and $RMSD_{test} = 56.9$), and the support vector machine (SVM) model ($R^2_{train} = 0.99$, $RMSD_{train} = 24.7$, $R^2_{loo} = 0.96$, $R^2_{test} = 0.96$ and $RMSD_{test} = 51.4$).

In 2011, Mohammadhosseini et al. [12] performed a reliable three–parametric MLR model for a training set of 29 compounds ($R^2_{train} = 0.98, R^2_{loo} = 0.95$ and $R^2_{lmo} = 0.94$), which was used for predicting retention indices of 14 test molecules ($R^2_{test} = 0.95$). One year later, Noorizadeh and Noorizadeh [13] studied the two–dimensional gas chromatography retention time for 69 opiate and sedative drugs. GAs were used to select the best Dragon molecular descriptors in a training set of 56 molecules, and to perform linear relationships: a MLR model with six descriptors ($R^2_{train} = 0.91$, $RMSD_{train} = 95.61$, $R^2_{test} = 0.87$ and $RMSD_{test} = 163.6$) and a PLS model with eight descriptors in five latent variables ($R^2_{train} = 0.92$, $RMSD_{train} = 90.5$, $R^2_{test} = 0.88$ and $RMSD_{test} = 157.9$); as well as nonlinear relationships: kernel PLS (KPLS) model with seven descriptors in five latent variables ($R^2_{train} = 0.93$, $RMSD_{train} = 81.9$, $R^2_{test} = 0.90$ and $RMSD_{test} = 141.3$) and Levenberg–Marquardt artificial neural network (L–M ANN) model using the KPLS descriptors ($R^2_{train} = 0.96$, $RMSD_{train} = 68.4$, $R^2_{test} = 0.93$ and $RMSD_{test} = 117.7$). The same year, 2012, Zhao et al. [14] used a dataset of 178 volatile organic compounds to predict the gas chromatographic programmed–temperature retention indices by means of topological indices divided into several blocks. The model was built on the basis of the subspace orthogonal projection (MSOP) with the *Monte–Carlo* cross–validation (MCCV). The best model with 9 blocks of descriptors ($R^2_{train} = 0.99$, $RMSD_{train} = 36.0$) clearly indicated good prediction, which was verified by an external test set of 20 molecules ($RMSD_{test} = 32.6$).

Subsequently, Qin et al. [15] developed a QSPR model for the retention indices of essential oil constituents. They used the Kennard–Stone procedure for splitting the dataset into a training set and a test set of 83 molecules each. Three Dragon descriptors were selected by means of the ordered predictors selection (OPS) procedure, exhibiting good stability in fitting ($R^2_{train} = 0.94$, $R^2_{loo} = 0.93$ and $R^2_{lmo} = 0.93$) and prediction ($R^2_{test} = 0.90$ and $RMSD_{test} = 94$). More recently, Goudarzi et al. [16] used the random forests (RF) approach to predict the retention indices of 83 polycyclic aromatic hydrocarbon (PAH) compounds. RF, which is intrinsically a variable selection, was used for selecting Dragon descriptors. The best RF model with five descriptors ($R^2 = 0.99$) is given with 70 randomly selected variables to split each node and 460 trees. RF descriptors were also used to perform a back–propagation artificial neural network (BPANN) ($R^2 = 0.99$). In addition, the stepwise variable subset selection was used to find both a

six–parametric MLR model ($R^2 = 0.98$) and a BPANN model ($R^2 = 0.98$). Finally, in a recent study, Mohammadhosseini [17] calibrated a MLR model for the retention index of 80 terpenoid derivatives based on the particle swarm optimization (PSO) descriptor selection ($R^2_{train} = 0.94$, $R^2_{loo} = 0.93$, $R^2_{lmo} = 0.92$), which was used to predict the *I* for 28 test compounds ($R^2_{test} = 0.90$).

Similarly, retention indices measured in the non–polar HP–1 stationary phase were subjected to several QSPR studies. In 2000, Héberger et al. [18] used a dataset of 35 aliphatic ketones and aldehydes and their retention indices were measured in the HP–1, HP–50, DB–210, and HP–INNOWax columns at four different temperatures (50, 70, 90 and 110°C) [19]. The model, constructed using the PLS2 [20] approach, included four latent variables and 16 retention indices, and exhibited a good fitting quality ($R^2 = 0.99$) and prediction ($R^2 = 0.97$). Subsequently, Körtvélyesi et al. [21] used the Hébergers' dataset [18, 19] to perform two QSPR models for the retention index. The first model was based on two quantum–chemical descriptors ($R^2 = 0.99$ and $RMSD = 12.04$), while the second model was based on three physicochemical descriptors ($R^2 = 1.00$ and $RMSD = 10.93$).

In 2003, da Silva Junkes et al. [22] used a new semi–empirical topological index (I_{ET}) to perform SLR for the retention index of 632 VOCs measured on stationary phases of low–polarity (e.g. Squalane, DB–1, HP–1 and OV–1). In a first attempt, they developed a QSPR model using the entire dataset ($R^2 = 0.99$ and $RMSD = 17.71$), and after the exclusion of 84 outliers, they recalibrated the model ($R^2 = 1.00$, $RMSD = 7.01$ and $R^2_{loo} = 0.99$). In the same year, Ren [23] proposed a QSPR model for the *I* parameter of 33 saturated chemicals (14 aldehydes and 19 ketones) in the HP–1 stationary phase at 50°C. A modified index and three atom–types were used to calibrate a MLR model ($R^2 = 0.99$, $RMSD = 7.73$, $R^2_{loo} = 0.99$ and $RMSD_{loo} = 9.43$). Subsequently, in 2004, Junkes et al. [24] used 31 oxo compounds from the Hébergers' dataset [18, 19] to build a SLR model between the retention index and the I_{ET}

($R^2 = 0.99$, $RMSD = 5.47$ and $R^2_{loo} = 0.99$).

In 2007, Zhou and Nie [25] used the Hébergers' dataset [18, 19] to perform a QSPR model based on a novel topological index and a path number ($R^2 = 1.00$ and $RMSD = 10.8$). One year later, Konoz et al. [26] used 140 molecules from the Hébergers' dataset [18, 19] to calibrate a MLR model ($R^2_{train} = 0.94$, $RMSD_{train} = 47.4$, $R^2_{val} = 0.94$, $RMSD_{val} = 46.8$, $R^2_{test} = 0.94$, $RMSD_{test} = 50.5$), and an artificial neural network (ANN) model ($R^2_{train} = 1.00$, $RMSD_{train} = 8.0$,

$R^2_{val} = 1.00$, $RMSD_{val} = 9.9$, $R^2_{test} = 1.00$, $RMSD_{test} = 10.5$). Finally, Souza et al. [27] used 15 aldehydes and 42 ketones from the Hébergers' dataset [18, 19] to develop two models based on the semi–empirical electrotopological index (I_{SET}). The model for aldehydes ($R^2 = 0.99$, $RMSD = 10.31$ and $R^2_{loo} = 0.99$) and the model for ketones ($R^2 = 0.99$, $RMSD = 11.72$ and $R^2_{loo} = 0.99$) exhibited good performances.

Given these premises, the purpose of the work presented here was the development of conformation–independent quantitative structure–property relationships for retention indices of 312 aromatic compounds measured in GC-MS, using three stationary phases of different polarities, such as the polar DB–225MS, weakly polar HP5–MS and the non–polar HP–1. To the best of our knowledge, these datasets have not been used to perform QSPR studies, nor have QSPR models been reported for the DB–225MS stationary phase. To make the models applicable, the five principles defined by the Organization for Economic Co–operation and Development (OECD) [28] were followed. In brief, the modeled property and algorithm should be clearly defined, the applicability domain of the QSPR model should be defined, the goodness–of–fit and predictive ability of the model should be evaluated through appropriate strategies, and a mechanistic interpretation of model descriptors should be given (if possible). Thus, we paid attention to the curation of the experimental data, which led to the definition of a dataset of 269 VOCs. Subsequently, the replacement method (RM) variable subset selection was used to search for the optimal descriptor models, and the applicability domains (AD) of such models were properly defined. The models' predictive ability was measured by internal and external validation procedures. Finally, an explanation of the mechanistic effect of molecular descriptors in each QSPR model is presented.

2. Materials and Methods

2.1. Dataset description, molecular structure representation and data curing

Experimental GC–MS retention indices for 312 aromatic compounds were retrieved from the database developed by Yan et al. [29]. This property was measured on an Agilent 7890A Gas Chromatograph coupled with an Agilent 5975C Mass Spectrometer using three stationary phases: the polar DB–225MS, weakly polar HP5–MS and the non–polar HP–1.

Chemical names, CAS numbers and retention indices of the 312 VOCs were merged using the KNIME platform [30] in order to cure the dataset. Subsequently, the Chemical Identifier

Resolver [31] node was used to obtain the canonical SMILES (simplified molecular input line entry system) strings from both the chemical name and CAS number. SMILES strings of all the compounds were then verified for the correct match between CAS number and chemical name.
For those compounds exhibiting two different SMILES notations, the correct molecular structure was manually checked within available libraries, such as PubChem [32], NIST Chemistry WebBook [33] and ChemSpider [34]. During data curation, 41 VOCs were identified as duplicates. Moreover, due to the fact that the purpose of this work was the development of conformation–independent QSPR models, molecules exhibiting the same canonical SMILES string were merged as a single compound. These fragrances were: 1) the *Bicyclo[2.2.1]heptan–2–ol, 1,7,7–trimethyl–, acetate, (1S–endo)–* (CAS number 005655–61–8) and the Isobornyl acetate (CAS number 000125–12–2), and 2) the *Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl ester* (CAS number 000077–83–8) and the *Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl ester, cis–* (CAS number 019464–95–0). For both the duplicates and the merged conformation–independent compounds, the average retention index was used. Thus, the final dataset for the three stationary phases was comprised of 269 VOCs. Details of the cured datasets were given in Table 1S.

2.2. Molecular Descriptors

Since the modeling of the retention indices property was affected by only a conformation–independent representation of the fragrances [35, 36], only these kinds of descriptors were calculated in order to develop the QSPR models. Descriptors were the final result of a logical and mathematical procedure that transformed chemical information encoded within a symbolic representation of a molecule into a numerical quantity or into the result of some standardized experiment [37]. Thus, 3808 conformation–independent molecular descriptors were computed using Dragon version 7 [38] node implemented in KNIME.

2.3. Model Development

2.3.1. Molecular Descriptor Reduction and Selection

In a first attempt to develop the QSPR model, non–informative molecular descriptors were excluded; that is, molecular descriptors affected by missing values or descriptors with constant and near–constant values. Thus, 1819 descriptors were subjected to subsequent supervised selection of the optimal descriptors by means of the replacement method (RM) variable subset

selection [39, 40]. This crucial step facilitated the interpretation and prediction of QSPR in multiple linear regression (MLR) models [41]. This approach was a sequential method which generated subsets of descriptors of a given dimension (*d*) from the pool of 1819 descriptors (*D*), by minimizing the root–mean–square deviation (*RMSD*) as a fitness function.

2.3.2. Model Validation

In order to avoid the risk of overfitting during the RM selection, the QSPR models were validated by means of a validation set; an external test set was also used to evaluate the 'real' predictive ability of the QSPR models. To this end, each dataset was divided into training, validation, and test sets based on the Balanced Subsets Method (BSM) approach [35]. This method has been used elsewhere to study several properties [35, 36, 42]. The splitting of a dataset should be done in order to achieve a similar structure–property relationships design in the training, validation and test sets [43]. Consequently, the BSM considered the retention indices and conformation–independent molecular descriptors only to create *k*–clusters of fragrances in terms of the Euclidean distance, in such a way as to group similar compounds in the same cluster. The steps involved in the BSM procedure were:

a. prepare a matrix ($\mathbf{C_1}$) that included the experimental retention indices for all the fragrances in each stationary phase (*N* molecules) and the 1819 conformation–independent molecular descriptors.

b. remove the linearly dependent variables from $\mathbf{C_1}$. The new size of the reduced matrix $\mathbf{C_2}$ was $N \times 268$.

c. standardize the reduced matrix $\mathbf{C_2}$ for centering and scaling its matrix elements.

d. create N_{train}^{0} clusters through the *k*–MCA method, for which the $\mathbf{C_2}$ standardized matrix was used together with the Euclidean metrics, and several runs for optimizing the *k*–MCA algorithm. This step calculated N_{train}^{0} cluster centroid locations with dimensions of 1×268. $N_{train}^{0} = N_{train} - N_{\min\max}$, where N_{train} was the number of fragrances in the training set and $N_{\min\max}$ was the number of molecules having the maximum or minimum retention index.

e. the training set (N_{train}) was constructed by including one compound per cluster (i.e., the nearer molecule to the centroid in each cluster). $N_{\min\max}$ molecules were also included in the training set.

f. generate N_{val} clusters with the remaining $N - N_{train}$ molecules through the *k*–MCA method with same numerical conditions as described above. This calculated N_{val} cluster centroid locations.

g. the validation set (N_{val}) was designed by including one molecule per cluster (i.e., the nearer compound to the centroid in each cluster).

h. Finally, the test set (N_{test}) included the remaining $N - N_{train} - N_{val}$ fragrances.

During the RM variable selection procedure, molecules in the training set were used during the supervised selection of molecular descriptors and to calibrate the models, while the validation set was used for the cross–validation of the model in order to avoid the presence of overfitting. Finally, test molecules were used to measure the predictive ability of the selected QSPR model.

The leave–one–out (loo) and leave–many–out (lmo) cross–validation techniques were further evaluated. In the leave–one–out cross–validation procedure, one molecule was excluded at a time from the model, and the remaining *n–1* molecules were used to re–calibrate the model and then used to predict the retention index of the excluded molecule. On the other hand, in the leave–many–out procedure, 20% of the molecules were randomly removed (as many times as possible), and the remaining molecules were used to re–calibrate the model and to predict the retention indices of the excluded molecules. This resampling procedure was repeated 50000 times.

The absence of chance correlation in the models was also evaluated through the Y–randomization procedure [44], which consisted of randomly scrambling the *I* values so that they did not correspond to the respective molecules. After 10000 iterations, the Y–randomization quality (R^2_{rand} or $RMSD_{rand}$) had to be poorer with respect to model parameters (R^2_{train} or $RMSD_{train}$). Finally, several other validation criteria were also evaluated in order to thoroughly validate the model and to avoid the proposal of an overoptimistic and perhaps erroneous, "predictive" QSPR model [45]. These additional validation criteria are defined in Table 2S.

2.3.3. Applicability Domain assessment

The applicability domain (AD) assessment of the QSPR models was based on the definition of a theoretical space that depended on the selected molecular descriptors and the experimental retention indices of the molecules [46]. The leverage approach [47] was one of the various

strategies proposed in the literature for defining the model's AD. This distance–based method measured the distance of each *i*th test molecule (h_i) from the centroid of the training set defined by a warning threshold (h^*). Thus, each model was confined to a chemical space defined by the chemical information provided by the molecular descriptors of the training molecules. Only test molecules with leverage values lower than this defined threshold ($h_i \leq h^*$) were considered to be reliably predicted for model interpolation; otherwise, the molecule fell outside the AD and its prediction was considered an extrapolation of the model ($h_i > h^*$).

2.3.4. Molecular descriptor interpretation

An important issue to be addressed in QSPR studies was the mechanistic interpretation of each molecular descriptor included in the MLR model. That is, how were these descriptors related to the retention indices of the compounds. Due to the fact that MLR models provided numerical coefficients for each *j*th molecular descriptor, the degree of the contribution of each descriptors was calculated by standardizing its regression coefficients (b_j^s). Thus, the larger the absolute value of b_j^s was for a given descriptor, the greater the importance of such descriptor was in modeling the experimental retention indices [48].

2.4. Software

A user–programmed workflow on KNIME [30] was used for the dataset description, molecular structure representation, and data curation. Molecular descriptors were computed using the node of Dragon version 7 [38] implemented in KNIME. The Balanced Subsets Method, the Replacement Method variable subset selection, and model fitting along with validation were performed in the MatLab environment [49] by means of routines programmed by the authors.

3. Results and Discussion

3.1. DB–225MS Stationary phase

The DB–225MS dataset comprised of 269 compounds was split by the BSM into a training set (90), a validation set (90), and a test set (89). Then, the supervised RM variable subset selection was used to explore the pool of 1819 conformation–independent molecular descriptors. The selection of the best model was performed by minimizing the RMSD. The number of descriptors *(d)* was kept as small as possible according to the principle of parsimony (Ockham's razor) [50].

Consequently, a four–parametric quantitative structure–property relationship was found to be the optimal model:

$$I_{DB-225MS} = 702.6 + 65.9 \times SpAbs\_B(e) + 1465.6 \times MATS1v - 105.9 \times C-001 + 235.7 \times C-008 \quad \text{(Eq. 1)}$$

$N_{train} = 90$, $R^2_{train} = 0.82$, $RMSD_{train} = 205.2$
$N_{val} = 90$, $R^2_{val} = 0.84$, $RMSD_{val} = 173.1$
$N_{test} = 89$, $R^2_{test} = 0.76$, $RMSD_{test} = 179.9$
$R^2_{loo} = 0.82$, $RMSD_{loo} = 191.5$, $R^2_{lmo} = 0.88$, $RMSD_{lmo} = 212.4$
$RMSD_{rand} = 414.8$, $o(3S) = 3$, $R^2_{ij\,max} = 0.16$

Moreover, the additional validation criteria presented below confirmed that a predictive QSPR model was achieved for the *I* prediction of fragrances in the DB–225MS stationary phase:

$R^2_{loo} > 0.5$ (0.82) and the $R^2_{test} > 0.6$ (0.76)
$1 - R^2_0 / R^2_{test} < 0.1$ (0.000) or $1 - R'^2_0 / R^2_{test} < 0.1$ (0.10)
$0.85 \le k(0.99) \le 1.15$ and $0.85 \le k'(0.99) \le 1.15$
$R^2_m > 0.5$ (0.75)

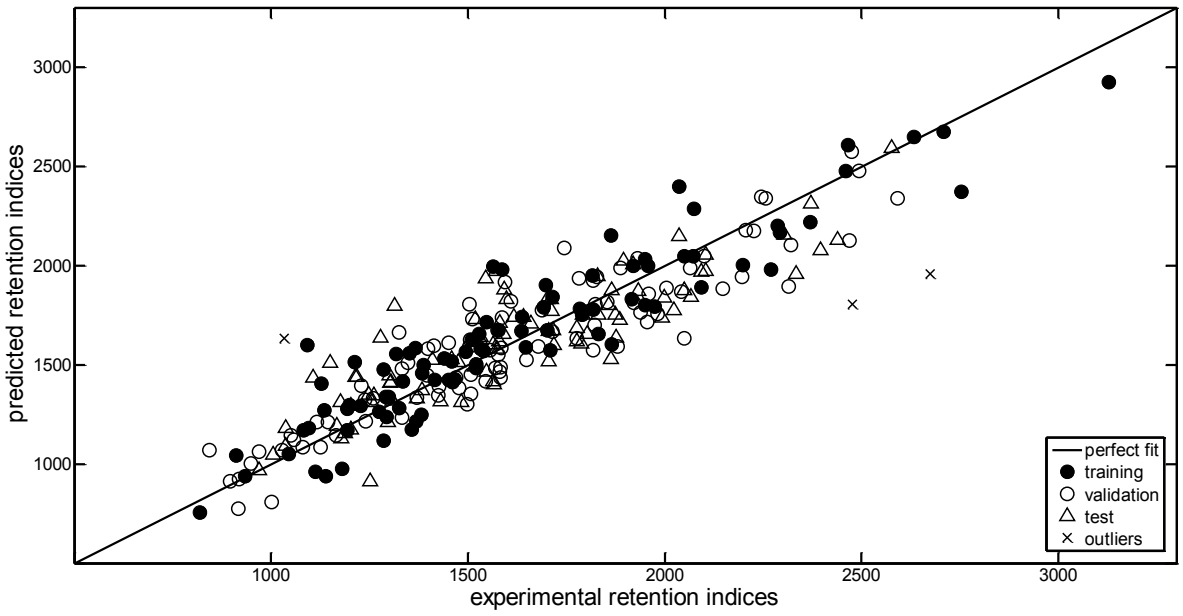

Figure 1. Experimental *vs.* predicted retention indices for fragrances in DB–225MS stationary phase

Figure 1 shows the plot between the experimental retention indices against the predicted *I* calculated with Eq. 1. The fulfillment of all the validation criteria and Figure 1 confirmed that a predictive QSPR was developed. Numerical values of both molecular descriptors and predicted retention indices provided by Eq. 1 are shown in Table 3S.

Three fragrances fell outside the threshold of three standard deviations (*o(3S)*): *2(4H)–Benzofuranone, 5,6,7,7a–tetrahydro–4,4,7a–trimethyl–*, *2–Butanone, 4–(4–hydroxyphenyl)–*, and *Benzyl alcohol*. Since their chemical structure and the *I* value at the source were correct, the irregular behavior of these fragrances may be attributed to the wide chemical diversity of the VOCs considered in the dataset, as well as to analytical aspects during the measurement of the retention index [36, 51, 52].

The QSPR model presented in Eq. 1 was composed by a 2D matrix–based descriptor (*SpAbs_B(e)*), one 2D autocorrelations (*MATS1v*) and two atom–centered fragments (*C–001* and *C–008*). Moreover, the maximum coefficient of determination ($R^2_{ij\,max} = 0.16$), indicated a low correlation between the *SpAbs_B(e)* and *MATS1v* descriptors, indicating that the model did not exhibit multicollinearity. Thus, each descriptor described particular aspects of the retention index mechanism. In addition, the degree of contribution of each molecular descriptor was assessed by standardizing the regression coefficients of the four descriptors involved in modeling the *I* in the DB–225MS Stationary phase: 0.78 (*SpAbs_B(e)*) > 0.28 (*C–001*) > 0.20 (*MATS1v*) > 0.17 (*C–008*).

The graph energy from the Burden matrix weighted by Sanderson electronegativity (*SpAbs_B(e)*) [53, 54] was a topological index calculated as the sum of absolute values of all the eigenvalues (graph invariant) from the Burden Matrix weighted by Sanderson electronegativity. A relationship was established such that the larger the value of the *SpAbs_B(e)* descriptor for a given fragrance, a higher retention index was found. This descriptor indicated that the presence of electronegative atoms (e.g., O, N, S, P, Se or halogens) or groups of atoms potentially involved in hydrogen bonds in the fragrance tended to increase its retention index due to the ability to attract electrons to itself. Consequently, the fragrance was able to create polar interactions with the components of the DB–225MS polar stationary phase. This fact was confirmed by the presence of CH fragments bonded to two carbon atoms and an electronegative atom in the molecule (*C–008*) [55].

MATS1v was a descriptor calculated by applying the Moran coefficient [56] to the H–filled molecular graph weighted by the van der Waals volume (v), and provided information on the

distribution of this property along with the topological structure of the volatile compound. High retention indices were related to positive values of the Moran coefficient (positive spatial autocorrelations), that is, compounds containing atoms with similar van der Waals volume at lag 1. Finally, the negative regression coefficient of the *C–001* [55] descriptor indicated that retention indices were inversely related to the presence of CH_3 fragments (bonded to a carbon atom) in the fragrance. These descriptors measured the presence of methyl–branched carbons and the presence of terminal methyl fragments in the scaffold (i.e., the length of the carbon chain). For example, the *Menthyl isovalerate* and the *α–Irone* both had five methyl radicals. This behavior of the retention indices had been observed in methyl–branched hydrocarbons [57, 58].
In addition, the applicability domain assessment provided information regarding the limitation of four molecular descriptors to generate reliable predictions, i.e., predicted retention indices were restricted only for fragrances exhibiting a leverage value below the warning leverage of the model ($h^* = 0.083$). For the DB–225MS Stationary phase, there were no test molecules with leverage values above the warning leverage of the model; therefore, this QSPR model was able to reliably predict the retention index of all the test molecules.

3.2. HP5–MS Stationary phase

For the HP5–MS dataset, the same workflow as for the DB–225MS dataset was used. Therefore, the BSM was used to split the 266 VOCs into a training set of 88 compounds, along with validation and test sets containing 89 molecules each. The RM led to the following optimal three–parametric QSPR model:

$$I_{HP5-MS} = 286.6 + 74.3 \times SpPos\_B(p) - 56.6 \times C-001 + 108.7 \times C-008 \qquad \text{(Eq. 2)}$$

$N_{train} = 88$, $R^2_{train} = 0.94$, $RMSD_{train} = 76.9$
$N_{val} = 89$, $R^2_{val} = 0.95$, $RMSD_{val} = 72.8$
$N_{test} = 89$, $R^2_{test} = 0.90$, $RMSD_{test} = 76.2$
$R^2_{loo} = 0.94$, $RMSD_{loo} = 75.1$, $R^2_{lmo} = 0.96$, $RMSD_{lmo} = 83.0$
$RMSD_{rand} = 294.0$, $o(3S) = 1$, $R^2_{ij\,max} = 0.03$

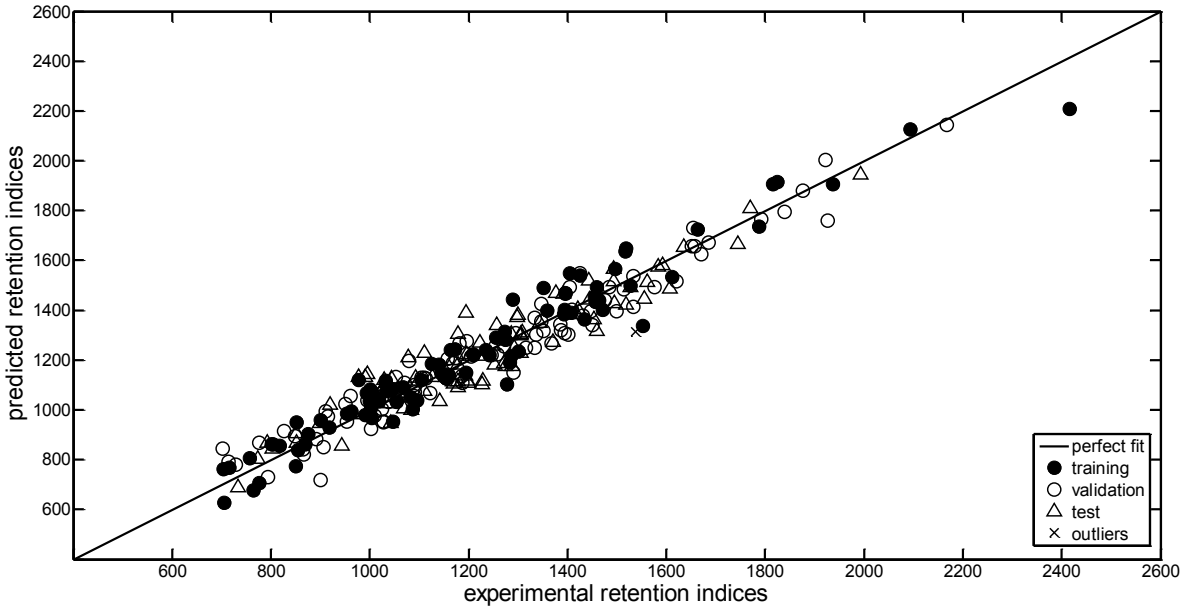

Figure 2. Experimental *vs.* predicted retention indices for fragrances in HP5–MS stationary phase.

Moreover, the additional validation criteria presented below confirmed that a predictive QSPR model was achieved for the *I* prediction of fragrances in the HP5–MS stationary phase. Figure 2 confirms the linear relationship between the experimental and the predicted retention indices.

$R^2_{loo} > 0.5$ (0.94) and the $R^2_{test} > 0.6$ (0.90)
$1 - R^2_0 / R^2_{test} < 0.1$ (0.00) or $1 - R'^2_0 / R^2_{test} < 0.1$ (0.01)
$0.85 \le k(1.00) \le 1.15$ and $0.85 \le k'(0.99) \le 1.15$
$R^2_m > 0.5$ (0.89)

In the HP5–MS model, there was only one outlier compound, *2(4H)–Benzofuranone, 5,6,7,7a–tetrahydro–4,4,7a–trimethyl–*. Its chemical structure and *I* value at the source were correct. In addition, numerical data on the predicted retention indexes provided by Eq. 2, as well as descriptor values involved in the QSPR model are shown in Table 4S.

The three molecular descriptors involved in Eq. 2 corresponded to one 2D matrix–based descriptor (*SpPos_B(p)*) and descriptors related to two atom–centered fragments (*C–001* and *C–008*) selected for the DB–225MS QSPR phase. In addition, the maximum coefficient of determination ($R^2_{ij\,max} = 0.03$) between the *SpPos_B(p)* and *C–001* descriptors, indicated the absence of multicollinearity. Moreover, the degree of contribution of each descriptor indicated

that the spectral positive sum from the Burden matrix weighted by polarizability was the more important descriptor: 0.98 (*SpPos_B(p)*) > 0.21 (*C–001*) > 0.11 (*C–008*).

SpPos_B(p) [53, 54] was a topological index calculated as the sum of the positive eigenvalues (graph invariant) from the Burden matrix weighted by polarizability. The higher the sum of the positive eigenvalues, the greater the presence of atoms exhibiting high polarizability. Thus, the fragrance had the ability to interact with the stationary phase through dipoles and consequently, it was retained for more time (e.g. the *Ethyl oleate* fatty acid ester). As presented for the DB–225MS stationary phase, retention indices of fragrances were inversely related to the presence of CH_3 fragments (*C–001*), and directly related to the existence of CH fragments bonded to two carbon atoms and an electronegative atom (*C–008*).

The AD assessment of the HP5–MS model indicated that predictions of retention indices were restricted to only molecules having a leverage value below the warning leverage ($h^* = 0.068$). No test molecules fell outside the AD. Thus, the QSPR model of Eq. 2 was reliable to predict the retention indices of the whole test set. Subsequently, the model developed in Eq. 2 was used to predict the retention index of the *1–Propanol, 2–methyl–* ($I = 589.9$ and $h_{ii} = 0.031$), the *Ethyl Acetate* ($I = 681.0$ and $h_{ii} = 0.024$) and *Isopropyl acetate* ($I = 831.5$ and $h_{ii} = 0.076$) for which this property was extrapolated by Yan et al.

The only compound exhibiting a leverage value above the warning leverage, i.e., falling outside the AD of the model was *Isopropyl acetate*. Consequently, for the other two compounds, the predicted retention indices were reliable and they were close to the extrapolated ones (refer to Table 1S).

3.3. HP–1 Stationary phase

Following the same workflow described for both the DB–225MS and the HP5–MS datasets, the 262 molecules in the HP–1 dataset were divided into a training set of 88 fragrances, and validation and test sets containing 87 compounds each. An optimal three–parametric quantitative structure–property relationship was found by the RM variable subset selection.

$$I_{HP-1} = 296.1 + 153.7 \times SpPos\_A + 543.4 \times MATS1p - 9.5 \times F02[C-C] \qquad \text{(Eq. 3)}$$

$N_{train} = 88$, $R^2_{train} = 0.95$, $RMSD_{train} = 75.1$

$N_{val} = 87$, $R^2_{val} = 0.95$, $RMSD_{val} = 62.2$

$N_{test} = 87$, $R^2_{test} = 0.92$, $RMSD_{test} = 65.1$
$R^2_{loo} = 0.95$, $RMSD_{loo} = 69.7$, $R^2_{lmo} = 0.97$, $RMSD_{lmo} = 74.4$
$RMSD_{rand} = 282.4$, $o(3S) = 2$, $R^2_{ij\,max} = 0.60$

The additional validation criteria presented below indicated that the developed QSPR model was predictive of retention indices measured in the HP–1 stationary phase. Figure 3 confirms the linear relation between experimental and predicted retention indices.

$R^2_{loo} > 0.5$ (0.95) and the $R^2_{test} > 0.6$ (0.92)
$1 - R^2_0 / R^2_{test} < 0.1$ (0.00) or $1 - R'^2_0 / R^2_{test} < 0.1$ (0.00)
$0.85 \le k(1.00) \le 1.15$ and $0.85 \le k'(0.99) \le 1.15$
$R^2_m > 0.5$ (0.89)

Numerical values of the three molecular descriptors and the predicted retention indices calculated in Eq. 3 are presented in Table 5S. In this model, there were two molecules that exhibited values above three times the *RMSD*: *2H–Pyran–2–one, tetrahydro–6–nonyl–* and *Cinnamyl cinnamate*.

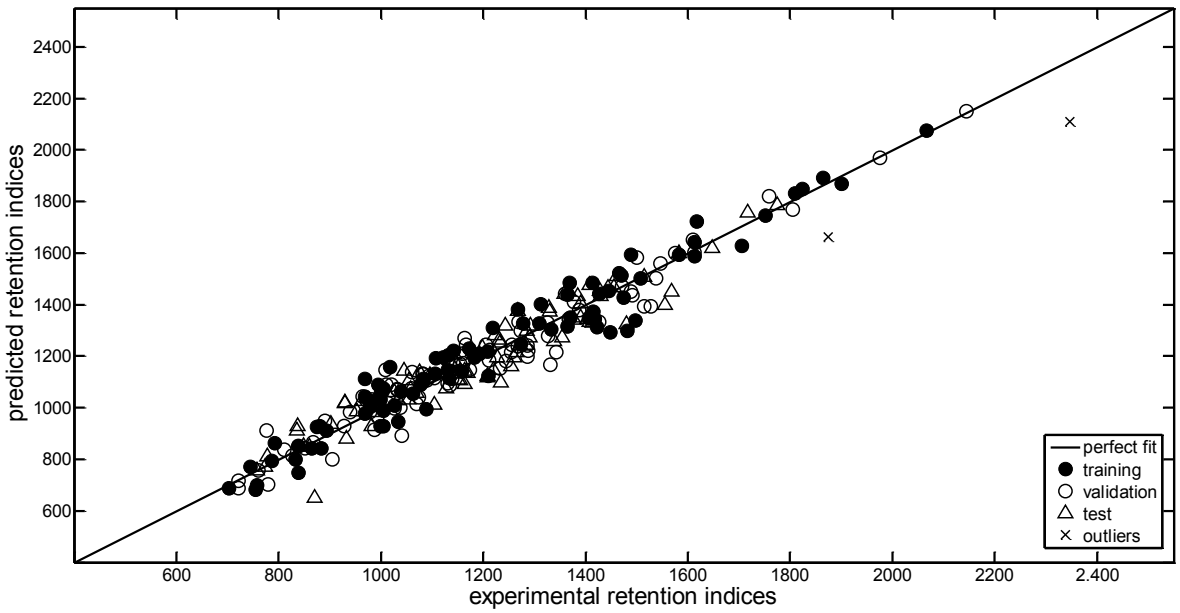

Figure 3. Experimental *vs.* predicted retention indices for fragrances in HP–1 stationary phase.

Descriptors presented in Eq. 3 correspond to 2D matrix–based descriptors (*SpPos_A*), 2D autocorrelations (*MATS1p*), and 2D Atom Pairs (*F02[C–C]*). Further evaluation of the maximum

coefficient of determination ($R^2_{ij\,max} = 0.60$), indicated a moderate correlation between the *SpPos_A* and *F02[C–C]* descriptors. Moreover, the degree of contribution of each of the descriptors in modeling the *I* property in the HP–1 stationary phase were: 1.06 (*SpPos_A*) > 0.18 (*MATS1p*) > 0.15 (*F02[C–C]*).

The spectral positive sum from the adjacency matrix (*SpPos_A*) [54] was a topological index obtained by summing the positive eigenvalues (graph invariant) from the adjacency matrix. This descriptor demonstrated that the retention index of a molecule was related to its complexity (e.g. *Ethyl Oleate* or *Cinnamyl cinnamate*). The Moran autocorrelation of lag 1 weighted by polarizability (*MATS1p*) was a descriptor calculated by applying the Moran coefficient [56] to the H–filled molecular graph weighted by atomic polarizabilities (*p*). This descriptor provided information regarding the distribution of polarizability along the topological structure of a fragrance. High retention indices were related to positive values of the Moran coefficient (positive spatial autocorrelations); that is, compounds containing atoms with similarly polarizability at lag 1. Finally, the *F02[C–C]* descriptor [59] indicated that the *I* was inversely related to the number of carbon–carbon atom pairs in the molecule separated by two topological distances; in other words, the retention index tended to diminish when increasing the size of the hydrocarbon skeleton in the fragrance (e.g. *Cedrol*).

The applicability domain analysis for Eq. 3 defined a warning leverage of $h^* = 0.069$. The only test molecule exhibiting a leverage value outside the AD was *Cyclohexane, 1–ethenyl–1–methyl–2, 4–bis (1–methylethenyl)–,[1S–(1α,2β,4β)]–* with a leverage value (h_{ii}) equal to 0.072. Since Yan et al. had not provided experimental retention indices for seven compounds, the QSPR model of Eq. 3 was used to predict them. The predicted values were: *2–Propenoic acid, ethyl ester* (I = 700.9 and h_{ii} = 0.048), *Acetoin* (I = 691.4 and h_{ii} =0.004), *Pentanal* (I = 725.2 and h_{ii} = 0.007), *Propanoic acid* (I = 619.7 and h_{ii} = 0.006), *1–Propanol, 2–methyl–* (I = 611.5 and h_{ii} = 0.004), *Ethyl Acetate* (I = 524.8 and h_{ii} = 0.097), and *Isopropyl acetate* (I = 617.5 and h_{ii} =0.055). The only compound exhibiting a leverage value above the warning leverage was *Ethyl Acetate*. Consequently, for all the other six compounds the predicted retention indices were reliable. In addition, these QSPR predicted retention index values were close to the extrapolated ones by Yan et al. (refer to Table 1S).

4. Conclusions

In this work, we developed conformation–independent QSPR models for the retention indices of fragrances measured in three stationary phases of different polarity. The supervised method enabled the selection of optimal subsets of four Dragon descriptors for the DB–225MS stationary phase and three Dragon descriptors for both the HP5–MS and the HP–1 stationary phases. The three models were validated by means of cross–validation procedures, the prediction of the test molecules, as well as other recommended validation criteria. Thus, these models could be useful for researchers working on the prediction of retention indices of fragrances. Moreover, the conformation–independent QSPR approach represents a useful alternative when modeling gas–chromatographic retention indices with models solely based on topological and constitutional representations of the fragrances.

Acknowledgments

Cristian Rojas is grateful for his PhD Fellowship from the National Secretary of Higher Education, Science, Technology and Innovation (SENESCYT) from the Republic of Ecuador. Pablo R. Duchowicz wishes to thank the National Scientific and Technical Research Council of Argentina (CONICET) for the project grant PIP11220130100311, and the Minister of Science, Technology and Productive Innovation for the use of the electronic library facilities. Pablo R. Duchowicz and Reinaldo Pis Diez are members of the Scientific Researcher Career of CONICET.

Supporting information description

Supporting information for this work is presented in Supplementary Tables. Table 1S details the names, CAS registry number, and the experimental retention indices for the 269 fragrances measured in the DB–225MS, HP5–MS and HP–1 stationary phases. Table 2S lists the mathematical equations used in the present study. In addition, the predicted retention indices, molecular descriptors values and the training, validation and test set assignments are reported in Tables 3S, 4S and 5S for the DB–225MS, HP5–MS and HP–1 stationary phases, respectively.

References

[1] H. Surburg, J. Panten, Common fragrance and flavor materials: preparation, properties and uses, John Wiley & Sons, **2016**.

[2] T. Cserháti, Chromatography of aroma compounds and fragrances, Springer, **2010**.

[3] R. Kaliszan, Chromatographia, 10 (**1977**) 529–531.

[4] R. Kaliszan, H. Foks, Chromatographia, 10 (**1977**) 346–349.

[5] Y. Michotte, D. Massart, Journal of Pharmaceutical Sciences, 66 (**1977**) 1630–1632.

[6] R. Kaliszan, Chemical Reviews, 107 (**2007**) 3212–3246.

[7] K. Héberger, Journal of Chromatography A, 1158 (**2007**) 273–305.

[8] W.P. Eckel, T. Kind, Analytica chimica acta, 494 (**2003**) 235–243.

[9] D. Wang, X. Xu, S. Chu, D. Zhang, Chemosphere, 53 (**2003**) 495–503.

[10] L.–T. Qin, S.–S. Liu, H.–L. Liu, J. Tong, Journal of Chromatography A, 1216 (**2009**) 5302–5312.

[11] S. Riahi, E. Pourbasheer, M.R. Ganjali, P. Norouzi, Journal of Hazardous Materials, 166 (**2009**) 853–859.

[12] M. Mohammadhosseini, H.A. Zamani, H. Akhlaghi, M. Nekoei, Journal of Essential Oil Bearing Plants, 14 (**2011**) 559–573.

[13] H. Noorizadeh, M. Noorizadeh, Medicinal Chemistry Research, 21 (**2012**) 1997–2005.

[14] C. Zhao, Y. Liang, X. Wang, L. Liu, D. Yuan, T. Wu, Q. Xu, Current Analytical Chemistry, 8 (**2012**) 168–179.

[15] L.–T. Qin, S.–S. Liu, F. Chen, Q.–F. Xiao, Q.–S. Wu, Chemosphere, 90 (**2013**) 300–305.

[16] N. Goudarzi, D. Shahsavani, F. Emadi–Gandaghi, M.A. Chamjangali, Journal of Chromatography A, 1333 (**2014**) 25–31.

[17] M. Mohammadhosseini, Journal of Chemical Health Risks, 4 (**2014**) 75–95.

[18] K. Héberger, M. Görgényi, M. Sjöström, Chromatographia, 51 (**2000**) 595–600.

[19] K. Héberger, M. Görgényi, Journal of Chromatography A, 845 (**1999**) 21–31.

[20] D. Ballabio, V. Consonni, PLS–DA, Analytical Methods, 5 (**2013**) 3790–3798.

[21] T. Körtvélyesi, M. Görgényi, K. Héberger, Analytica chimica acta, 428 (**2001**) 73–82.

[22] B. da Silva Junkes, R.D.d.M.C. Amboni, R. Augusto, Internet Electronic Journal of Molecular Design, 2 (**2003**) 33–49.

[23] B. Ren, Chemometrics and Intelligent Laboratory Systems, 66 (**2003**) 29–39.

[24] B.d.S. Junkes, R.D.d. Amboni, R.A. Yunes, V.E. Heinzen, Journal of the Brazilian Chemical Society, 15 (**2004**) 183–189.

[25] C. Zhou, C. Nie, Chromatographia, 66 (2007) 545–554.

[26] E. Konoz, M.H. Fatemi, R. Faraji, Journal of chromatographic science, 46 (**2008**) 406–412.

[27] É.S. Souza, C.A. Kuhnen, B.d.S. Junkes, R.A. Yunes, V.E.F. Heinzen, Journal of Chemometrics, 23 (**2009**) 229–235.

[28] Organisation for Economic Co–operation and Development, Guidance Document On The Validation Of (Quantitative)Structure–Activity Relationships [(Q)SAR] Models, OECD Publishing. Paris, **2007**.

[29] J. Yan, X.–B. Liu, W.–W. Zhu, X. Zhong, Q. Sun, Y.–Z. Liang, Chromatographia, 78 (**2015**) 89–108.

[30] M. Berthold, N. Cebron, F. Dill, T. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel, KNIME: The Konstanz Information Miner, in: C. Preisach, H. Burkhardt, L. Schmidt–Thieme, R. Decker (Eds.) Data Analysis, Machine Learning and Applications, Springer Berlin Heidelberg, **2008**, pp. 319–326.

[31] NCI/CADD Group, Chemical Identifier Resolver, **2013**.

[32] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, S.H. Bryant, Nucleic acids research, 44 (**2015**) D1202–D1213.

[33] P.J. Linstrom, W.G. Mallard, NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg MD, **2001**.

[34] H.E. Pence, A. Williams, Journal of Chemical Education, 87 (**2010**) 1123–1124.

[35] C. Rojas, P.R. Duchowicz, P. Tripaldi, R. Pis Diez, Chemometrics and Intelligent Laboratory Systems, 140 (**2015**) 126–132.

[36] C. Rojas, P.R. Duchowicz, P. Tripaldi, R. Pis Diez, Journal of Chromatography A, 1422 (**2015**) 277–288.

[37] R. Todeschini, V. Consonni, Molecular Descriptors for Chemoinformatics, WILEY–VCH, Weinheim, 2009.

[38] Kode srl., Dragon (version 7). Software for Molecular Descriptor Calculation, https://chm.kode–solutions.net, **2016**.

[39] P.R. Duchowicz, E.A. Castro, F.M. Fernández, M.P. González, Chemical Physics Letters, 412 (**2005**) 376–380.

[40] P.R. Duchowicz, E.A. Castro, F.M. Fernández, MATCH Communications in Mathematical and in Computer Chemistry, 55 (**2006**) 179–192.

[41] M. Eklund, U. Norinder, S. Boyer, L. Carlsson, Journal of Chemical Information and Modeling, 54 (**2014**) 837–843.

[42] C. Rojas, P. Tripaldi, P.R. Duchowicz, International Journal of Quantitative Structure–Property Relationships, 1 (**2016**) 76–90.

[43] T.M. Martin, P. Harten, D.M. Young, E.N. Muratov, A. Golbraikh, H. Zhu, A. Tropsha, Journal of Chemical Information and Modeling, 52 (**2012**) 2570–2578.

[44] C. Rücker, G. Rücker, M. Meringer, Journal of Chemical Information and Modeling, 47 (**2007**) 2345–2357.

[45] A. Golbraikh, A. Tropsha, Journal of Molecular Graphics and Modelling, 20 (**2002**) 269–276.

[46] P. Gramatica, QSAR & Combinatorial Science, 26 (**2007**) 694–701.

[47] L. Eriksson, J. Jaworska, A.P. Worth, M.T. Cronin, R.M. McDowell, P. Gramatica, Environmental health perspectives, 111 (**2003**) 1361–1375.

[48] N.R. Draper, H. Smith, Applied Regression Analysis, New York, **1981**.

[49] The MathWorks Inc., MatLab, http://www.mathworks.com.

[50] R. Hoffmann, V.I. Minkin, B.K. Carpenter, Bulletin de la Société chimique de France, 133 (**1996**) 117–130.

[51] L. Soják, V.G. Berezkin, J. Janák, Journal of Chromatography A, 209 (**1981**) 15–20.

[52] V.J. Barwick, Journal of Chromatography A, 849 (**1999**) 13–33.

[53] I. Gutrnan, D. Vidović, N. Cmiljanović, S. Milosavljević, S. Radenković, Indian Journal of Chemistry, 42A (**2003**) 1309–1311.

[54] V. Consonni, R. Todeschini, MATCH Communications in Mathematical and in Computer Chemistry, 60 (**2008**) 3–14.

[55] A.K. Ghose, V.N. Viswanadhan, J.J. Wendoloski, The Journal of Physical Chemistry A, 102 (**1998**) 3762–3772.

[56] P.A. Moran, Biometrika, 37 (**1950**) 17–23.

[57] D.A. Carlson, U.R. Bernier, B.D. Sutton, Journal of Chemical Ecology, 24 (**1998**) 1845–1865.

[58] A.R. Katritzky, K. Chen, U. Maran, D.A. Carlson, Analytical Chemistry, 72 (**2000**) 101–109.

[59] R.E. Carhart, D.H. Smith, R. Venkataraghavan, Journal of Chemical Information and Computer Sciences, 25 (**1985**) 64–73.

APENDIX to *Quantitative structure–property relationships for predicting the retention indices of fragrances on stationary phases of different polarity, C. Rojas et al.*

Table 1S. Chemical names, CAS registry numbers, and experimental retention indices for the 269 fragrances measured in the DB–225MS, HP5–MS and HP–1 stationary phases.

Name	CAS number	DB–225MS	HP5–MS	HP–1
1,2–Cyclopentanedione, 3–methyl–	000765–70–8	1547	1027	994
1,6,10–Dodecatrien–3–ol, 3,7,11–trimethyl–	007212–44–4	1866	1534	1515
1,6–Octadien–3–ol, 3,7–dimethyl–	000078–70–6	1396	1099	1083
10–Undecenal	000112–45–8	1609	1299	1274
1–Butanol, 3–methyl–	000123–51–3	1251	734	722
1–Butanol, 3–methyl–, acetate	000123–92–2	1080.5	876	859.5
1–Butanol, 3–methyl–, formate	000110–45–2	1006	792	774
1–Butanol, 3–methyl–, propanoate	000105–68–0	1168	968	952
1–Cyclohexene–1–methanol, 4–(1–methylethenyl)–	000536–59–4	1791	1301	1274
1–Decanol	000112–30–1	1577	1271	1254
1–Hexanol	000111–27–3	1178.5	867	852
1–Hexanol, 2–ethyl–	000104–76–7	1333	1028	1014
1–Hexanol, 3,5,5–trimethyl–	003452–97–9	1358	1047	1033
1–Octanol	000111–87–5	1248.7	1069.3	1054
1–Octen–3–ol	003391–86–4	1278	978	963
2(3H)–Furanone, 5–butyldihydro–	000104–50–7	1949	1259	1210
2(3H)–Furanone, 5–ethyldihydro–	000695–06–7	1709	1054.5	1005
2(3H)–Furanone, 5–heptyldihydro–	000104–67–6	2321	1576	1527.5
2(3H)–Furanone, 5–hexyldihydro–	000706–14–9	2197.5	1471	1422
2(3H)–Furanone, 5–methyl–	000591–12–8	1325	869	833
2(3H)–Furanone, dihydro–5–methyl–	000108–29–2	1569	953	905
2(3H)–Furanone, dihydro–5–propyl–	000105–21–5	1822	1154	1105
2(4H)–Benzofuranone, 5,6,7,7a–tetrahydro–4,4,7a–trimethyl–	015356–74–8	2477	1537	1481
2(5H)–Furanone, 5–ethyl–3–hydroxy–4–methyl–	000698–10–2	1974	1195	1158
2,4–Decadienal, (E,E)–	025152–84–5	1776	1316	1287
2,4–Heptadienal, (E,E)–	004313–03–5	1430	1010	981
2,6,6–Trimethyl–2–cyclohexene–1,4–dione	001125–21–9	1646	1144	1104
2,6–Octadien–1–ol, 3,7–dimethyl–, (Z)–	000106–25–2	1582	1228	1208
2,6–Octadien–1–ol, 3,7–dimethyl–, formate, (E)–	000105–86–2	1564	1306	1284
2–Acetyl–5–methylfuran	001193–79–9	1507	1038	1005
2–Acetylthiazole	024295–03–2	1476	1019	981
2–Butanone, 4–(4–hydroxyphenyl)–	005471–51–2	2675	1553	1498
2–Buten–1–ol, 3–methyl–	000556–82–1	1113	776	758
2–Buten–1–one, 1–(2,6,6–trimethyl–1–cyclohexen–1–yl)–	035044–68–9	1832	1418	1390
2–Buten–1–one, 1–(2,6,6–trimethyl–2–cyclohexen–1–yl)–, (E)–	024720–09–0	1789	1393	1370
2–Cyclopenten–1–one, 3–methyl–2–(2–pentenyl)–, (Z)–	000488–10–8	1938	1401	1364
2–Ethyl–3–methoxypyrazine	025680–58–4	1301	1053	1032
2–Furancarboxaldehyde, 5–methyl–	000620–02–0	1451	963	928
2–Furanmethanol, acetate	000623–17–6	1367	995	964

2–Furfurylthiol	000098–02–2	1259	911	883
2–Heptanone	000110–43–0	1166	891	868
2–Hexen–1–ol, (E)–	000928–95–0	1202	865	848
2–Hexen–1–ol, acetate, (E)–	002497–18–9	1261	1015	995
2–Hexenal, (E)–	006728–26–3	1187	852	827
2H–Pyran, tetrahydro–4–methyl–2–(2–methyl–1–propenyl)–	016409–43–1	1326	1111	1095
2H–Pyran–2–one, 6–hexyltetrahydro–	000710–04–3	2396	1607	1555
2H–Pyran–2–one, tetrahydro–6–methyl–	000823–22–3	1818	1095	1041
2H–Pyran–2–one, tetrahydro–6–nonyl–	002721–22–4	2753	1926	1874
2H–Pyran–2–one, tetrahydro–6–pentyl–	000705–86–2	2270	1499	1448
2H–Pyran–2–one, tetrahydro–6–propyl–	000698–76–0	2024	1287	1235
2–Methylheptanoic acid	001188–02–9	1564	1141	1128
2–Naphthyl methyl ketone	000093–08–3	2440	1620	1568
2–Nonanone	000821–55–6	1383	1091	1070
2–Nonen–1–ol, (E)–	031502–14–4	1517	1168	1150
2–Nonenal, (Z)–	060784–31–8	1520	1159	1134
2–Octanone	000111–13–7	1274.5	991	969
2–Octenal, (E)–	002548–87–0	1427.5	1057	1026.5
2–Octynoic acid, methyl ester	000111–12–6	1557	1202	1169
2–Propanone, 1–(4–methoxyphenyl)–	000122–84–9	2066	1386	1339
2–Propen–1–ol, 3–phenyl–	000104–54–1	1999	1306	1268
2–Propen–1–ol, 3–phenyl–, propanoate	000103–56–0	2105	1555	1515
2–Propenal, 3–(2–methoxyphenyl)–	001504–74–1	2333	1533	1480
2–Tridecanone	000593–08–8	1818	1495	1474
2–Undecanone	000112–12–9	1573	1293	1272
3–(4–Isopropylphenyl)–2–methylpropionaldehyde	000103–95–7	1959	1464	1426
3–(Methylthio)propanoic acid methyl ester	013532–18–8	1424	1024	993
3,4–Hexanedione	004437–51–8	1083	802	777
3–Hexen–1–ol	000544–12–7	1193	855	838
3–Hexen–1–ol, acetate, (Z)–	003681–71–8	1298	1006	979.3
3–Hexen–1–ol, formate, (Z)–	033467–73–1	1177	920	902
3–Hexenoic acid, (E)–	001577–18–0	1508	1003	988
3–Phenylpropanol	000122–97–4	1824	1233	1198
4–Heptenal, (Z)–	006728–31–0	1194	901	874
4H–Pyran–4–one, 2–ethyl–3–hydroxy–	004940–11–8	1777	1197	1161
4–Methylthiazole	000693–95–8	1136	817	793
5,6,7,8–Tetrahydroquinoxaline	034413–35–9	1637	1209	1172
5,9–Undecadien–2–one, 6,10–dimethyl–	000689–67–8	1804	1434	1408
5,9–Undecadien–2–one, 6,10–dimethyl–, (E)–	003796–70–1	1831	1453	1426
5H–5–Methyl–6,7–dihydrocyclopentapyrazine	023747–48–0	1529	1140	1107
5–Heptenal, 2,6–dimethyl–	000106–72–9	1498	1054	1033
5–Methyl–2–phenyl–2–hexenal	021834–92–4	2065	1493	1455
5–Thiazoleethanol, 4–methyl–	000137–00–8	2049	1277	1233
6–Octen–1–ol, 3,7–dimethyl–, acetate	000150–84–5	1584	1352	1333
6–Octen–1–ol, 3,7–dimethyl–, formate	000105–85–1	1540	1275	1256

6–Octen–1–ol, 3,7–dimethyl–, propanoate	000141–14–0	1705	1448	1429
6–Octenal, 3,7–dimethyl–, (R)–	002385–77–5	1460	1153	1130
9,12–Octadecadienoic acid (Z,Z)–, methylester	000112–63–0	2466	2094	2067
Acetic acid, 2–phenylethyl ester	000103–45–7	1710.5	1257.5	1223
Acetic acid, 4–methylphenyl ester	000140–39–6	1590	1170	1138
Acetic acid, decyl ester	000112–17–4	1617	1409	1391
Acetic acid, heptyl ester	000112–06–1	1335	1111	1094
Acetic acid, nonyl ester	000143–13–5	1532	1309	1292
Acetic acid, octyl ester	000112–14–1	1440	1210	1193
Acetic acid, phenyl–, isopentyl ester	000102–19–2	1950	1497	1470
Acetophenone	000098–86–2	1532	1067	1032
Acetylpyrazine	022047–25–2	1470	1022	987
Allyl nonanoate	007493–72–3	1593	1377	1358
Anisyl propionate	007549–33–9	2092	1514	1471
Benzaldehyde	000100–52–7	1399	961	929
Benzaldehyde, 2–hydroxy–	000090–02–8	1504	1044	1009
Benzaldehyde, 4–(1–methylethyl)–	000122–03–2	1716	1242	1209
Benzaldehyde, 4–ethoxy–	010031–82–0	1955	1333	1288
Benzaldehyde, 4–ethyl–	004748–78–1	1636	1179	1145
Benzene, 1,1'–[oxybis(methylene)]bis–	000103–50–4	2256	1654	1610
Benzene, 1,2–dimethoxy–4–(1–propenyl)–	000093–16–3	2050	1457	1418
Benzene, 1,3–dimethoxy–	000151–10–0	1580	1168	1135
Benzene, 1–methoxy–4–methyl–	000104–93–8	1333	1021	997
Benzene, 2–methoxy–4–methyl–1–(1–methylethyl)–	001076–56–8	1503	1235	1213
Benzene, ethoxy–	000103–73–1	1286	993	969
Benzeneacetaldehyde, α–ethylidene–	004411–89–6	1858	1274	1233
Benzeneacetic acid, 2–methylpropyl ester	000102–13–6	1829	1392	1360
Benzeneacetic acid, 2–phenylethyl ester	000102–20–5	2709	1922	1865
Benzeneacetic acid, methyl ester	000101–41–7	1598	1178	1144
Benzeneacetic acid, phenylmethyl ester	000102–16–9	2576	1815	1759
Benzeneethanol, α,α–dimethyl–	000100–86–7	1581	1158	1130
Benzenemethanol, 4–methoxy–	000105–13–5	1994	1284	1244
Benzenemethanol, 4–methoxy–, acetate	000104–21–2	2004	1421	1377
Benzenemethanol, 4–methoxy–, formate	000122–91–8	1933	1334	1291
Benzenemethanol,α–methyl–, acetate	000093–92–5	1571	1194	1164
Benzenepropanal	000104–53–0	1693	1163	1122
Benzoic acid	000065–85–0	1314	1178	1155
Benzoic acid, 2–hydroxy–, 2–methylpropyl ester	000087–19–4	1920	1475	1444
Benzoic acid, 2–hydroxy–, ethyl ester	000118–61–6	1697	1273	1243
Benzoic acid, 2–hydroxy–, phenylmethyl ester	000118–58–1	2633.5	1876	1824
Benzoic acid, ethyl ester	000093–89–0	1547	1172	1142
Benzophenone	000119–61–9	2372	1635	1583
Benzyl alcohol	000100–51–6	1033	1034	1004
Benzyl Benzoate	000120–51–4	2493	1770	1717
Bicyclo[2.2.1]heptan–2–ol, 1,7,7–trimethyl–, acetate, (1S–endo)– Isobornyl acetate	005655–61–8 000125–12–2	1565	1289	1268

Butanal, 3–methyl–	000590–86–3	897	900	870
Butanedioic acid, dimethyl ester	000106–65–0	1467	1032	999
Butanoic acid	000107–92–6	1125	794	779
Butanoic acid, 2–methyl–, ethyl ester	007452–79–1	1027.5	849	835.5
Butanoic acid, 2–methylpropyl ester	000539–90–2	1145	954	939
Butanoic acid, 3,7–dimethyl–2,6–octadienyl ester, (E)–	000106–29–6	1854.5	1561.5	1537.5
Butanoic acid, 3–methyl–	000503–74–2	1286	850	839
Butanoic acid, 3–methyl–, 2–phenylethyl ester	000140–26–1	1931	1494	1461
Butanoic acid, 3–methyl–, 3–methylbutyl ester	000659–70–1	1292	1104.5	1090
Butanoic acid, 3–methyl–, 3–phenyl–2–propenyl ester	000140–27–2	2226	1686	1648
Butanoic acid, 3–methyl–, butyl ester	000109–19–3	1238	1045	1027
Butanoic acid, 3–methyl–, ethyl ester	000108–64–5	1038	852	838
Butanoic acid, 3–methyl–, phenylmethyl ester	000103–38–8	1816	1396	1364
Butanoic acid, 3–methylbutyl ester	000106–27–4	1255	1055	1039
Butanoic acid, butyl ester	000109–21–7	1199	995	979
Butanoic acid, ethyl ester	000105–54–4	1045	804	787
Butanoic acid, phenylmethyl ester	000103–37–7	1782	1347	1313
Butanoic acid, propyl ester	000105–66–8	1096	899	881
Caryophyllene	000087–44–5	1586	1426	1413
Cedrol	000077–53–2	2072	1611	1583
Cinnamaldehyde, (E)–	014371–10–9	1922	1272	1227
Cinnamaldehyde, α–pentyl–	000122–40–7	2206	1651	1613
Cinnamyl cinnamate	000122–69–0	3127	2416	2347
Citronellol	000106–22–9	1566	1227	1208
Citronellyl butyrate	000141–16–2	1784	1528	1508
Creosol	000093–51–6	1715	1193	1163
Cyclohexane, 1–ethenyl–1–methyl–2, 4–bis (1–methylethenyl)–,[1S–(1α,2β,4β)]–	000515–13–9	1545	1426	1384
Cyclohexanol, 5–methyl–2–(1–methylethyl)–, (1α,2α,5α)–	000491–02–1 015356–70–4	1512	1174	1155
Cyclohexanone, 2–(1–mercapto–1–methylethyl)–5–methyl–	038462–22–5	1880	1367	1331
Cyclohexanone, 5–methyl–2–(1–methylethyl)–, (2R–cis)–	001196–31–2	1521	1155	1130
Cyclohexene, 1–methyl–4–(1–methylethylidene)–	000586–62–9	1229	1089	1077
Cyclopentaneacetic acid, 3–oxo–2–pentyl–, methyl ester	024851–98–7	2287	1657	1613
D–Carvone	002244–16–8	1701	1246	1212
Decanal	000112–31–2	1494	1205	1183
Decanedioic acid, diethyl ester	000110–40–7	2293	1787	1751
Decanoic acid, ethyl ester	000110–38–3	1586	1394	1377
Diphenyl ether	000101–84–8	1863	1404	1369
D–Limonene	005989–27–5	1151	1029	1020
Dodecanal	000112–54–9	1713	1408	1387
Dodecanoic acid, ethyl ester	000106–33–2	1827	1593	1576
Ethane, 1,1–diethoxy–	000105–57–7	1002	729	721
Ethanone, 1–(2,4–dimethylphenyl)–	000089–74–7	1719	1253	1220
Ethanone, 1–(2–pyridinyl)–	001122–62–9	1458	1033	999

Ethanone, 1–(3–pyridinyl)–	000350–03–8	1705	1111	1074
Ethanone, 1–(4–methylphenyl)–	000122–00–9	1679	1186	1151
Ethyl Oleate	000111–62–6	2475	2167	2144
Ethyl Vanillin	000121–32–4	2315	1459	1407
Eucalyptol	000470–82–6	1212	1032	1018
Eugenol	000097–53–0	1920	1359	1325
Formic acid, pentyl ester	000638–49–3	1050	826	811
Formic acid, phenylmethyl ester	000104–57–4	1519	1078	1045
Furan, 2–[(methylthio)methyl]–	001438–91–1	1352	1001	972
Furan, 2–methyl–5–(methylthio)–	013678–59–6	1215	951	930
Heptanoic acid	000111–14–8	1545	1084	1070
Heptanoic acid, 3–methylbutyl ester	000109–25–1	1876	1347	1330
Heptanoic acid, ethyl ester	000106–30–9	1305	1127	1081
Hexadecanoic acid, ethyl ester	000628–97–7	2246	1993	1975
Hexanal	000066–25–1	1057	803	778
Hexanoic acid, 2–methyl–	004536–23–6	1484	1043	1027
Hexanoic acid, 2–propenyl ester	000123–68–2	1318	1079	1061
Hexanoic acid, ethyl ester	000123–66–0	1202	999	981
Hexanoic acid, pentyl ester	000540–07–8	1506	1287	1269
Humulene	006753–98–6	1916	1460	1445
Hydrocoumarin	000119–84–6	2197	1387	1327
Isoamyl cinnamate	007779–65–9	2304	1745	1705
Isobutyl acetate	000110–19–0	970	773	757
Isopentyl hexanoate	002198–61–0	1458	1249	1232
Isophorone	000078–59–1	1581	1122	1089
Isopropyl myristate	000110–27–0	2037	1824	1809
Menthyl isovalerate	016409–46–4	1744	1518	1500
Methional	003268–49–3	1368	907	865
Methyl isovalerate	000556–24–1	970	777	761
Methyl salicylate	000119–36–8	1594	1196	1166
Methyleugenol	000093–15–2	1888.5	1404.5	1367.5
Naphthalene, 2–ethoxy–	000093–18–5	2100	1528	1489
Naphthalene, 2–methoxy–	000093–04–9	2050.5	1454.5	1414.5
n–Decanoic acid	000334–48–5	1885	1370	1354
Nonanal	000124–19–6	1383	1104	1082
Nonanoic acid	000112–05–0	1777	1274	1260
Nonanoic acid, ethyl ester	000123–29–5	1509	1295	1278
Nonanoic acid, methyl ester	001731–84–6	1452	1223	1205
n–Propyl acetate	000109–60–4	918	715	703
Octanal, 7–hydroxy–3,7–dimethyl–	000107–75–5	1864	1288	1254
Octanoic acid	000124–07–2	1649.5	1179	1163.5
Octanoic acid, 3–methylbutyl ester	002035–99–6	1641	1445	1428
Octanoic acid, ethyl ester	000106–32–1	1410	1196	1179
Octanoic acid, methyl ester	000111–11–5	1347	1125	1106
Oxacycloheptadec–8–en–2–one, (8Z)	000123–69–3	2460.5	1937.5	1900.5
Oxacyclohexadecan–2–one	000106–02–5	2591	1839	1805

Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl ester Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl ester, cis–	000077–83–8 019464–95–0	2074.5	1517	1489
p–Cymene	000099–87–6	1218	1025	1010
Pentanoic acid, 3–methyl–	000105–43–1	1297	944	933
Pentanoic acid, butyl ester	000591–68–4	1302	1093	1075
Pentanoic acid, ethyl ester	000539–82–2	1037	852	837
Pentyl octanoate	000638–25–5	1714	1484	1466
Phenol, 2–(1–methylethyl)–	000088–69–7	1788	1199	1174
Phenol, 2–methoxy–	000090–05–1	1577	1090	1058
Phenol, 2–methyl–5–(1–methylethyl)–	000499–75–2	1866	1291	1265
Phenylethyl Alcohol	000060–12–8	1660	1114	1082
Piperonal	000120–57–0	2043	1336	1286
Propanedioic acid, diethyl ester	000105–53–3	1477	1070	1038
Propanethioic acid, S–(2–furanylmethyl) ester	059020–85–8	1689	1256	1219
Propanoic acid, 2–methyl–	000079–31–2	1181	765	754
Propanoic acid, 2–methyl–, 2–phenylethyl ester	000103–48–0	1818	1396	1367
Propanoic acid, 2–methyl–, 3–phenyl–2–propenyl ester	000103–59–3	2105	1584	1547
Propanoic acid, 2–methyl–, 4–formyl–2–methoxyphenyl ester	020665–85–4	2470	1671	1613
Propanoic acid, 2–methyl–, ethyl ester	000097–62–1	934	757	744
Propanoic acid, 2–methyl–, octyl ester	000109–15–9	1531	1344	1329
Propanoic acid, 2–methyl–, phenylmethyl ester	000103–28–6	1702	1298	1268
Pyrazine, 2,3–diethyl–5–methyl–	018138–04–0	1415	1155	1132
Pyrazine, 2,3–dimethyl–	005910–89–4	1240	918	893
Pyrazine, 2–ethyl–3,5–dimethyl–	013925–07–0	1370	1084	1060
Pyrazine, 2–ethyl–3–methyl–	015707–23–0	1307.5	1003	978
Pyrazine, 2–methoxy–3–(2–methylpropyl)–	024683–00–9	1414	1181	1160
Pyrazine, 2–methyl–3–(methylthio)–	002882–20–4	1521	1169	1139
Pyrazine, 2–methyl–6–(methylthio)–	002884–13–1	1583	1187	1155
Pyrazine, 3–ethyl–2,5–dimethyl–	013360–65–1	1370	1078	1055
Pyrazine, ethyl–	013925–00–3	1228	915	890
Pyrazine, tetramethyl–	001124–11–4	1381	1087	1063
Pyrazine, trimethyl–	014667–55–1	1293	1004	979
Tetradecanoic acid, ethyl ester	000124–06–1	2036	1792.3	1775
Thiazole, 5–ethenyl–4–methyl–	001759–28–0	1387	1026	999
trans–Isoeugenol	005932–68–3	2092	1451	1413
Triacetin	000102–76–1	1957	1352	1309
Triethyl citrate	000077–93–0	2369	1664	1618
Undecanal	000112–44–7	1575	1306	1285
α–Ionone	000127–41–3	1875	1431	1403
α–Irone	000079–69–6	1982	1519	1491
α–Phellandrene	000099–83–2	1128	1006	995
β–Myrcene	000123–35–3	1107	991	982
β–Phenylethyl butyrate	000103–52–6	1896	1443	1408
β–Pinene	000127–91–3	1092	978	968

δ–Nonalactone	003301–94–8	2148	1394	1343
2–Propenoic acid, ethyl ester	000140–88–5	911	702	649[a]
Acetoin	000513–86–0	1117	714	685[a]
Pentanal	000110–62–3	948	704	673[a]
Propanoic acid	000079–09–4	1138	706	648[a]
1–Propanol, 2–methyl–	000078–83–1	916	593[a]	564[a]
Ethyl Acetate	000141–78–6	819	568[a]	536[a]
Isopropyl acetate	000108–21–4	844	589[a]	603[a]

[a] extrapolated

Table 2S. List of mathematical equations used in the present study.

$$S = \sqrt{\frac{\sum_{i=1}^{N}\left(p_i^{exp} - p_i^{pred}\right)^2}{N-d-1}} \qquad RMSD = \sqrt{\frac{\sum_{i=1}^{N}\left(p_i^{exp} - p_i^{pred}\right)^2}{N}}$$

$$h_i = x_i(\mathbf{X}^T\mathbf{X})^{-1}x_i^T \qquad h^* = \frac{3(d+1)}{N_{train}}$$

$$b_j^s = \frac{S_j\, b_j}{S_{p_{exp}}} \qquad k = \frac{\sum_{i=1}^{N_{test}}\left(p_i^{exp}.p_i^{pred}\right)^2}{\sum_{i=1}^{N_{test}}\left(p_i^{pred}\right)^2}$$

$$k' = \frac{\sum_{i=1}^{N_{test}}\left(p_i^{exp}.p_i^{pred}\right)^2}{\sum_{i=1}^{N_{test}}\left(p_i^{\exp}\right)^2} \qquad p_{0i}^{\exp} = k.p^{pred}$$

$$p_{0i}^{pred} = k'.p^{exp} \qquad R_0^2 = 1 - \frac{\sum_{i=1}^{N_{test}}\left(p_i^{\exp} - p_{0i}^{\exp}\right)^2}{\sum_{i=1}^{N_{test}}\left(p_i^{\exp} - p_{av}^{\exp}\right)^2}$$

$$R_0'^2 = 1 - \frac{\sum_{i=1}^{N_{test}}\left(p_i^{pred} - p_{0i}^{pred}\right)^2}{\sum_{i=1}^{N_{test}}\left(p_i^{pred} - p_{av}^{pred}\right)^2} \qquad R_m^2 = R^2 \times \left(1 - \left|\sqrt{R^2 - R_0^2}\right|\right)$$

S: standard deviation; N: number of molecules; p_i^{exp}: experimental property for compound *i*; p_i^{pred}: predicted property for compound *i*; *d*: number of descriptors; $RMSD$: root mean square deviation; h_i: leverage for compound *i*; x_i: descriptor vector for *i*; $\mathbf{X}$: model matrix for the training set (train); N_{train}: number of molecules in train; N_{test}: number of molecules in test; h^*: warning leverage; b_j: regression coefficient for the *j*th descriptor; b_j^s: standardized b_j; S_j: standard deviation for the *j*th descriptor; $S_{p_{exp}}$: standard deviation for the experimental property; $p_{0i}^{\exp}$ and p_{0i}^{pred}: values of property in regressions through the origin of p_i^{exp} against p_i^{pred} and p_i^{pred} against p_i^{exp}; R_0^2 and $R_0'^2$: the correlation coefficients for regression of p_i^{exp} against p_i^{pred} and p_i^{pred} against p_i^{exp} through the origin; p_{av}^{pred}: average value for p^{pred} in test set; p_{av}^{exp}: average value for p^{exp} in test set; R_m^2: modified squared correlation coefficient; R^2: squared correlation coefficient between observed and predicted values for the test set with b intercept.

Table 3S. Chemical names, predicted retention indices, and numerical values for the four molecular descriptors appearing in the conformation–independent QSPR model (Eq. 1) for the DB–225MS stationary phase

No.	name	DB–225MS	SpAbs_B(e)	MATS1v	C–001	C–008
1^	1,2–Cyclopentanedione, 3–methyl–	1464.7	13.169	0	1	0
2^	1,6,10–Dodecatrien–3–ol, 3,7,11–trimethyl–	1877.2	24.85	–0.027	4	0
3*	1,6–Octadien–3–ol, 3,7–dimethyl–	1448.2	17.067	–0.042	3	0
4*	10–Undecenal	1819.5	18.257	–0.059	0	0
5^	1–Butanol, 3–methyl–	913.3	8.701	–0.103	2	0
6*	1–Butanol, 3–methyl–, acetate	1085.5	13.833	–0.144	3	0
7^	1–Butanol, 3–methyl–, formate	1045.3	12.127	–0.167	2	0
8^	1–Butanol, 3–methyl–, propanoate	1193.4	15.07	–0.126	3	0
9^	1–Cyclohexene–1–methanol, 4–(1–methylethenyl)–	1750.9	16.956	0.025	1	0
10*	1–Decanol	1551.4	15.597	–0.05	1	0
11^	1–Hexanol	1128.0	9.952	–0.085	1	0
12*	1–Hexanol, 2–ethyl–	1233.3	12.667	–0.063	2	0
13	1–Hexanol, 3,5,5–trimethyl–	1172.5	14.803	–0.056	4	0
14^	1–Octanol	1350.0	12.83	–0.063	1	0
15^	1–Octen–3–ol	1636.0	13.46	–0.057	1	1
16	2(3H)–Furanone, 5–butyldihydro–	1801.6	15.727	–0.046	1	1
17	2(3H)–Furanone, 5–ethyldihydro–	1574.9	12.6	–0.06	1	1
18*	2(3H)–Furanone, 5–heptyldihydro–	2101.9	20.016	–0.034	1	1
19	2(3H)–Furanone, 5–hexyldihydro–	2002.9	18.582	–0.037	1	1
20	2(3H)–Furanone, 5–methyl–	1282.5	12.029	–0.073	1	0
21*	2(3H)–Furanone, dihydro–5–methyl–	1483.0	11.473	–0.072	1	1
22*	2(3H)–Furanone, dihydro–5–propyl–	1698.5	14.297	–0.052	1	1
23	2(4H)–Benzofuranone, 5,6,7,7a–tetrahydro–4,4,7a–trimethyl–	1805.3	21.195	0.016	3	0
24	2(5H)–Furanone, 5–ethyl–3–hydroxy–4–methyl–	1793.3	16.431	–0.011	2	1
25*	2,4–Decadienal, (E,E)–	1633.7	17.112	–0.062	1	0
26^	2,4–Heptadienal, (E,E)–	1315.0	12.788	–0.085	1	0
27	2,6,6–Trimethyl–2–cyclohexene–1,4–dione	1588.5	18.262	0	3	0
28*	2,6–Octadien–1–ol, 3,7–dimethyl–, (Z)–	1435.5	16.875	–0.042	3	0
29^	2,6–Octadien–1–ol, 3,7–dimethyl–, formate, (E)–	1602.9	20.482	–0.09	3	0
30*	2–Acetyl–5–methylfuran	1449.2	15.72	–0.053	2	0
31^	2–Acetylthiazole	1514.9	14.153	–0.01	1	0
32	2–Butanone, 4–(4–hydroxyphenyl)–	1958.2	19.923	0.033	1	0
33	2–Buten–1–ol, 3–methyl–	960.9	9.113	–0.089	2	0
34^	2–Buten–1–one, 1–(2,6,6–trimethyl–1–cyclohexen–1–yl)–	1752.2	22.352	0	4	0
35	2–Buten–1–one, 1–(2,6,6–trimethyl–2–cyclohexen–1–yl)–, (E)–	1752.2	22.353	0	4	0
36*	2–Cyclopenten–1–one, 3–methyl–2–(2–pentenyl)–, (Z)–	1762.4	19.293	0	2	0
37^	2–Ethyl–3–methoxypyrazine	1447.8	15.915	–0.135	1	0
38	2–Furancarboxaldehyde, 5–methyl–	1425.5	13.953	–0.062	1	0
39	2–Furanmethanol, acetate	1583.2	17.28	–0.104	1	0
40*	2–Furfurylthiol	1331.3	11.65	–0.095	0	0
41*	2–Heptanone	1144.0	12.002	–0.094	2	0

42^	2–Hexen–1–ol, (E)–	1173.3	10.417	–0.075	1	0
43^	2–Hexen–1–ol, acetate, (E)–	1346.9	15.746	–0.124	2	0
44^	2–Hexenal, (E)–	1154.3	10.773	–0.104	1	0
45*	2H–Pyran, tetrahydro–4–methyl–2–(2–methyl–1–propenyl)–	1664.5	16.684	–0.038	3	1
46^	2H–Pyran–2–one, 6–hexyltetrahydro–	2075.9	19.622	–0.034	1	1
47*	2H–Pyran–2–one, tetrahydro–6–methyl–	1573.9	12.584	–0.06	1	1
48	2H–Pyran–2–one, tetrahydro–6–nonyl–	2370.8	23.94	–0.027	1	1
49	2H–Pyran–2–one, tetrahydro–6–pentyl–	1982.0	18.264	–0.037	1	1
50^	2H–Pyran–2–one, tetrahydro–6–propyl–	1773.1	15.295	–0.046	1	1
51^	2–Methylheptanoic acid	1416.0	15.216	–0.053	2	0
52^	2–Naphthyl methyl ketone	2129.4	22.453	0.036	1	0
53^	2–Nonanone	1372.3	15.02	–0.074	2	0
54^	2–Nonen–1–ol, (E)–	1491.6	14.711	–0.051	1	0
55*	2–Nonenal, (Z)–	1488.5	15.109	–0.071	1	0
56	2–Octanone	1263.3	13.567	–0.083	2	0
57*	2–Octenal, (E)–	1389.1	13.779	–0.079	1	0
58*	2–Octynoic acid, methyl ester	1572.6	17.208	–0.108	1	0
59^	2–Propanone, 1–(4–methoxyphenyl)–	1842.9	19.73	–0.037	1	0
60^	2–Propen–1–ol, 3–phenyl–	1839.8	16.385	0.039	0	0
61^	2–Propen–1–ol, 3–phenyl–, propanoate	1974.3	21.656	–0.034	1	0
62^	2–Propenal, 3–(2–methoxyphenyl)–	1957.8	19.887	–0.038	0	0
63	2–Tridecanone	1778.2	20.666	–0.051	2	0
64^	2–Undecanone	1580.3	17.864	–0.06	2	0
65*	3–(4–Isopropylphenyl)–2–methylpropionaldehyde	1857.7	22.346	0	3	0
66*	3–(Methylthio)propanoic acid methyl ester	1344.1	12.067	–0.105	0	0
67	3,4–Hexanedione	1169.5	12.5	–0.099	2	0
68	3–Hexen–1–ol	1170.6	10.375	–0.075	1	0
69	3–Hexen–1–ol, acetate, (Z)–	1339.7	15.638	–0.124	2	0
70^	3–Hexen–1–ol, formate, (Z)–	1311.8	13.962	–0.14	1	0
71*	3–Hexenoic acid, (E)–	1353.0	12.72	–0.056	1	0
72*	3–Phenylpropanol	1803.4	15.966	0.033	0	0
73	4–Heptenal, (Z)–	1278.2	12.341	–0.09	1	0
74^	4H–Pyran–4–one, 2–ethyl–3–hydroxy–	1683.5	16.533	–0.002	1	0
75	4–Methylthiazole	1269.6	10.477	–0.012	1	0
76	5,6,7,8–Tetrahydroquinoxaline	1743.1	16.097	–0.014	0	0
77^	5,9–Undecadien–2–one, 6,10–dimethyl–	1655.4	21.952	–0.048	4	0
78	5,9–Undecadien–2–one, 6,10–dimethyl–, (E)–	1655.4	21.952	–0.048	4	0
79	5H–5–Methyl–6,7–dihydrocyclopentapyrazine	1656.1	16.384	–0.014	1	0
80*	5–Heptenal, 2,6–dimethyl–	1299.6	15.458	–0.071	3	0
81*	5–Methyl–2–phenyl–2–hexenal	1987.6	22.709	0	2	0
82*	5–Thiazoleethanol, 4–methyl–	1632.4	14.98	0.033	1	0
83*	6–Octen–1–ol, 3,7–dimethyl–, acetate	1584.3	21.673	–0.084	4	0
84	6–Octen–1–ol, 3,7–dimethyl–, formate	1568.9	19.988	–0.091	3	0
85^	6–Octen–1–ol, 3,7–dimethyl–, propanoate	1669.2	22.827	–0.078	4	0
86	6–Octenal, 3,7–dimethyl–, (R)–	1411.3	16.997	–0.064	3	0

87	9,12–Octadecadienoic acid (Z,Z)–, methylester	2607.8	31.689	–0.053	1	0
88^	Acetic acid, 2–phenylethyl ester	1831.6	19.558	–0.037	1	0
89^	Acetic acid, 4–methylphenyl ester	1654.6	18.592	–0.042	2	0
90^	Acetic acid, decyl ester	1741.9	20.872	–0.085	2	0
91	Acetic acid, heptyl ester	1416.2	16.553	–0.113	2	0
92^	Acetic acid, nonyl ester	1640.5	19.512	–0.093	2	0
93	Acetic acid, octyl ester	1530.7	18.046	–0.102	2	0
94	Acetic acid, phenyl–, isopentyl ester	2032.0	24.027	–0.029	2	0
95	Acetophenone	1582.6	14.958	0	1	0
96*	Acetylpyrazine	1434.9	14.963	–0.101	1	0
97^	Allyl nonanoate	1880.6	21.346	–0.084	1	0
98^	Anisyl propionate	1969.5	22.317	–0.067	1	0
99*	Benzaldehyde	1582.4	13.348	0	0	0
100*	Benzaldehyde, 2–hydroxy–	1803.3	15.231	0.066	0	0
101^	Benzaldehyde, 4–(1–methylethyl)–	1674.2	17.955	0	2	0
102*	Benzaldehyde, 4–ethoxy–	1715.4	17.906	–0.042	1	0
103	Benzaldehyde, 4–ethyl–	1671.7	16.31	0	1	0
104*	Benzene, 1,1'–[oxybis(methylene)]bis–	2338.1	24.656	0.007	0	0
105^	Benzene, 1,2–dimethoxy–4–(1–propenyl)–	1875.1	20.907	–0.068	1	0
106^	Benzene, 1,3–dimethoxy–	1619.5	15.979	–0.093	0	0
107*	Benzene, 1–methoxy–4–methyl–	1480.2	14.449	–0.047	1	0
108*	Benzene, 2–methoxy–4–methyl–1–(1–methylethyl)–	1592.4	19.1	–0.035	3	0
109	Benzene, ethoxy–	1478.1	14.417	–0.047	1	0
110^	Benzeneacetaldehyde, α–ethylidene–	1801.3	18.276	0	1	0
111^	Benzeneacetic acid, 2–methylpropyl ester	1948.4	22.804	–0.031	2	0
112	Benzeneacetic acid, 2–phenylethyl ester	2676.1	29.807	0.006	0	0
113^	Benzeneacetic acid, methyl ester	1832.4	18.074	–0.042	0	0
114^	Benzeneacetic acid, phenylmethyl ester	2593.2	28.549	0.006	0	0
115^	Benzeneethanol, α,α–dimethyl–	1710.8	17.887	0.028	2	0
116^	Benzenemethanol, 4–methoxy–	1737.3	15.853	–0.007	0	0
117*	Benzenemethanol, 4–methoxy–, acetate	1887.2	21.224	–0.074	1	0
118^	Benzenemethanol, 4–methoxy–, formate	1873.0	19.58	–0.082	0	0
119^	Benzenemethanol,α–methyl–, acetate	1973.5	19.742	–0.037	2	1
120	Benzenepropanal	1789.1	16.484	0	0	0
121^	Benzoic acid	1799.1	15.167	0.066	0	0
122	Benzoic acid, 2–hydroxy–, 2–methylpropyl ester	2000.0	23.008	–0.005	2	0
123	Benzoic acid, 2–hydroxy–, ethyl ester	1901.8	19.823	–0.001	1	0
124	Benzoic acid, 2–hydroxy–, phenylmethyl ester	2650.1	28.745	0.036	0	0
125	Benzoic acid, ethyl ester	1716.9	17.929	–0.042	1	0
126^	Benzophenone	2311.4	23.695	0.032	0	0
127	Benzyl alcohol	1631.6	13.005	0.049	0	0
128*	Benzyl Benzoate	2475.3	26.782	0.005	0	0
129	Bicyclo[2.2.1]heptan–2–ol, 1,7,7–trimethyl–, acetate, (1S–endo)– Isobornyl acetate	1993.9	22.066	0.017	4	1
130*	Butanal, 3–methyl–	912.8	9.339	–0.132	2	0
131	Butanedioic acid, dimethyl ester	1427.4	15.644	–0.209	0	0

132*	Butanoic acid	1084.4	9.512	–0.095	1	0
133*	Butanoic acid, 2–methyl–, ethyl ester	1069.4	13.589	–0.144	3	0
134*	Butanoic acid, 2–methylpropyl ester	1211.6	15.346	–0.126	3	0
135*	Butanoic acid, 3,7–dimethyl–2,6–octadienyl ester, (E)–	1817.0	24.915	–0.071	4	0
136	Butanoic acid, 3–methyl–	1116.5	11.273	–0.08	2	0
137*	Butanoic acid, 3–methyl–, 2–phenylethyl ester	2036.0	24.089	–0.029	2	0
138	Butanoic acid, 3–methyl–, 3–methylbutyl ester	1337.1	18.323	–0.102	4	0
139*	Butanoic acid, 3–methyl–, 3–phenyl–2–propenyl ester	2175.8	26.164	–0.027	2	0
140*	Butanoic acid, 3–methyl–, butyl ester	1322.9	16.745	–0.113	3	0
141^	Butanoic acid, 3–methyl–, ethyl ester	1093.0	13.946	–0.144	3	0
142	Butanoic acid, 3–methyl–, phenylmethyl ester	1950.8	22.841	–0.031	2	0
143^	Butanoic acid, 3–methylbutyl ester	1310.7	16.56	–0.113	3	0
144	Butanoic acid, butyl ester	1295.3	15.009	–0.126	2	0
145	Butanoic acid, ethyl ester	1048.9	12.182	–0.167	2	0
146*	Butanoic acid, phenylmethyl ester	1936.9	21.089	–0.034	1	0
147	Butanoic acid, propyl ester	1182.7	13.7	–0.144	2	0
148	Caryophyllene	1979.9	23.311	0.04	3	0
149	Cedrol	2047.9	24.771	0.093	4	0
150*	Cinnamaldehyde, (E)–	1812.5	16.838	0	0	0
151*	Cinnamaldehyde, α–pentyl–	2179.8	24.018	0	1	0
152	Cinnamyl cinnamate	2923.9	33.61	0.004	0	0
153^	Citronellol	1401.4	16.447	–0.046	3	0
154	Citronellyl butyrate	1783.5	24.429	–0.072	4	0
155*	Creosol	1667.8	16.406	–0.007	1	0
156^	Cyclohexane, 1–ethenyl–1–methyl–2, 4–bis (1–methylethenyl)–,[1S–(1α,2β,4β)]–	1933.6	23.498	0	3	0
157*	Cyclohexanol, 5–methyl–2–(1–methylethyl)–, (1α,2α,5α)–	1729.4	16.38	0.02	3	1
158*	Cyclohexanone, 2–(1–mercapto–1–methylethyl)–5–methyl–	1592.6	18.636	–0.014	3	0
159	Cyclohexanone, 5–methyl–2–(1–methylethyl)–, (2R–cis)–	1501.1	16.936	0	3	0
160*	Cyclohexene, 1–methyl–4–(1–methylethylidene)–	1392.5	15.289	0	3	0
161	Cyclopentaneacetic acid, 3–oxo–2–pentyl–, methyl ester	2201.3	24.966	–0.028	1	0
162	D–Carvone	1672.3	17.926	0	2	0
163	Decanal	1566.6	16.182	–0.066	1	0
164	Decanedioic acid, diethyl ester	2166.7	27.561	–0.096	2	0
165*	Decanoic acid, ethyl ester	1736.0	20.783	–0.085	2	0
166	Diphenyl ether	2153.7	21.904	0.005	0	0
167^	D–Limonene	1508.7	15.444	0	2	0
168^	Dodecanal	1771.0	19.039	–0.055	1	0
169*	Dodecanoic acid, ethyl ester	1941.1	23.627	–0.073	2	0
170*	Ethane, 1,1–diethoxy–	808.3	11.584	–0.232	3	0
171^	Ethanone, 1–(2,4–dimethylphenyl)–	1601.0	18.451	0	3	0
172	Ethanone, 1–(2–pyridinyl)–	1516.8	14.961	–0.045	1	0
173^	Ethanone, 1–(3–pyridinyl)–	1516.9	14.962	–0.045	1	0
174*	Ethanone, 1–(4–methylphenyl)–	1592.0	16.708	0	2	0

175*	Ethyl Oleate	2574.7	32.75	–0.051	2	0
176*	Ethyl Vanillin	1895.5	19.727	–0.001	1	0
177	Eucalyptol	1513.9	16.619	0.023	3	0
178^	Eugenol	2004.7	19.866	–0.005	0	0
179*	Formic acid, pentyl ester	1145.2	12.035	–0.167	1	0
180^	Formic acid, phenylmethyl ester	1731.9	16.661	–0.047	0	0
181	Furan, 2–[(methylthio)methyl]–	1558.6	13.365	–0.017	0	0
182^	Furan, 2–methyl–5–(methylthio)–	1438.6	13.152	–0.017	1	0
183*	Heptanoic acid	1418.3	13.799	–0.06	1	0
184^	Heptanoic acid, 3–methylbutyl ester	1635.2	20.861	–0.085	3	0
185^	Heptanoic acid, ethyl ester	1411.3	16.479	–0.113	2	0
186*	Hexadecanoic acid, ethyl ester	2344.8	29.395	–0.057	2	0
187*	Hexanal	1121.3	10.406	–0.11	1	0
188^	Hexanoic acid, 2–methyl–	1310.1	13.766	–0.06	2	0
189	Hexanoic acid, 2–propenyl ester	1555.0	17.008	–0.111	1	0
190^	Hexanoic acid, ethyl ester	1293.9	14.987	–0.126	2	0
191*	Hexanoic acid, pentyl ester	1625.3	19.281	–0.093	2	0
192	Humulene	1830.2	23.536	0	4	0
193*	Hydrocoumarin	1942.5	18.632	0.008	0	0
194^	Isoamyl cinnamate	2154.9	25.847	–0.027	2	0
195^	Isobutyl acetate	969.1	12.578	–0.167	3	0
196^	Isopentyl hexanoate	1529.5	19.435	–0.093	3	0
197*	Isophorone	1464.4	16.38	0	3	0
198	Isopropyl myristate	2398.4	28.33	–0.061	3	1
199*	Menthyl isovalerate	2088.5	26.042	–0.025	5	1
200	Methional	1215.9	9.1	–0.059	0	0
201*	Methyl isovalerate	1060.5	12.358	–0.167	2	0
202*	Methyl salicylate	1916.7	18.353	0.003	0	0
203*	Methyleugenol	1986.4	20.989	–0.068	0	0
204*	Naphthalene, 2–ethoxy–	2048.2	21.822	0.009	1	0
205	Naphthalene, 2–methoxy–	2046.5	20.232	0.007	0	0
206^	n–Decanoic acid	1726.6	18.098	–0.043	1	0
207	Nonanal	1458.9	14.726	–0.074	1	0
208^	Nonanoic acid	1619.6	16.586	–0.048	1	0
209	Nonanoic acid, ethyl ester	1627.4	19.312	–0.093	2	0
210*	Nonanoic acid, methyl ester	1612.1	17.674	–0.102	1	0
211*	n–Propyl acetate	922.0	10.946	–0.198	2	0
212^	Octanal, 7–hydroxy–3,7–dimethyl–	1530.4	18.336	–0.043	3	0
213*	Octanoic acid	1524.8	15.259	–0.053	1	0
214^	Octanoic acid, 3–methylbutyl ester	1742.4	22.353	–0.079	3	0
215^	Octanoic acid, ethyl ester	1524.9	17.958	–0.102	2	0
216*	Octanoic acid, methyl ester	1508.1	16.341	–0.113	1	0
217	Oxacycloheptadec–8–en–2–one, (8Z)	2475.0	27.423	–0.024	0	0
218*	Oxacyclohexadecan–2–one	2339.3	25.386	–0.025	0	0
219	Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl ester Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl	2286.1	24.085	–0.019	2	1

	ester, cis–					
220^	p–Cymene	1437.5	15.971	0	3	0
221^	Pentanoic acid, 3–methyl–	1212.9	12.468	–0.068	2	0
222^	Pentanoic acid, butyl ester	1409.3	16.448	–0.113	2	0
223^	Pentanoic acid, ethyl ester	1181.1	13.676	–0.144	2	0
224	Pentyl octanoate	1841.7	22.252	–0.079	2	0
225^	Phenol, 2–(1–methylethyl)–	1606.2	16.19	0.033	2	0
226	Phenol, 2–methoxy–	1672.5	14.782	–0.003	0	0
227	Phenol, 2–methyl–5–(1–methylethyl)–	1602.5	17.851	0.028	3	0
228^	Phenylethyl Alcohol	1708.9	14.4	0.039	0	0
229*	Piperonal	1867.3	18.581	–0.041	0	0
230*	Propanedioic acid, diethyl ester	1382.7	17.579	–0.182	2	0
231*	Propanethioic acid, S–(2–furanylmethyl) ester	1776.4	18.187	–0.013	1	0
232	Propanoic acid, 2–methyl–	976.3	9.48	–0.095	2	0
233*	Propanoic acid, 2–methyl–, 2–phenylethyl ester	1924.3	22.439	–0.031	2	0
234^	Propanoic acid, 2–methyl–, 3–phenyl–2–propenyl ester	2056.2	24.395	–0.029	2	0
235*	Propanoic acid, 2–methyl–, 4–formyl–2–methoxyphenyl ester	2126.6	26.174	–0.061	2	0
236	Propanoic acid, 2–methyl–, ethyl ester	939.6	12.131	–0.167	3	0
237^	Propanoic acid, 2–methyl–, octyl ester	1628.8	20.763	–0.085	3	0
238^	Propanoic acid, 2–methyl–, phenylmethyl ester	1831.0	21.089	–0.034	2	0
239	Pyrazine, 2,3–diethyl–5–methyl–	1424.2	17.303	–0.069	3	0
240*	Pyrazine, 2,3–dimethyl–	1215.6	13.244	–0.101	2	0
241^	Pyrazine, 2–ethyl–3,5–dimethyl–	1329.0	16.037	–0.077	3	0
242^	Pyrazine, 2–ethyl–3–methyl–	1319.4	14.507	–0.087	2	0
243*	Pyrazine, 2–methoxy–3–(2–methylpropyl)–	1596.1	19.149	–0.107	2	0
244	Pyrazine, 2–methyl–3–(methylthio)–	1484.4	14.535	–0.048	1	0
245*	Pyrazine, 2–methyl–6–(methylthio)–	1485.9	14.559	–0.048	1	0
246*	Pyrazine, 3–ethyl–2,5–dimethyl–	1334.7	16.124	–0.077	3	0
247	Pyrazine, ethyl–	1291.5	12.788	–0.101	1	0
248	Pyrazine, tetramethyl–	1247.2	16.403	–0.077	4	0
249	Pyrazine, trimethyl–	1237.0	14.864	–0.087	3	0
250^	Tetradecanoic acid, ethyl ester	2147.1	26.552	–0.064	2	0
251	Thiazole, 5–ethenyl–4–methyl–	1498.4	13.88	–0.009	1	0
252	trans–Isoeugenol	1891.2	19.751	–0.005	1	0
253	Triacetin	1999.2	24.831	–0.176	3	1
254	Triethyl citrate	2219.5	30.325	–0.112	3	0
255*	Undecanal	1672.8	17.661	–0.06	1	0
256^	α–Ionone	1753.5	22.372	0	4	0
257*	α–Irone	1754.8	23.999	0	5	0
258	α–Phellandrene	1404.1	15.465	0	3	0
259^	β–Myrcene	1435.4	15.755	–0.064	2	0
260^	β–Phenylethyl butyrate	2025.2	22.362	–0.031	1	0
261	β–Pinene	1598.0	15.487	0.059	2	0
262*	δ–Nonalactone	1882.5	16.844	–0.041	1	1
263	2–Propenoic acid, ethyl ester	1043.8	11.053	–0.192	1	0

264*	Acetoin	1210.4	9.455	–0.095	2	1
265*	Pentanal	1002.3	9.089	–0.132	1	0
266	Propanoic acid	939.8	7.852	–0.119	1	0
267*	1–Propanol, 2–methyl–	774.7	7.222	–0.131	2	0
268	Ethyl Acetate	756.2	9.454	–0.244	2	0
269*	Isopropyl acetate	1071.2	11.241	–0.198	3	1

* validation set. ^ test set

Table 4S. Chemical names, predicted retention indices, and numerical values for the three molecular descriptors appearing in the conformation–independent QSPR model (Eq. 2) for the HP5–MS stationary phase

No.	name	HP5–MS	SpPos_B(p)	C–001	C–008
1*	1,2–Cyclopentanedione, 3–methyl–	954.4	9.744	1	0
2*	1,6,10–Dodecatrien–3–ol, 3,7,11–trimethyl–	1537.9	19.876	4	0
3^	1,6–Octadien–3–ol, 3,7–dimethyl–	1117.9	13.465	3	0
4^	10–Undecenal	1382.0	14.736	0	0
5^	1–Butanol, 3–methyl–	689.8	6.945	2	0
6	1–Butanol, 3–methyl–, acetate	903.3	10.578	3	0
7^	1–Butanol, 3–methyl–, formate	867.2	9.332	2	0
8^	1–Butanol, 3–methyl–, propanoate	981.3	11.628	3	0
9	1–Cyclohexene–1–methanol, 4–(1–methylethenyl)–	1234.6	13.513	1	0
10^	1–Decanol	1186.9	12.872	1	0
11*	1–Hexanol	821.4	7.955	1	0
12*	1–Hexanol, 2–ethyl–	949.6	10.44	2	0
13	1–Hexanol, 3,5,5–trimethyl–	952.6	12.002	4	0
14^	1–Octanol	1007.0	10.451	1	0
15^	1–Octen–3–ol	1131.4	10.663	1	1
16*	2(3H)–Furanone, 5–butyldihydro–	1221.8	11.879	1	1
17	2(3H)–Furanone, 5–ethyldihydro–	1032.3	9.33	1	1
18*	2(3H)–Furanone, 5–heptyldihydro–	1492.9	15.526	1	1
19	2(3H)–Furanone, 5–hexyldihydro–	1403.1	14.318	1	1
20	2(3H)–Furanone, 5–methyl–	862.4	8.507	1	0
21*	2(3H)–Furanone, dihydro–5–methyl–	951.7	8.246	1	1
22^	2(3H)–Furanone, dihydro–5–propyl–	1132.5	10.678	1	1
23	2(4H)–Benzofuranone, 5,6,7,7a–tetrahydro–4,4,7a–trimethyl–	1314.7	16.112	3	0
24	2(5H)–Furanone, 5–ethyl–3–hydroxy–4–methyl–	1148.5	11.654	2	1
25*	2,4–Decadienal, (E,E)–	1249.4	13.712	1	0
26*	2,4–Heptadienal, (E,E)–	977.4	10.054	1	0
27	2,6,6–Trimethyl–2–cyclohexene–1,4–dione	1145.5	13.836	3	0
28^	2,6–Octadien–1–ol, 3,7–dimethyl–, (Z)–	1118.6	13.475	3	0
29^	2,6–Octadien–1–ol, 3,7–dimethyl–, formate, (E)–	1298.4	15.894	3	0
30*	2–Acetyl–5–methylfuran	1031.5	11.542	2	0
31	2–Acetylthiazole	1036.4	10.847	1	0
32	2–Butanone, 4–(4–hydroxyphenyl)–	1338.5	14.911	1	0
33	2–Buten–1–ol, 3–methyl–	706.9	7.176	2	0
34^	2–Buten–1–one, 1–(2,6,6–trimethyl–1–cyclohexen–1–yl)–	1385.1	17.821	4	0
35	2–Buten–1–one, 1–(2,6,6–trimethyl–2–cyclohexen–1–yl)–, (E)–	1384.2	17.808	4	0
36*	2–Cyclopenten–1–one, 3–methyl–2–(2–pentenyl)–, (Z)–	1301.3	15.171	2	0
37*	2–Ethyl–3–methoxypyrazine	1131.5	12.127	1	0
38	2–Furancarboxaldehyde, 5–methyl–	994.7	10.286	1	0
39^	2–Furanmethanol, acetate	1145.0	12.308	1	0
40*	2–Furfurylthiol	994.0	9.516	0	0
41*	2–Heptanone	881.4	9.523	2	0

42*	2–Hexen–1–ol, (E)–	841.0	8.218	1	0
43^	2–Hexen–1–ol, acetate, (E)–	1055.4	11.863	2	0
44^	2–Hexenal, (E)–	869.0	8.595	1	0
45^	2H–Pyran, tetrahydro–4–methyl–2–(2–methyl–1–propenyl)–	1228.6	13.492	3	1
46^	2H–Pyran–2–one, 6–hexyltetrahydro–	1487.5	15.454	1	1
47	2H–Pyran–2–one, tetrahydro–6–methyl–	1038.9	9.418	1	1
48*	2H–Pyran–2–one, tetrahydro–6–nonyl–	1759.0	19.106	1	1
49*	2H–Pyran–2–one, tetrahydro–6–pentyl–	1394.9	14.208	1	1
50	2H–Pyran–2–one, tetrahydro–6–propyl–	1216.4	11.806	1	1
51^	2–Methylheptanoic acid	1035.5	11.596	2	0
52*	2–Naphthyl methyl ketone	1516.6	17.307	1	0
53*	2–Nonanone	1069.0	12.047	2	0
54^	2–Nonen–1–ol, (E)–	1112.9	11.876	1	0
55	2–Nonenal, (Z)–	1140.6	12.249	1	0
56	2–Octanone	978.9	10.835	2	0
57*	2–Octenal, (E)–	1049.5	11.023	1	0
58^	2–Octynoic acid, methyl ester	1217.8	13.287	1	0
59*	2–Propanone, 1–(4–methoxyphenyl)–	1343.5	14.979	1	0
60^	2–Propen–1–ol, 3–phenyl–	1229.6	12.685	0	0
61^	2–Propen–1–ol, 3–phenyl–, propanoate	1446.7	16.367	1	0
62*	2–Propenal, 3–(2–methoxyphenyl)–	1413.7	15.162	0	0
63^	2–Tridecanone	1427.2	16.865	2	0
64^	2–Undecanone	1250.0	14.481	2	0
65	3–(4–Isopropylphenyl)–2–methylpropionaldehyde	1439.1	17.786	3	0
66*	3–(Methylthio)propanoic acid methyl ester	1001.9	9.622	0	0
67	3,4–Hexanedione	863.4	9.28	2	0
68	3–Hexen–1–ol	838.4	8.183	1	0
69	3–Hexen–1–ol, acetate, (Z)–	1055.2	11.861	2	0
70^	3–Hexen–1–ol, formate, (Z)–	1019.7	10.622	1	0
71*	3–Hexenoic acid, (E)–	923.6	9.33	1	0
72^	3–Phenylpropanol	1215.7	12.499	0	0
73	4–Heptenal, (Z)–	958.8	9.803	1	0
74^	4H–Pyran–4–one, 2–ethyl–3–hydroxy–	1116.1	11.92	1	0
75	4–Methylthiazole	854.7	8.403	1	0
76	5,6,7,8–Tetrahydroquinoxaline	1221.1	12.571	0	0
77	5,9–Undecadien–2–one, 6,10–dimethyl–	1362.3	17.514	4	0
78^	5,9–Undecadien–2–one, 6,10–dimethyl–, (E)–	1362.3	17.514	4	0
79	5H–5–Methyl–6,7–dihydrocyclopentapyrazine	1182.8	12.816	1	0
80*	5–Heptenal, 2,6–dimethyl–	1036.9	12.376	3	0
81^	5–Methyl–2–phenyl–2–hexenal	1516.0	18.06	2	0
82	5–Thiazoleethanol, 4–methyl–	1101.1	11.718	1	0
83*	6–Octen–1–ol, 3,7–dimethyl–, acetate	1318.1	16.919	4	0
84	6–Octen–1–ol, 3,7–dimethyl–, formate	1282.2	15.676	3	0
85*	6–Octen–1–ol, 3,7–dimethyl–, propanoate	1395.9	17.966	4	0
86	6–Octenal, 3,7–dimethyl–, (R)–	1129.3	13.619	3	0

87	9,12–Octadecadienoic acid (Z,Z)–, methylester	2126.1	25.507	1	0
88^	Acetic acid, 2–phenylethyl ester	1340.2	14.934	1	0
89^	Acetic acid, 4–methylphenyl ester	1200.2	13.811	2	0
90*	Acetic acid, decyl ester	1401.3	16.517	2	0
91*	Acetic acid, heptyl ester	1129.7	12.863	2	0
92^	Acetic acid, nonyl ester	1310.7	15.298	2	0
93	Acetic acid, octyl ester	1222.2	14.108	2	0
94	Acetic acid, phenyl–, isopentyl ester	1565.1	18.721	2	0
95	Acetophenone	1090.7	11.578	1	0
96*	Acetylpyrazine	1064.0	11.218	1	0
97^	Allyl nonanoate	1470.4	16.685	1	0
98*	Anisyl propionate	1482.7	16.851	1	0
99*	Benzaldehyde	1054.2	10.326	0	0
100^	Benzaldehyde, 2–hydroxy–	1125.5	11.285	0	0
101	Benzaldehyde, 4–(1–methylethyl)–	1220.5	14.084	2	0
102*	Benzaldehyde, 4–ethoxy–	1248.3	13.698	1	0
103^	Benzaldehyde, 4–ethyl–	1183.0	12.819	1	0
104*	Benzene, 1,1'–[oxybis(methylene)]bis–	1731.5	19.437	0	0
105	Benzene, 1,2–dimethoxy–4–(1–propenyl)–	1432.7	16.178	1	0
106*	Benzene, 1,3–dimethoxy–	1188.3	12.13	0	0
107^	Benzene, 1–methoxy–4–methyl–	1071.6	11.321	1	0
108	Benzene, 2–methoxy–4–methyl–1–(1–methylethyl)–	1241.8	15.132	3	0
109	Benzene, ethoxy–	1067.8	11.269	1	0
110^	Benzeneacetaldehyde, α–ethylidene–	1295.4	14.331	1	0
111^	Benzeneacetic acid, 2–methylpropyl ester	1467.3	17.405	2	0
112*	Benzeneacetic acid, 2–phenylethyl ester	2002.1	23.077	0	0
113^	Benzeneacetic acid, methyl ester	1305.3	13.704	0	0
114	Benzeneacetic acid, phenylmethyl ester	1907.6	21.806	0	0
115*	Benzeneethanol, α,α–dimethyl–	1204.7	13.872	2	0
116	Benzenemethanol, 4–methoxy–	1189.1	12.141	0	0
117^	Benzenemethanol, 4–methoxy–, acetate	1405.2	15.809	1	0
118*	Benzenemethanol, 4–methoxy–, formate	1369.7	14.57	0	0
119^	Benzenemethanol,α–methyl–, acetate	1389.6	14.897	2	1
120	Benzenepropanal	1241.1	12.84	0	0
121^	Benzoic acid	1112.9	11.115	0	0
122*	Benzoic acid, 2–hydroxy–, 2–methylpropyl ester	1440.5	17.044	2	0
123	Benzoic acid, 2–hydroxy–, ethyl ester	1314.4	14.587	1	0
124*	Benzoic acid, 2–hydroxy–, phenylmethyl ester	1880.1	21.436	0	0
125	Benzoic acid, ethyl ester	1243.2	13.629	1	0
126^	Benzophenone	1652.8	18.378	0	0
127^	Benzyl alcohol	1026.0	9.947	0	0
128^	Benzyl Benzoate	1809.3	20.484	0	0
129	Bicyclo[2.2.1]heptan–2–ol, 1,7,7–trimethyl–, acetate, (1S–endo)– Isobornyl acetate	1441.6	17.118	4	1
130*	Butanal, 3–methyl–	717.2	7.314	2	0
131	Butanedioic acid, dimethyl ester	1116.2	11.16	0	0

132*	Butanoic acid	730.4	6.731	1	0
133*	Butanoic acid, 2–methyl–, ethyl ester	894.4	10.459	3	0
134	Butanoic acid, 2–methylpropyl ester	985.7	11.687	3	0
135^	Butanoic acid, 3,7–dimethyl–2,6–octadienyl ester, (E)–	1514.3	19.559	4	0
136	Butanoic acid, 3–methyl–	775.3	8.096	2	0
137^	Butanoic acid, 3–methyl–, 2–phenylethyl ester	1565.1	18.721	2	0
138*	Butanoic acid, 3–methyl–, 3–methylbutyl ester	1128.2	14.365	4	0
139*	Butanoic acid, 3–methyl–, 3–phenyl–2–propenyl ester	1672.0	20.158	2	0
140	Butanoic acid, 3–methyl–, butyl ester	1083.0	12.995	3	0
141^	Butanoic acid, 3–methyl–, ethyl ester	906.0	10.614	3	0
142	Butanoic acid, 3–methyl–, phenylmethyl ester	1470.6	17.449	2	0
143^	Butanoic acid, 3–methylbutyl ester	1083.2	12.999	3	0
144*	Butanoic acid, butyl ester	1037.9	11.628	2	0
145	Butanoic acid, ethyl ester	860.9	9.247	2	0
146*	Butanoic acid, phenylmethyl ester	1425.8	16.086	1	0
147^	Butanoic acid, propyl ester	947.3	10.409	2	0
148	Caryophyllene	1540.9	19.155	3	0
149	Cedrol	1533.7	19.819	4	0
150*	Cinnamaldehyde, (E)–	1254.2	13.016	0	0
151*	Cinnamaldehyde, α–pentyl–	1655.9	19.181	1	0
152	Cinnamyl cinnamate	2208.6	25.856	0	0
153^	Citronellol	1102.9	13.264	3	0
154	Citronellyl butyrate	1498.2	19.342	4	0
155^	Creosol	1138.5	12.221	1	0
156*	Cyclohexane, 1–ethenyl–1–methyl–2, 4–bis (1–methylethenyl)–,[1S–(1α,2β,4β)]–	1547.8	19.249	3	0
157*	Cyclohexanol, 5–methyl–2–(1–methylethyl)–, (1α,2α,5α)–	1208.2	13.218	3	1
158*	Cyclohexanone, 2–(1–mercapto–1–methylethyl)–5–methyl–	1265.4	15.45	3	0
159	Cyclohexanone, 5–methyl–2–(1–methylethyl)–, (2R–cis)–	1124.9	13.559	3	0
160*	Cyclohexene, 1–methyl–4–(1–methylethylidene)–	1056.9	12.644	3	0
161*	Cyclopentaneacetic acid, 3–oxo–2–pentyl–, methyl ester	1656.6	19.19	1	0
162*	D–Carvone	1218.9	14.063	2	0
163*	Decanal	1213.7	13.232	1	0
164	Decanedioic acid, diethyl ester	1734.7	21.002	2	0
165	Decanoic acid, ethyl ester	1402.4	16.531	2	0
166	Diphenyl ether	1548.9	16.981	0	0
167^	D–Limonene	1119.2	12.722	2	0
168	Dodecanal	1391.1	15.619	1	0
169^	Dodecanoic acid, ethyl ester	1579.7	18.917	2	0
170*	Ethane, 1,1–diethoxy–	780.3	8.924	3	0
171^	Ethanone, 1–(2,4–dimethylphenyl)–	1180.3	14.305	3	0
172	Ethanone, 1–(2–pyridinyl)–	1077.5	11.4	1	0
173^	Ethanone, 1–(3–pyridinyl)–	1077.6	11.401	1	0
174*	Ethanone, 1–(4–methylphenyl)–	1135.6	12.942	2	0

175*	Ethyl Oleate	2143.4	26.5	2	0
176^	Ethyl Vanillin	1318.0	14.636	1	0
177	Eucalyptol	1108.7	13.342	3	0
178	Eugenol	1399.1	14.966	0	0
179*	Formic acid, pentyl ester	915.4	9.219	1	0
180^	Formic acid, phenylmethyl ester	1210.2	12.424	0	0
181	Furan, 2–[(methylthio)methyl]–	1081.0	10.687	0	0
182*	Furan, 2–methyl–5–(methylthio)–	1024.0	10.68	1	0
183^	Heptanoic acid	1000.6	10.365	1	0
184*	Heptanoic acid, 3–methylbutyl ester	1353.6	16.636	3	0
185^	Heptanoic acid, ethyl ester	1131.3	12.884	2	0
186^	Hexadecanoic acid, ethyl ester	1945.1	23.832	2	0
187^	Hexanal	845.6	8.28	1	0
188^	Hexanoic acid, 2–methyl–	946.6	10.4	2	0
189*	Hexanoic acid, 2–propenyl ester	1197.7	13.017	1	0
190	Hexanoic acid, ethyl ester	1034.1	11.577	2	0
191^	Hexanoic acid, pentyl ester	1304.6	15.216	2	0
192	Humulene	1492.7	19.268	4	0
193*	Hydrocoumarin	1318.3	13.879	0	0
194^	Isoamyl cinnamate	1665.5	20.071	2	0
195^	Isobutyl acetate	804.6	9.25	3	0
196^	Isopentyl hexanoate	1257.0	15.337	3	0
197*	Isophorone	1066.2	12.77	3	0
198	Isopropyl myristate	1914.5	22.719	3	1
199	Menthyl isovalerate	1648.2	20.658	5	1
200*	Methional	850.5	7.586	0	0
201*	Methyl isovalerate	868.4	9.348	2	0
202*	Methyl salicylate	1277.0	13.323	0	0
203*	Methyleugenol	1492.9	16.227	0	0
204^	Naphthalene, 2–ethoxy–	1491.5	16.97	1	0
205	Naphthalene, 2–methoxy–	1454.9	15.716	0	0
206^	n–Decanoic acid	1271.6	14.011	1	0
207	Nonanal	1118.7	11.954	1	0
208^	Nonanoic acid	1176.5	12.732	1	0
209*	Nonanoic acid, ethyl ester	1307.3	15.252	2	0
210^	Nonanoic acid, methyl ester	1270.8	14.001	1	0
211	n–Propyl acetate	766.8	7.981	2	0
212^	Octanal, 7–hydroxy–3,7–dimethyl–	1177.1	14.262	3	0
213^	Octanoic acid	1090.5	11.575	1	0
214^	Octanoic acid, 3–methylbutyl ester	1443.9	17.851	3	0
215^	Octanoic acid, ethyl ester	1221.5	14.098	2	0
216	Octanoic acid, methyl ester	1184.3	12.837	1	0
217	Oxacycloheptadec–8–en–2–one, (8Z)	1906.2	21.787	0	0
218*	Oxacyclohexadecan–2–one	1794.1	20.28	0	0
219	Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl ester Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl	1635.6	18.207	2	1

	ester, cis–				
220^	p–Cymene	1082.2	12.985	3	0
221^	Pentanoic acid, 3–methyl–	855.9	9.18	2	0
222^	Pentanoic acid, butyl ester	1127.7	12.836	2	0
223	Pentanoic acid, ethyl ester	950.6	10.454	2	0
224*	Pentyl octanoate	1491.9	17.735	2	0
225^	Phenol, 2–(1–methylethyl)–	1109.3	12.588	2	0
226*	Phenol, 2–methoxy–	1100.3	10.946	0	0
227*	Phenol, 2–methyl–5–(1–methylethyl)–	1149.7	13.893	3	0
228*	Phenylethyl Alcohol	1126.7	11.301	0	0
229*	Piperonal	1303.4	13.678	0	0
230*	Propanedioic acid, diethyl ester	1107.2	12.56	2	0
231	Propanethioic acid, S–(2–furanylmethyl) ester	1289.7	14.255	1	0
232	Propanoic acid, 2–methyl–	676.9	6.772	2	0
233*	Propanoic acid, 2–methyl–, 2–phenylethyl ester	1466.4	17.393	2	0
234^	Propanoic acid, 2–methyl–, 3–phenyl–2–propenyl ester	1573.7	18.836	2	0
235*	Propanoic acid, 2–methyl–, 4–formyl–2–methoxyphenyl ester	1623.9	19.511	2	0
236	Propanoic acid, 2–methyl–, ethyl ester	807.2	9.285	3	0
237^	Propanoic acid, 2–methyl–, octyl ester	1348.8	16.571	3	0
238^	Propanoic acid, 2–methyl–, phenylmethyl ester	1373.0	16.136	2	0
239^	Pyrazine, 2,3–diethyl–5–methyl–	1134.6	13.69	3	0
240	Pyrazine, 2,3–dimethyl–	929.4	10.169	2	0
241	Pyrazine, 2–ethyl–3,5–dimethyl–	1049.0	12.538	3	0
242	Pyrazine, 2–ethyl–3–methyl–	1009.9	11.251	2	0
243*	Pyrazine, 2–methoxy–3–(2–methylpropyl)–	1266.0	14.697	2	0
244^	Pyrazine, 2–methyl–3–(methylthio)–	1105.5	11.777	1	0
245*	Pyrazine, 2–methyl–6–(methylthio)–	1108.5	11.817	1	0
246^	Pyrazine, 3–ethyl–2,5–dimethyl–	1048.3	12.529	3	0
247*	Pyrazine, ethyl–	972.6	9.989	1	0
248	Pyrazine, tetramethyl–	1003.9	12.692	4	0
249	Pyrazine, trimethyl–	968.3	11.452	3	0
250*	Tetradecanoic acid, ethyl ester	1764.1	21.398	2	0
251*	Thiazole, 5–ethenyl–4–methyl–	1055.9	11.109	1	0
252*	trans–Isoeugenol	1339.4	14.923	1	0
253	Triacetin	1489.6	17.003	3	1
254	Triethyl citrate	1725.4	21.638	3	0
255^	Undecanal	1304.1	14.449	1	0
256*	α–Ionone	1379.3	17.743	4	0
257^	α–Irone	1422.2	19.081	5	0
258	α–Phellandrene	1063.5	12.733	3	0
259^	β–Myrcene	1130.7	12.877	2	0
260^	β–Phenylethyl butyrate	1520.2	17.355	1	0
261	β–Pinene	1120.8	12.744	2	0
262*	δ–Nonalactone	1308.5	13.045	1	1
263*	2–Propenoic acid, ethyl ester	843.0	8.246	1	0

264*	Acetoin	791.9	6.857	2	1
265	Pentanal	762.0	7.156	1	0
266	Propanoic acid	626.9	5.338	1	0
267	1–Propanol, 2–methyl–	589.9[a]	5.601	2	0
268	Ethyl Acetate	681.0[a]	6.827	2	0
269	Isopropyl acetate	831.5[a]	8.15	3	1

* validation set. ^ test set. [a] predicted retention indices using Eq. 2

Table 5S. Chemical names, predicted retention indices, and numerical values for the three molecular descriptors appearing in the conformation–independent QSPR model (Eq. 3) for the HP–1 stationary phase

No.	name	HP–1	SpPos_A	MATS1p	F02[C–C]
1^	1,2–Cyclopentanedione, 3–methyl–	982.5	4.899	0	7
2^	1,6,10–Dodecatrien–3–ol, 3,7,11–trimethyl–	1506.8	8.902	–0.01	16
3*	1,6–Octadien–3–ol, 3,7–dimethyl–	1099.7	5.914	–0.019	10
4*	10–Undecenal	1298.7	7.296	–0.061	9
5*	1–Butanol, 3–methyl–	688.7	3.078	–0.078	4
6^	1–Butanol, 3–methyl–, acetate	847.6	4.732	–0.236	5
7^	1–Butanol, 3–methyl–, formate	770.8	4.381	–0.278	5
8^	1–Butanol, 3–methyl–, propanoate	986.6	5.589	–0.205	6
9	1–Cyclohexene–1–methanol, 4–(1–methylethenyl)–	1248.0	6.81	0.053	13
10*	1–Decanol	1216.9	6.596	–0.031	8
11*	1–Hexanol	843.8	4.027	–0.061	4
12*	1–Hexanol, 2–ethyl–	1018.7	5.283	–0.042	7
13	1–Hexanol, 3,5,5–trimethyl–	945.5	5.033	–0.036	11
14^	1–Octanol	1032.9	5.314	–0.042	6
15^	1–Octen–3–ol	1025.8	5.236	–0.033	6
16	2(3H)–Furanone, 5–butyldihydro–	1124.1	6.213	–0.111	7
17	2(3H)–Furanone, 5–ethyldihydro–	929.1	4.955	–0.149	5
18*	2(3H)–Furanone, 5–heptyldihydro–	1393.5	8.045	–0.081	10
19	2(3H)–Furanone, 5–hexyldihydro–	1311.6	7.479	–0.089	9
20	2(3H)–Furanone, 5–methyl–	799.4	4.159	–0.18	4
21*	2(3H)–Furanone, dihydro–5–methyl–	800.0	4.159	–0.179	4
22^	2(3H)–Furanone, dihydro–5–propyl–	1010.7	5.47	–0.127	6
23	2(4H)–Benzofuranone, 5,6,7,7a–tetrahydro–4,4,7a–trimethyl–	1299.0	7.661	–0.024	17
24	2(5H)–Furanone, 5–ethyl–3–hydroxy–4–methyl–	1142.3	6.13	–0.054	7
25*	2,4–Decadienal, (E,E)–	1198.5	6.596	–0.065	8
26^	2,4–Heptadienal, (E,E)–	930.0	4.759	–0.092	5
27*	2,6,6–Trimethyl–2–cyclohexene–1,4–dione	1115.6	6.136	0	13
28^	2,6–Octadien–1–ol, 3,7–dimethyl–, (Z)–	1123.2	6.067	–0.019	10
29^	2,6–Octadien–1–ol, 3,7–dimethyl–, formate, (E)–	1245.5	7.367	–0.144	11
30	2–Acetyl–5–methylfuran	989.3	5.334	–0.128	6
31*	2–Acetylthiazole	1031.0	4.837	0.037	3
32	2–Butanone, 4–(4–hydroxyphenyl)–	1338.3	7.228	0.066	11
33	2–Buten–1–ol, 3–methyl–	700.1	3.078	–0.057	4
34^	2–Buten–1–one, 1–(2,6,6–trimethyl–1–cyclohexen–1–yl)–	1351.7	7.981	0	18
35	2–Buten–1–one, 1–(2,6,6–trimethyl–2–cyclohexen–1–yl)–, (E)–	1351.7	7.981	0	18
36	2–Cyclopenten–1–one, 3–methyl–2–(2–pentenyl)–, (Z)–	1315.4	7.436	0	13
37*	2–Ethyl–3–methoxypyrazine	1073.5	6.432	–0.301	5
38*	2–Furancarboxaldehyde, 5–methyl–	928.6	4.955	–0.15	5
39*	2–Furanmethanol, acetate	1043.7	6.086	–0.258	5
40	2–Furfurylthiol	842.8	4.459	–0.185	4
41*	2–Heptanone	866.4	4.381	–0.102	5

42^	2–Hexen–1–ol, (E)–	852.0	4.027	–0.046	4
43^	2–Hexen–1–ol, acetate, (E)–	1009.8	5.671	–0.203	5
44*	2–Hexenal, (E)–	815.0	4.027	–0.114	4
45^	2H–Pyran, tetrahydro–4–methyl–2–(2–methyl–1–propenyl)–	1125.8	6.39	–0.088	11
46^	2H–Pyran–2–one, 6–hexyltetrahydro–	1400.1	8.088	–0.081	10
47*	2H–Pyran–2–one, tetrahydro–6–methyl–	892.4	4.716	–0.149	5
48	2H–Pyran–2–one, tetrahydro–6–nonyl–	1663.1	9.921	–0.063	13
49	2H–Pyran–2–one, tetrahydro–6–pentyl–	1291.6	7.349	–0.089	9
50^	2H–Pyran–2–one, tetrahydro–6–propyl–	1098.9	6.049	–0.111	7
51^	2–Methylheptanoic acid	1073.7	5.62	–0.036	7
52^	2–Naphthyl methyl ketone	1452.1	8.438	0.038	17
53^	2–Nonanone	1058.7	5.671	–0.078	7
54*	2–Nonen–1–ol, (E)–	1140.1	6.027	–0.029	7
55	2–Nonenal, (Z)–	1115.1	6.027	–0.075	7
56	2–Octanone	979.0	5.126	–0.088	6
57	2–Octenal, (E)–	1009.6	5.314	–0.085	6
58^	2–Octynoic acid, methyl ester	1135.9	6.523	–0.177	7
59^	2–Propanone, 1–(4–methoxyphenyl)–	1259.4	7.263	–0.089	11
60*	2–Propen–1–ol, 3–phenyl–	1239.9	6.466	0.083	10
61*	2–Propen–1–ol, 3–phenyl–, propanoate	1392.0	8.097	–0.081	11
62^	2–Propenal, 3–(2–methoxyphenyl)–	1324.9	7.689	–0.089	11
63	2–Tridecanone	1428.5	8.236	–0.053	11
64^	2–Undecanone	1245.2	6.955	–0.063	9
65	3–(4–Isopropylphenyl)–2–methylpropionaldehyde	1441.2	8.502	0	17
66*	3–(Methylthio)propanoic acid methyl ester	967.3	4.705	–0.043	3
67*	3,4–Hexanedione	911.6	4.666	–0.117	4
68	3–Hexen–1–ol	852.0	4.027	–0.046	4
69	3–Hexen–1–ol, acetate, (Z)–	1009.8	5.671	–0.203	5
70^	3–Hexen–1–ol, formate, (Z)–	938.6	5.314	–0.233	5
71*	3–Hexenoic acid, (E)–	913.4	4.381	–0.033	4
72^	3–Phenylpropanol	1232.9	6.466	0.07	10
73	4–Heptenal, (Z)–	927.2	4.759	–0.097	5
74*	4H–Pyran–4–one, 2–ethyl–3–hydroxy–	1171.6	6.209	–0.04	6
75	4–Methylthiazole	863.0	3.733	0.04	3
76	5,6,7,8–Tetrahydroquinoxaline	1229.2	6.842	–0.078	8
77^	5,9–Undecadien–2–one, 6,10–dimethyl–	1331.7	7.719	–0.05	13
78*	5,9–Undecadien–2–one, 6,10–dimethyl–, (E)–	1331.7	7.719	–0.05	13
79	5H–5–Methyl–6,7–dihydrocyclopentapyrazine	1192.8	6.667	–0.078	9
80^	5–Heptenal, 2,6–dimethyl–	1028.8	5.589	–0.075	9
81^	5–Methyl–2–phenyl–2–hexenal	1467.5	8.611	0	16
82*	5–Thiazoleethanol, 4–methyl–	1151.5	5.511	0.103	5
83	6–Octen–1–ol, 3,7–dimethyl–, acetate	1305.6	7.719	–0.133	11
84*	6–Octen–1–ol, 3,7–dimethyl–, formate	1245.0	7.367	–0.145	11
85^	6–Octen–1–ol, 3,7–dimethyl–, propanoate	1433.6	8.575	–0.122	12
86*	6–Octenal, 3,7–dimethyl–, (R)–	1096.5	6.067	–0.068	10

87	9,12–Octadecadienoic acid (Z,Z)–, methylester	2075.0	12.916	–0.082	17
88^	Acetic acid, 2–phenylethyl ester	1281.4	7.344	–0.089	10
89^	Acetic acid, 4–methylphenyl ester	1172.8	6.611	–0.099	9
90^	Acetic acid, decyl ester	1403.5	8.236	–0.134	9
91*	Acetic acid, heptyl ester	1123.1	6.392	–0.181	6
92^	Acetic acid, nonyl ester	1317.6	7.661	–0.147	8
93	Acetic acid, octyl ester	1210.4	6.955	–0.162	7
94	Acetic acid, phenyl–, isopentyl ester	1513.0	9.024	–0.068	14
95*	Acetophenone	1070.5	5.595	0	9
96^	Acetylpyrazine	1004.9	5.595	–0.208	4
97^	Allyl nonanoate	1443.1	8.49	–0.133	9
98*	Anisyl propionate	1470.0	8.895	–0.163	11
99^	Benzaldehyde	1021.1	5.212	0	8
100*	Benzaldehyde, 2–hydroxy–	1146.2	5.602	0.12	8
101	Benzaldehyde, 4–(1–methylethyl)–	1219.2	6.81	0	13
102*	Benzaldehyde, 4–ethoxy–	1220.3	6.92	–0.099	9
103^	Benzaldehyde, 4–ethyl–	1179.6	6.429	0	11
104*	Benzene, 1,1'–[oxybis(methylene)]bis–	1650.4	9.955	–0.026	17
105	Benzene, 1,2–dimethoxy–4–(1–propenyl)–	1344.7	8.138	–0.162	12
106*	Benzene, 1,3–dimethoxy–	1085.6	6.424	–0.224	8
107^	Benzene, 1–methoxy–4–methyl–	1010.1	5.595	–0.111	9
108*	Benzene, 2–methoxy–4–methyl–1–(1–methylethyl)–	1223.7	7.184	–0.08	14
109	Benzene, ethoxy–	1044.0	5.692	–0.111	7
110^	Benzeneacetaldehyde, α–ethylidene–	1253.0	6.968	0	12
111*	Benzeneacetic acid, 2–methylpropyl ester	1442.1	8.522	–0.074	13
112	Benzeneacetic acid, 2–phenylethyl ester	1891.9	11.636	–0.022	19
113^	Benzeneacetic acid, methyl ester	1207.4	6.898	–0.099	10
114*	Benzeneacetic acid, phenylmethyl ester	1821.7	11.124	–0.024	18
115*	Benzeneethanol, α,α–dimethyl–	1185.1	6.314	0.06	12
116*	Benzenemethanol, 4–methoxy–	1180.7	6.429	–0.033	9
117^	Benzenemethanol, 4–methoxy–, acetate	1341.7	8.055	–0.179	10
118^	Benzenemethanol, 4–methoxy–, formate	1273.3	7.684	–0.2	10
119*	Benzenemethanol,α–methyl–, acetate	1268.9	7.263	–0.089	10
120	Benzenepropanal	1194.8	6.466	0	10
121^	Benzoic acid	1145.2	5.595	0.12	8
122^	Benzoic acid, 2–hydroxy–, 2–methylpropyl ester	1465.2	8.441	–0.026	12
123^	Benzoic acid, 2–hydroxy–, ethyl ester	1318.5	7.301	–0.026	9
124	Benzoic acid, 2–hydroxy–, phenylmethyl ester	1850.2	11.043	0.034	17
125	Benzoic acid, ethyl ester	1220.3	6.92	–0.099	9
126^	Benzophenone	1598.8	9.407	0.034	17
127	Benzyl alcohol	1076.5	5.212	0.102	8
128^	Benzyl Benzoate	1757.7	10.657	–0.027	17
129	Bicyclo[2.2.1]heptan–2–ol, 1,7,7–trimethyl–, acetate, (1S–endo)– Isobornyl acetate	1383.0	8.379	–0.02	20
130^	Butanal, 3–methyl–	651.8	3.078	–0.146	4
131	Butanedioic acid, dimethyl ester	929.2	5.916	–0.438	4

132*	Butanoic acid	703.4	3.078	–0.086	2
133^	Butanoic acid, 2–methyl–, ethyl ester	912.7	5.156	–0.236	5
134*	Butanoic acid, 2–methylpropyl ester	982.8	5.564	–0.205	6
135*	Butanoic acid, 3,7–dimethyl–2,6–octadienyl ester, (E)–	1503.0	9.053	–0.112	13
136	Butanoic acid, 3–methyl–	748.0	3.414	–0.064	4
137^	Butanoic acid, 3–methyl–, 2–phenylethyl ester	1512.2	9.019	–0.068	14
138*	Butanoic acid, 3–methyl–, 3–methylbutyl ester	1107.2	6.407	–0.162	9
139^	Butanoic acid, 3–methyl–, 3–phenyl–2–propenyl ester	1620.6	9.768	–0.063	15
140^	Butanoic acid, 3–methyl–, butyl ester	1061.9	6.056	–0.181	7
141*	Butanoic acid, 3–methyl–, ethyl ester	850.9	4.754	–0.236	5
142	Butanoic acid, 3–methyl–, phenylmethyl ester	1439.9	8.508	–0.074	13
143	Butanoic acid, 3–methylbutyl ester	1063.6	6.067	–0.181	7
144	Butanoic acid, butyl ester	1015.5	5.715	–0.205	5
145	Butanoic acid, ethyl ester	794.9	4.414	–0.278	3
146	Butanoic acid, phenylmethyl ester	1402.4	8.165	–0.081	11
147	Butanoic acid, propyl ester	927.7	5.192	–0.236	4
148	Caryophyllene	1485.1	8.955	0.04	22
149	Cedrol	1593.0	9.831	0.113	29
150^	Cinnamaldehyde, (E)–	1194.8	6.466	0	10
151*	Cinnamaldehyde, α–pentyl–	1604.0	9.499	0	16
152	Cinnamyl cinnamate	2108.8	13.167	–0.021	21
153^	Citronellol	1119.4	6.067	–0.026	10
154	Citronellyl butyrate	1502.5	9.053	–0.113	13
155^	Creosol	1143.7	6.188	–0.033	9
156^	Cyclohexane, 1–ethenyl–1–methyl–2, 4–bis (1–methylethenyl)–,[1S–(1α,2β,4β)]–	1434.7	8.768	0	22
157*	Cyclohexanol, 5–methyl–2–(1–methylethyl)–, (1α,2α,5α)–	1170.0	6.338	0.043	13
158*	Cyclohexanone, 2–(1–mercapto–1–methylethyl)–5–methyl–	1166.4	6.629	–0.046	13
159	Cyclohexanone, 5–methyl–2–(1–methylethyl)–, (2R–cis)–	1146.7	6.338	0	13
160	Cyclohexene, 1–methyl–4–(1–methylethylidene)–	1089.8	5.968	0	13
161	Cyclopentaneacetic acid, 3–oxo–2–pentyl–, methyl ester	1642.3	9.927	–0.068	15
162*	D–Carvone	1183.2	6.576	0	13
163	Decanal	1195.7	6.596	–0.07	8
164	Decanedioic acid, diethyl ester	1745.5	10.686	–0.18	10
165*	Decanoic acid, ethyl ester	1411.3	8.287	–0.134	9
166	Diphenyl ether	1484.0	8.65	–0.033	13
167*	D–Limonene	1089.8	5.968	0	13
168*	Dodecanal	1379.8	7.875	–0.058	10
169*	Dodecanoic acid, ethyl ester	1599.9	9.567	–0.114	11
170*	Ethane, 1,1–diethoxy–	717.9	4.414	–0.437	2
171^	Ethanone, 1–(2,4–dimethylphenyl)–	1146.7	6.338	0	13
172	Ethanone, 1–(2–pyridinyl)–	1051.1	5.595	–0.088	6
173*	Ethanone, 1–(3–pyridinyl)–	1041.6	5.595	–0.088	7
174^	Ethanone, 1–(4–methylphenyl)–	1108.8	5.968	0	11

175*	Ethyl Oleate	2150.9	13.399	–0.079	17
176	Ethyl Vanillin	1346.9	7.486	–0.026	9
177	Eucalyptol	1158.9	6.557	–0.022	14
178*	Eugenol	1328.8	7.481	–0.023	11
179*	Formic acid, pentyl ester	838.4	4.759	–0.278	4
180^	Formic acid, phenylmethyl ester	1143.4	6.466	–0.112	9
181*	Furan, 2–[(methylthio)methyl]–	1028.0	5.026	0.013	5
182^	Furan, 2–methyl–5–(methylthio)–	1017.1	4.955	0.013	5
183*	Heptanoic acid	1013.5	5.126	–0.042	5
184^	Heptanoic acid, 3–methylbutyl ester	1374.0	8.106	–0.134	10
185*	Heptanoic acid, ethyl ester	1132.7	6.455	–0.181	6
186*	Hexadecanoic acid, ethyl ester	1968.9	12.123	–0.088	15
187^	Hexanal	811.7	4.027	–0.12	4
188*	Hexanoic acid, 2–methyl–	998.2	5.088	–0.042	6
189*	Hexanoic acid, 2–propenyl ester	1139.0	6.489	–0.179	6
190^	Hexanoic acid, ethyl ester	1015.5	5.715	–0.205	5
191*	Hexanoic acid, pentyl ester	1331.7	7.753	–0.147	8
192	Humulene	1453.6	8.768	0	20
193*	Hydrocoumarin	1278.9	7.213	–0.039	11
194	Isoamyl cinnamate	1627.9	9.816	–0.063	15
195^	Isobutyl acetate	758.0	4.236	–0.278	4
196^	Isopentyl hexanoate	1262.9	7.367	–0.147	9
197	Isophorone	994.7	5.349	0	13
198	Isopropyl myristate	1831.3	11.187	–0.094	14
199*	Menthyl isovalerate	1581.6	9.686	–0.059	18
200	Methional	842.8	3.494	0.053	2
201*	Methyl isovalerate	755.7	4.283	–0.278	5
202*	Methyl salicylate	1245.1	6.816	–0.024	9
203*	Methyleugenol	1344.7	8.138	–0.162	12
204*	Naphthalene, 2–ethoxy–	1450.1	8.542	–0.03	15
205	Naphthalene, 2–methoxy–	1373.2	8.056	–0.034	15
206^	n–Decanoic acid	1273.7	6.955	–0.028	8
207	Nonanal	1113.4	6.027	–0.078	7
208^	Nonanoic acid	1195.1	6.392	–0.031	7
209	Nonanoic acid, ethyl ester	1326.9	7.722	–0.147	8
210*	Nonanoic acid, methyl ester	1244.1	7.236	–0.162	8
211	n–Propyl acetate	686.7	3.864	–0.339	2
212^	Octanal, 7–hydroxy–3,7–dimethyl–	1162.1	6.352	–0.028	10
213^	Octanoic acid	1091.0	5.671	–0.036	6
214^	Octanoic acid, 3–methylbutyl ester	1454.3	8.655	–0.124	11
215^	Octanoic acid, ethyl ester	1217.9	7.004	–0.162	7
216	Octanoic acid, methyl ester	1133.7	6.523	–0.181	7
217	Oxacycloheptadec–8–en–2–one, (8Z)	1867.9	11.349	–0.055	15
218*	Oxacyclohexadecan–2–one	1768.5	10.655	–0.059	14
219	Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl ester Oxiranecarboxylic acid, 3–methyl–3–phenyl–, ethyl	1593.1	9.586	–0.097	13

	ester, cis–				
220*	p–Cymene	1089.8	5.968	0	13
221^	Pentanoic acid, 3–methyl–	879.1	4.283	–0.051	5
222^	Pentanoic acid, butyl ester	1138.0	6.489	–0.181	6
223^	Pentanoic acid, ethyl ester	927.7	5.192	–0.236	4
224	Pentyl octanoate	1523.0	9.04	–0.124	10
225*	Phenol, 2–(1–methylethyl)–	1147.6	5.973	0.07	11
226^	Phenol, 2–methoxy–	1072.6	5.602	–0.033	7
227^	Phenol, 2–methyl–5–(1–methylethyl)–	1215.8	6.576	0.06	13
228*	Phenylethyl Alcohol	1130.5	5.692	0.083	9
229*	Piperonal	1240.1	7.302	–0.153	10
230*	Propanedioic acid, diethyl ester	1001.6	6.092	–0.372	3
231	Propanethioic acid, S–(2–furanylmethyl) ester	1310.3	6.928	0.012	6
232	Propanoic acid, 2–methyl–	681.9	3	–0.086	3
233^	Propanoic acid, 2–methyl–, 2–phenylethyl ester	1451.6	8.584	–0.074	13
234*	Propanoic acid, 2–methyl–, 3–phenyl–2–propenyl ester	1558.9	9.323	–0.068	14
235	Propanoic acid, 2–methyl–, 4–formyl–2–methoxyphenyl ester	1588.1	9.741	–0.15	13
236	Propanoic acid, 2–methyl–, ethyl ester	771.5	4.324	–0.278	4
237^	Propanoic acid, 2–methyl–, octyl ester	1386.8	8.189	–0.134	10
238^	Propanoic acid, 2–methyl–, phenylmethyl ester	1376.8	8.06	–0.081	12
239	Pyrazine, 2,3–diethyl–5–methyl–	1203.3	6.813	–0.135	7
240	Pyrazine, 2,3–dimethyl–	912.6	4.977	–0.203	4
241*	Pyrazine, 2–ethyl–3,5–dimethyl–	1074.4	5.973	–0.152	6
242	Pyrazine, 2–ethyl–3–methyl–	1014.9	5.602	–0.174	5
243*	Pyrazine, 2–methoxy–3–(2–methylpropyl)–	1210.6	7.269	–0.233	8
244*	Pyrazine, 2–methyl–3–(methylthio)–	1111.9	5.602	–0.013	4
245^	Pyrazine, 2–methyl–6–(methylthio)–	1107.2	5.571	–0.013	4
246^	Pyrazine, 3–ethyl–2,5–dimethyl–	1107.5	6.188	–0.152	6
247*	Pyrazine, ethyl–	948.7	5.212	–0.203	4
248	Pyrazine, tetramethyl–	1054.3	5.842	–0.152	6
249*	Pyrazine, trimethyl–	983.3	5.396	–0.174	5
250^	Tetradecanoic acid, ethyl ester	1785.1	10.846	–0.1	13
251	Thiazole, 5–ethenyl–4–methyl–	1030.9	4.97	0.034	5
252^	trans–Isoeugenol	1328.8	7.481	–0.023	11
253	Triacetin	1327.7	8.343	–0.391	4
254	Triethyl citrate	1722.6	10.668	–0.252	8
255^	Undecanal	1297.6	7.296	–0.063	9
256^	α–Ionone	1345.2	7.939	0	18
257*	α–Irone	1435.1	8.647	0	20
258	α–Phellandrene	1089.8	5.968	0	13
259	β–Myrcene	1025.3	5.589	–0.064	10
260^	β–Phenylethyl butyrate	1475.5	8.678	–0.074	12
261	β–Pinene	1112.4	6.092	0.059	16
262*	δ–Nonalactone	1215.2	6.825	–0.099	8
263	2–Propenoic acid, ethyl ester	700.9[a]	3.939	–0.334	2

264	Acetoin	691.4[a]	3	–0.086	2
265	Pentanal	725.2[a]	3.494	–0.146	3
266	Propanoic acid	619.7[a]	2.613	–0.126	1
267	1–Propanol, 2–methyl–	611.5[a]	2.613	–0.106	3
268	Ethyl Acetate	524.8[a]	3.078	–0.432	1
269	Isopropyl acetate	617.5[a]	3.414	–0.339	2

* validation set. ^ test set. [a] predicted retention indices using Eq. 3