# Predicting Sorbent-Air Partition Coefficients for Terpenoids at Multiple Temperatures

Kavita M. Jeerage\* and Elijah N. Holland

Applied Chemicals and Materials Division, Material Measurement Laboratory, National Institute of Standards and Technology (NIST)

325 Broadway, Boulder, CO 80305

\*kavita.jeerage@nist.gov

## Abstract

Partition coefficients describe the relative concentration of a chemical equilibrated between two phases. In the design of air samplers, the sorbent-air partition coefficient is a critical parameter, as is the ability to extrapolate or predict partitioning at a variety of temperatures. Our specific interest is the partitioning of plant-derived terpenes (hydrocarbons formed from isoprene building blocks) and terpenoids (with oxygen-containing functional groups) in polydimethylsiloxane (PDMS) sorbents. To predict  $K_{PDMS/AIR}$  as a function of temperature for compounds containing carbon, hydrogen, and oxygen, we developed a group contribution model that explicitly incorporates the van't Hoff equation. For the 360 training compounds, predicted  $K_{PDMS/AIR}$  values strongly correlate ( $\mathbb{R}^2 > 0.987$ ) with  $K_{PDMS/AIR}$  values measured at temperatures from 60 °C to 200 °C. To validate the model with available literature data, we compared predictions for 50 additional C<sub>10</sub> compounds, including 6 terpenes and 22 terpenoids, with  $K_{PDMS/AIR}$  values measured at 100 °C and determined an average relative error of 3.1 %. We also compared predictions with  $K_{PDMS/AIR}$  values measured at 25 °C. The modeling approach developed here is advantageous for properties with limited experimental values at a single temperature.



## Introduction

Passive air sampling is an important technique for characterizing exposure to hazardous chemicals in indoor and outdoor environments.<sup>1-3</sup> Commercially-available personal exposure badges utilize activated carbon sorbents and are generally intended to capture industrial chemicals such as benzene, toluene, ethylbenzene, and xylenes (BTEX). Direct capture of vapor samples (e.g., with evacuated canisters) can also be effective for volatile organic chemicals but may not be effective for semi-volatile chemicals present in lower concentrations. For these chemicals, sampling schemes rely on capture and concentration by a sorbent material. Passive air samplers may utilize activated carbon, polyurethane foam, styrene-divinylbenzene copolymer resins, semi-permeable membranes filled with triolein, or polydimethylsiloxane (PDMS) sheets.<sup>1-5</sup> With the exception of activated carbon, these materials are intended to capture persistent organic pollutants (POPs) such as polybrominated diphenyl ethers (PBDEs), polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and organochlorine pesticides (OCPs). Interestingly, butter acts as an indirect sampling matrix for POPs that are captured from the air by pasture crops and fed to livestock, allowing them to concentrate in milk fats.<sup>6</sup>

Passive headspace sampling is a related technique that permits volatile or semi-volatile organic chemicals to be extracted from a complex solid or liquid matrix, concentrated, and released for identification. Sampling schemes often rely on capture by a sorbent material because direct capture of vapor samples (e.g., with gas-tight syringes) does not permit the detection of trace chemicals. For example, in current forensic science practice, the headspace of fire debris is sampled with activated carbon, which must be eluted with solvent to recover the adsorbed chemicals. Headspace solid phase microextraction (SPME) was developed as a solvent-free alternative in which adsorbed chemicals can be recovered by heating.<sup>7</sup> This technique utilizes a short (1 cm long), sorbentcoated glass fiber that is retracted into a stainless steel needle for protection once adsorption is complete. PDMS or other sorbents coat the fiber; PDMS is valued for its ability to capture a wide range of organic chemicals and withstand high injector temperatures. SPME has been applied to simulated fire debris samples<sup>8</sup> and illicit drugs and explosives.<sup>9</sup> Capillary microextraction of volatiles (CMV) is a related technique that utilizes a wide (2 mm diameter, 2 cm long) glass capillary packed with PDMS-coated glass microfibers, resulting in high surface area and sampling capacity.<sup>10</sup> Equilibration is accelerated by forcing headspace air through the capillary. Depending on the target analyte and matrix, the equilibration time can be reduced to as little as 30 s. Breath collection devices were recently created by connecting CMVs to a mouth piece to capture organic chemicals from the exhaled breath of cigarette smokers.<sup>11</sup>

The uptake profile of a passive air sampler has three regions. Mathematical models to predict uptake have been described in detail by others<sup>4</sup> and will only be discussed briefly here. The concentration of a target analyte in the sorbent initially increases linearly with time (kinetic region) but eventually equilibrates (thermodynamic region). The linear uptake region is often considered to end when the sampler has accumulated 25 % of the eventual equilibrium value. Similarly, when the sampler has accumulated 95 % of the eventual equilibrium value, it is in

the thermodynamic region. The transition region from approximately 25 % to 95 % of capacity requires consideration of kinetics and thermodynamics. In the kinetic region, Eq. 1 defines the mass of target analyte captured by the sorbent ( $M_{sorbent}$ ).

$$M_{sorbent} = k_{air} A_{sorbent} C_{air} t$$
<sup>[1]</sup>

In this equation,  $k_{air}$  is the air-side mass transfer coefficient (cm/s),  $A_{sorbent}$  is the planar area of the exposed sorbent material (cm<sup>2</sup>),  $C_{air}$  is the concentration of target analyte in air (ng/cm<sup>3</sup>) and t is the sampling time (s). The sampling rate,  $R_s = k_{air}A_{sorbent}$ , which has units of cm<sup>3</sup>/s, provides a sense of how much air is sampled by the sorbent. In the thermodynamic region, Eq. 2 defines the sorbent-air partition coefficient ( $K_{SORBENT/AIR}$ ), which is dimensionless.  $C_{sorbent}$  is the measured concentration of target analyte in the sorbent (ng/cm<sup>3</sup>) and one can calculate  $C_{air}$  if  $K_{SORBENT/AIR}$  is known.

$$K_{SORBENT/AIR} = \frac{C_{sorbent}}{C_{air}}$$
[2]

Passive air samplers can be designed to operate in either region. To sample in the kinetic region, the capacity of the sorbent must be high enough to avoid reaching the transition (or curvilinear) region. Personal exposure badges generally operate in the kinetic region; manufacturers provide  $R_s$  values and maximum sampling times for each chemical that can be sampled by their badges. Generic, chemical-independent  $R_s$  values have also been proposed for chemicals predominantly found in the vapor-phase rather than the particle-phase.<sup>5,12</sup> To sample in the thermodynamic region, the sorbent configuration must facilitate rapid equilibration. PDMS-coated glass fibers (100 µm sorbent thickness) equilibrate with volatile organic chemicals within minutes to hours,<sup>13,14</sup> whereas PDMS sheets (10<sup>4</sup> µm sorbent thickness) might require years to equilibrate with semi-volatile chemicals.<sup>15</sup> Ethylene vinyl acetate has been coated onto glass to create thin films that equilibrate with POPs within days or weeks.<sup>16</sup> For environmental monitoring, one advantage to operating in the thermodynamic region is that once equilibrium is reached, samplers can be retrieved at any time. More analyte mass is generally recovered, leading to a lower detection limit. However, the same sorbent may be in the kinetic region with respect to some analytes and in the thermodynamic region with respect to others.<sup>4</sup> Partition coefficients are needed for two purposes: to estimate the duration of the kinetic region and to calculate equilibrium concentrations in the thermodynamic region during adsorption.

There are several approaches to determining partition coefficients, depending on the sorbent and analyte(s) of interest.  $K_{PDMS/AIR}$  can be measured by equilibrating PDMS-coated fibers<sup>13,14,17,18</sup> or other PDMS-coated materials<sup>19</sup> with single compounds or mixtures. Sorbent thickness does not matter if equilibrium is achieved.<sup>17</sup> For semi-volatile compounds, constant linear airflow has been utilized to reduce the equilibration time.<sup>17</sup>  $K_{PDMS/AIR}$  can also be measured by isothermal gas chromatography and extrapolated to lower temperatures within the linear range of the van't Hoff equation. For example, Okeme et al. determined  $K_{PDMS/AIR}$  for 76 semi-volatile

organic compounds at temperatures from 60 °C to 190 °C and extrapolated these values to 25 °C.<sup>15</sup> Such measurements, again, do not depend on the stationary phase (sorbent) thickness because the specific retention volume is normalized by the stationary phase volume; however, the accuracy of stationary phase dimensions must be verified.<sup>20</sup> Furthermore, extrapolation can introduce errors because the relationship between  $\log K_{PDMS/AIR}$  and 1/T may not be linear over a large range of temperatures.

Semi-empirical prediction approaches have also been developed – most commonly a polyparameter linear free energy relationship (ppLFER) equation that models solute transfer between two phases at a single temperature.<sup>21,22</sup> For gas-liquid partition coefficients, this approach includes five solute descriptors: excess molar refractivity (E), dipolarity/polarity (S), hydrogen bond acidity (A) and basicity (B), and the logarithm of the Ostwald partition coefficient (L) into hexadecane at 298 K. The predictive equation has the form  $\log K = c + eE$ +sS + aA + bB + lL, where c, e, s, a, b, and l are sorbent-specific constants. Solute descriptors must be measured or predicted; sorbent-specific constants are determined by regression with training compounds for which the partition coefficient is known. Starting with  $K_{PDMS/AIR}$  values for 142 training compounds, Sprunger et al. developed a ppLFER equation to predict  $K_{PDMS/AIR}$  at a single temperature (25 °C).<sup>23</sup> Okeme et al. applied the equation developed by Sprunger et al. to 76 compounds; the authors concluded that the semi-empirical ppLFER approach was more successful than the theoretical COSMO-RS approach in predicting  $K_{PDMS/AIR}$  at 25 °C.<sup>15</sup> To predict  $K_{PDMS/AIR}$  at other temperatures, sorbent-specific constants must be determined from  $K_{PDMS/AIR}$  values at these temperatures, essentially developing a ppLFER equation at each temperature. Analogous ppLFER equations have been developed to predict the enthalpy of sorption ( $\Delta H_s$ ) on polyurethane foam<sup>24</sup> and styrenedivinvlbenzene resin.<sup>25</sup> Although  $\Delta H_S$  predictions can be used to adjust log K values to different temperatures, thus far training compounds have been limited (N < 55) and the quality of the correlation has been similarly limited ( $\mathbb{R}^2 \sim 0.85$ ).<sup>24,25</sup>

Group contribution models are based on the concept that pure component thermophysical and transport properties can be predicted solely by molecular structure. Group contribution models are additive – the contributions of each group, multiplied by its frequency within the compound, are summed to give the value for the pure compound.<sup>26</sup> While property data for training compounds are required to determine group contributions by regression, solute descriptors are not required. Joback & Reid<sup>27</sup> predicted temperature-independent properties such as the enthalpy of vaporization (H<sub>v</sub>, widely used for the design of vapor-liquid equilibrium-based processes) with 41 first-order groups. Their approach was expanded to treat compounds of greater complexity by additional first-order groups, second-order groups, second-order groups typically do not describe the entire molecule and may overlap. Second-order groups were created to provide more information for aliphatic and aromatic compounds with one ring, and to distinguish between isomers. Similarly, third-order groups were created for polycyclic compounds. Octanol-water partition coefficients (Kow, a measure of lipophilicity that influences

biodistribution and environmental fate) have been predicted at 25 °C with models containing second-order groups<sup>31</sup> and third-order groups.<sup>32</sup>  $K_{PDMS/AIR}$  has not been modelled by group contribution methods.<sup>26</sup>

Identification or quantitation of plant-derived compounds by vapor-phase analysis is important for determining intoxication, monitoring the air quality of indoor cannabis production facilities, and enforcing legal limits for cannabis possession. Cannabis plant material can be distinguished from similar plants by its major cannabinoids ( $\Delta$ -9-tetrahydrocannabinol or cannabidiol); however, cannabinoids have low vapor pressures<sup>33</sup> and only small quantities will be captured at ambient temperatures. Furthermore, highly-odorous compounds which may be important for cannabis detection by humans or trained canines are not necessarily the compounds present in the highest concentration in the vapor phase.<sup>34,35</sup> Recent investigations indicated that three sesquiterpenes ( $\alpha$ -santalene, valencene, and  $\beta$ -bisabolene) are unique to cannabis,<sup>36</sup> suggesting that terpenoids may be effective markers for cannabis for some applications. Cigarette smokers were distinguished from non-smokers by twelve chemicals in their exhaled breath, including the terpenes  $\beta$ -myrcene and limonene and the terpenoid citral.<sup>11</sup> Nicotine alone was a poor indicator of recent smoking, whereas detection of multiple chemicals in combination was more successful.<sup>11</sup> Recent cannabis users are likely to exhibit analogous differences in their breath profiles compared to non-users.

We are interested in quantitative measurements of exhaled breath, indoor environments such as greenhouses, isolated plant material, and thermal desorption of adsorbed compounds. These applications require partition coefficients at breath temperature (34 °C), greenhouse temperature (21 °C to 27 °C), or ambient temperature (-40 °C to 40 °C). Higher temperatures (60 °C to 80 °C) may be employed for passive headspace sampling in the lab, while 200 °C may be an appropriate desorption temperature for many compounds. Our goal is to enable prediction within the range 20 °C to 200 °C. In this work, we developed a group contribution model with 18 first-order groups to predict  $K_{PDMS/AIR}$  as a function of temperature. We created the model with data from 360 training compounds and validated the model at 100 °C with 50 additional C<sub>10</sub> compounds.  $K_{PDMS/AIR}$  values for the training compounds were calculated from Kovats retention indices and isothermal gas chromatography measurements at temperatures from 60 °C to 200 °C. The 360 training compounds contain only carbon, hydrogen, and oxygen, reflecting our interest in phytochemicals such as terpenoids. For the 360 training compounds,  $K_{PDMS/AIR}$  values predicted by the resulting model correlate with measured  $K_{PDMS/AIR}$  values ( $\mathbb{R}^2 > 0.987$ ). For the 50 additional C<sub>10</sub> validation compounds,  $K_{PDMS/AIR}$  values predicted at 100 °C have an average relative error of 3.1% compared to measured  $K_{PDMS/AIR}$  values. We find that the modeling approach developed here is advantageous for properties with limited experimental values at a single temperature.

### **Materials and Methods**

**Determining partition coefficients by IGC (isothermal gas chromatography).** We can calculate  $K_{PDMS/AIR}$  from isothermal retention times when the stationary phase is the sorbent of interest (Eq. 3).<sup>37</sup>  $K_{PDMS/AIR}$  values have been determined by this method for volatile organic compounds (n-alkanes and substituted benzenes)<sup>20</sup> and semi-volatile organic compounds containing bromine, chlorine, and/or phosphate functional groups.<sup>15</sup>

$$K_{PDMS/AIR} = \frac{F(t_{analyte} - t_{tracer})}{V_{PDMS}}$$
[3]

In Eq. 3, F is the flow rate of mobile phase (mL/min),  $t_{analyte}$  is the retention time of the analyte (min),  $t_{tracer}$ is the retention time of a non-retained chemical (min), and  $V_{PDMS}$  is the volume of the liquid stationary phase (mL). F was calculated at each temperature by measuring the flow rate of nitrogen carrier gas with a bubble flowmeter and applying corrections for water vapor and gas compressibility.  $V_{PDMS}$  was calculated from the column dimensions and the manufacturer-specified stationary phase thickness. Retention times were determined with a flame ionization detector and the non-retained tracer was methane for all experiments. For isothermal measurements, the practical temperature range depends on the compound of interest, because long retention times result in wide chromatographic peaks that are difficult to distinguish from baseline and short retention times result in peaks that cannot be distinguished from the non-retained tracer. Retention time measurements for C<sub>6</sub> to C<sub>16</sub> nalkanes were made at a series of temperatures from 60 °C to 200 °C, with at least 5 temperatures spanning a range of at least 60 °C for each chemical. For example, hexane was measured at temperatures from 60 °C to 140 °C, whereas hexadecane was measured from 140 °C to 200 °C. We employed an Agilent 6890 gas chromatograph with a flame ionization detector and ChemStation software. The J&W Scientific DB-1 capillary column had an inner diameter of 0.25 mm and a stationary phase thickness of 0.1 µm. The column was 29.0 m in length. Nine replicate measurements were made at each temperature. We calculated average retention times ( $t_x = t_{analyte}$  –  $t_{tracer}$ ) and average  $K_{PDMS/AIR}$  values for use in Eq. 4 and Eq. 8, respectively, as will be described.

Determining partition coefficients by combining IGC with literature  $I_x$  values. Kovats retention indices  $(I_x)$  convert isothermal retention times into dimensionless values by normalizing the retention time of any analyte by the retention times of the n-alkanes that elute before and after it.<sup>38</sup> We can calculate the retention time for our experimental conditions if  $I_x$  has been reported for a column with equivalent stationary phase chemistry (Eq. 4). This equation is simply a rearrangement of the equation defining the Kovats retention index. In Eq. 4,  $t_n$  and  $t_N$  are the retention times of n-alkanes with  $I_n < I_x < I_N$ , where n and N are the number of carbons in the smaller and larger n-alkanes, respectively. Importantly, data from capillary or packed columns can be employed and experimental parameters such as column dimensions, stationary phase thickness, mobile phase flow rate, and/or pressure can be different. Once  $t_x$  is obtained for our experimental conditions, we calculate  $K_{PDMS/AIR}$  from Eq.

$$logt_{x} = \left(\frac{I_{x}}{100} - n\right)(logt_{N} - logt_{n}) + logt_{n}$$

$$[4]$$

We determined  $K_{PDMS/AIR}$  values for a series of n-alkylbenzenes by both methods and compared them to values reported by Kloskowski et al.<sup>20</sup> (Fig. 1). Note that benzene could not be measured at 200 °C because it did not separate sufficiently from the non-retained tracer. At higher temperatures (150 °C and 200 °C), there is greater variability in the measured  $K_{PDMS/AIR}$  values, which suggests that temperature is the greatest source of uncertainty. The correspondence between our  $K_{PDMS/AIR}$  values for benzene, toluene, and ethylbenzene and values for these compounds measured with three stationary phases by Kloskowski et al.<sup>20</sup> provides verification of our stationary phase volume. The data in Fig. 1 also clearly demonstrates that  $K_{PDMS/AIR}$  values determined by combining IGC measurements with literature  $I_x$  values (Eq. 4 and Eq. 3) are equivalent to  $K_{PDMS/AIR}$  values for a diverse set of training compounds and the second approach enables us to utilize Kovats retention indices reported by other laboratories. Kovats retention indices were identified through the NIST Chemistry Webbook<sup>39</sup> for stationary phases equivalent to 100% polydimethylsiloxane.



**Figure 1.** Log  $K_{PDMS/AIR}$  as a function of 1 / T [K] for a series of n-alkylbenzenes. IGC (Eq. 3) was conducted at 50 °C, 75 °C, 100 °C, 150 °C, and 200 °C. Nine measurements are shown at each temperature. IGC was combined with literature I<sub>x</sub> values (Eq. 4 and Eq. 3) at 60 °C, 80 °C, 100 °C, 120 °C, 140 °C, 160 °C, and 180 °C. Only the regression lines are shown. Average values reported by Kloskowski et al. include columns with 1 µm, 5 µm, and 18 µm thick stationary phases.<sup>20</sup> Abbreviations: B = benzene; T = toluene; EB = ethylbenzene, PB = propylbenzene; BB = butylbenzene.

Model development and evaluation. Fig. 1 demonstrates that  $\log K_{PDMS/AIR}$  is proportional to 1/T over the temperatures investigated, in accordance with the van't Hoff equation (Eq. 5). Therefore, the slope (proportional

to  $\Delta$ H<sub>s</sub>, the enthalpy of sorption) and intercept (proportional to  $\Delta$ S<sub>s</sub>, the entropy of sorption) can be determined by linear regression, permitting  $K_{PDMS/AIR}$  values to be calculated at temperatures of interest within the linear range of the van't Hoff equation. Eq. 5 provides the starting point for a group contribution model that explicitly incorporates temperature dependence by predicting the slope and intercept as a function of molecular structure. We determined  $K_{PDMS/AIR}$  values by Eq. 3 for 21 training compounds, which included 11 n-alkanes and 10 compounds with aldehyde and/or ether groups. We determined  $K_{PDMS/AIR}$  values for all other training compounds by Eq. 4 and Eq. 3. To do this, we first identified compounds with  $I_x = 600 - 1600$  at three or more temperatures from the following: 60 °C, 80 °C, 100 °C, 120 °C, 140 °C, 150 °C, 160 °C, 180 °C, and 200 °C. We calculated  $K_{PDMS/AIR}$  from each  $I_x$  value and plotted log  $K_{PDMS/AIR}$  vs. 1/T for each compound. We used these plots to verify that  $K_{PDMS/AIR}$  values are consistent with the van't Hoff equation by examining the R<sup>2</sup> values resulting from linear regression for the slope and intercept. We eliminated compounds with R<sup>2</sup> < 0.95 from the training data. One compound was eliminated by this quality control check:  $\alpha$ -phellandrene (R<sup>2</sup> = 0.87). The remaining 339 compounds were added to the 21 compounds described above, resulting in 360 training compounds (Supplementary Table S1). Apart from the n-alkanes which ranged from C<sub>6</sub> to C<sub>16</sub>, training compounds ranged in size from C<sub>4</sub> alcohols to C<sub>15</sub> terpenes.

$$logK_{PRED} = \frac{Slope}{T} - Intercept$$
[5]

$$Slope = S + \sum_{i=1}^{18} n_i S_i$$
 [6]

$$Intercept = I + \sum_{i=1}^{18} n_i I_i$$
<sup>[7]</sup>

$$\sum_{j=1}^{N} (log K_{PRED} - log K_{MEAS})_j^2$$
[8]

We selected 18 first-order groups to describe the molecular structure of each training compound. For each training compound at each temperature, we predicted  $\log K_{PDMS/AIR}$  with the van't Hoff equation (Eq. 5). In this equation, the slope (Eq. 6) and intercept (Eq. 7) are the sum of contributions by first-order structural groups,  $S_i$  and  $I_i$ , respectively, multiplied by their frequency,  $n_i$ , and T is the absolute temperature. To determine values of  $S_i$  and  $I_i$  for the 18 first-order groups, we minimized the sum of the squared errors (Eq. 8). We applied a multistart approach to run a generalized reduced gradient (GRG) local solver from multiple starting points to reach a solution with high probability of being a global solution. N = 1625 and is greater than the number of training compounds, because each compound at each temperature generates a squared error. We chose this approach to minimize the error of the value we wish to predict ( $K_{PDMS/AIR}$ ) without reducing the model input to two values per compound.

In this way, compounds measured at more temperatures were weighted by the model more than compounds measured at fewer temperatures.

### **Results and Discussion**

**Group Contribution Model with 18 Groups.** The groups utilized here were compared with first-order groups utilized for carbon, hydrogen, and oxygen containing compounds by Joback & Reid (25 groups total),<sup>27</sup> Marrero & Gani (64 groups total),<sup>32</sup> and Stefanis et al. (37 groups total)<sup>31</sup> (Supplementary Table S2). There are significant differences in group definitions. Joback & Reid<sup>27</sup> define five cyclic groups for aliphatic and aromatic rings. Marrero & Gani<sup>32</sup> use cyclic groups for aliphatic rings only, whereas Stefanis et al.<sup>31</sup> do not define any cyclic groups for aliphatic rings. Marrero & Gani<sup>32</sup> define a variety of additional groups for substituents that replace hydrogen (e.g. aromatic carbon bound to alcohol, methoxy, acetyl, aldehyde, or acetate). Stefanis et al.<sup>31</sup> define only five such substituents and delineate fewer groups in general (e.g., compare ethers, ketones, and esters). Our groups match the Joback & Reid<sup>27</sup> groups, however, we utilized only one ether group and one ketone group due to the limited number of training compounds with these groups (18 and 12, respectively). Furthermore, the training compounds contain only one compound with a cyclic ether group (eucalyptol) and one compound with a cyclic ketone group (cyclohexanone).

Group	Slope (S <sub>i</sub> )	Intercept ( <i>I<sub>i</sub></i> )
CH <sub>3</sub>	144.865	0.221633
CH <sub>2</sub>	174.012	0.203179
СН	175.625	0.222433
С	135.718	0.127316
=CH <sub>2</sub>	177.947	0.348045
=CH	128.264	0.089427
=C	241.677	0.272969
CH <sub>2</sub> (cyc) <sup>a</sup>	160.692	0.170520
CH (cyc)	123.224	0.063547
C (cyc)	96.295	0.075337
=CH (cyc)	158.186	0.171851
=C (cyc)	223.060	0.225132
OH (alcohol)	339.926	0.415119
OH (phenol)	480.871	0.783309
0	192.570	0.209433
СО	411.400	0.398990
СНО	398.877	0.431350
COO	459.731	0.519668
CONSTANT	578.918	1.500359

Table 1. First-order groups and their contributions to the slope and intercept.

<sup>a</sup>Specifies groups found within a ring structure, which includes aromatic rings.

Group contributions (Table 1) are provided with six digits to avoid roundoff error when these values are used for prediction. To qualitatively examine model performance, correlation plots were separated based on functional groups (Fig. 2). Compounds in the plots for alkanes-alkenes, isoalkanes, cycloalkanes, or aromatic hydrocarbons appear only once, whereas compounds may appear more than once in the plots for alcohols-phenols, aldehydesketone-ethers, esters, or terpenoids, based on multiple oxygen-containing groups. For example, vanillin contains a phenol group, an aldehyde group, and an ether group, and therefore appears in two plots. Furthermore, each compound generates several pairs of predicted/measured values – one at each temperature included in the training data. Fig. 2 indicates the absolute error between model predictions and measurements in the form of vertical distance from the solid black line. For an individual compound, higher  $\log K_{PDMS/AIR}$  values are associated with lower temperatures. Predictions for the simplest compounds, alkanes and alkenes, require only three groups and, not surprisingly, have small absolute errors. Predictions for esters, which includes linear esters and benzoates, and therefore many more groups, also have small absolute errors. The model systematically underpredicts a subset of isoalkanes, whereas aromatic hydrocarbon outliers are both under- and over-predicted. Note that aromatic hydrocarbons include fused ring compounds such as indane and naphthalene. Compounds with alcohol, phenol, aldehyde, ketone, or ether functional groups have the largest absolute errors, which may reflect the simplified group assignments, the limited number of training compounds with these functional groups, or the presence of multiple oxygen-containing groups.



**Figure 2.** Correlation between predicted  $\log K_{PDMS/AIR}$  values and measured  $\log K_{PDMS/AIR}$  values for 360 training compounds. Solid black lines indicate 1:1 correspondence. Solid blue lines are regression lines.

Model performance statistics (Table 2) indicate that the overall correlation between predicted and measured values is high ( $R^2 > 0.987$ ). Many of the qualitative observations from Fig. 2 are quantified here; for example, the

average absolute error (AAE) of 0.07 log units for compounds with alcohol or phenol groups indicates lower predictive capability for these compounds. However, the maximum absolute error (AE<sub>MAX</sub>) is equal to 0.3 log units for individual compounds with a variety of functional groups, suggesting that model performance could be improved by a variety of approaches. AAE is frequently utilized to compare the performance of different models when the training compounds remain the same. For example, Marrero & Gani<sup>32</sup> developed a three-level group contribution model to predict Kow at 25 °C with a diverse set of more than 9500 training compounds. AAE for the training data decreased from 0.35 log units when first-order groups were employed, to 0.27 log units and 0.24 log units when second-order and third-order groups, respectively, were included. AAE for the training data modeled here is nearly an order of magnitude smaller (0.05 log units) but cannot be directly compared because the training compounds were different. One advantage to creating small models for compounds of interest rather than large, inclusive models is minimizing absolute errors. Average relative error (ARE) provides an estimate of uncertainty for model predictions based on the functional groups involved, meaning that predictions for a compound with an ester group have lower uncertainty than predictions for a compound with aldehyde or alcohol groups.

Table 2. Statistical performance of the group contribution model.

COMPOUNDS	Ν	<b>R</b> <sup>2</sup>	SD <sup>a</sup>	AAE <sup>b</sup>	AE <sub>MAX</sub>	ARE <sup>c</sup>	RE > 5% <sup>d</sup>
Alkane-Alkene (18)	102	0.998	0.04	0.02	0.1	1.4	1
Isoalkane (29)	121	0.987	0.07	0.05	0.2	3.0	4
Cycloalkane (30)	128	0.985	0.06	0.05	0.2	2.6	2
Aromatic Hydrocarbon (71)	318	0.984	0.07	0.05	0.3	2.2	3
Alcohol-Phenol (66)	306	0.979	0.09	0.07	0.3	3.2	8
Aldehyde-Ketone-Ether (43)	183	0.974	0.09	0.06	0.3	3.0	5
Ester (95)	422	0.990	0.05	0.04	0.3	1.7	1
Terpenoid (38)	184	0.980	0.07	0.06	0.2	2.5	6
Training Compounds (360)	1625	0.987	0.07	0.05	0.3	2.3	23

<sup>a</sup>Standard deviation (SD) is estimated by calculating the root mean squared error: SD =

 $\sqrt{\frac{1}{N}\sum(log K_{PRED} - log K_{MEAS})^2}$ . <sup>b</sup>Average absolute error (AAE) measures the deviation of predicted values from measured values:  $AAE = \frac{1}{N}\sum|log K_{PRED} - log K_{MEAS}|$ . <sup>c</sup>Average relative error (ARE) normalizes each deviation by the measured value:  $ARE = \frac{1}{N}\sum 100|(log K_{PRED} - log K_{MEAS})/log K_{MEAS}|$ . <sup>d</sup>At three or more temperatures.

Relative error (RE) can be used to identify compounds that are poorly described by the model. We identified compounds with RE > 5 % at three or more temperatures to provide direction for selecting second-order groups. Twenty-three compounds fell into this category; most had multiple substituents, such as multiple methyl groups or multiple oxygen-containing groups. Our model consistently underpredicts  $K_{PDMS/AIR}$  for four isoalkanes with multiple methyl groups: 2,2,3-trimethylbutane, 2,3,3-trimethylpentane, 2,3,4-trimethylpentane, and 2,2,3,3-

trimethylbutane. This leads to multiple points below the y = x line in Fig. 2. There are no alternative approaches to selecting first-order groups; however, the following second-order groups<sup>30</sup> provide additional structural information: (CH<sub>3</sub>)<sub>2</sub>CH, (CH<sub>3</sub>)<sub>3</sub>C, CH(CH<sub>3</sub>)CH(CH<sub>3</sub>), CH(CH<sub>3</sub>)C(CH<sub>3</sub>)<sub>2</sub>, and C(CH<sub>3</sub>)<sub>2</sub>C(CH<sub>3</sub>)<sub>2</sub>. Only nine of the isoalkanes do not include a second-order group and the four compounds identified above include up to four second-order groups (2 unique groups in each molecule). Second-order groups such as (CH<sub>3</sub>)<sub>2</sub>CH are also found in many terpenoids. Second-order groups have also been created to distinguish isomers, especially the isomers that result from substituted aromatic rings. Many training compounds with aromatic rings have multiple substitutions and predicted  $K_{PDMS/AIR}$  values for compounds such as o-, m-. and p-xylene or o-, m-, and pethyltoluene are approximately equal to the average  $K_{PDMS/AIR}$  value. For example, our model predicts  $K_{PDMS/AIR}$ for the isomers durene (1,2,4,5-tetramethylbenzene), isodurene (1,2,3,5-tetramethylbenzene), and prehnitene (1,2,3,4-tetramethylbenzene) identically. Durene and isodurene have similar (lower)  $K_{PDMS/AIR}$  values than prehnitene. Another example is o-hydroxybenzaldehyde, which our model overpredicts, and phydroxybenzaldehyde, which our model underpredicts. Second-order groups that distinguish these isomers are expected to improve predictions.

Model Predictions for Validation Compounds at 100 °C. Terpenes are based on a five-carbon isoprene unit (2-methyl-1,3-butadiene); monoterpenes consist of 2 isoprene units and sesquiterpenes consist of 3 isoprene units. We identified 50 ten-carbon (C<sub>10</sub>) validation compounds with Kovats retention indices at 100 °C, including 6 terpenes and 22 terpenoids (Supplementary Table S3). Compared to the training compounds, the validation compounds have more alcohol, ether, ketone, and aldehyde functional groups (Fig. 3). The reliability of  $K_{PDMS/AIR}$  values for these compounds is also lower because there was no possibility to verify consistency with the van't Hoff equation. Predicted values are correlated with measured values (Fig. 4). AAE and ARE for the validation compounds are comparable to values from the alcohol-phenol and ether-ketone-aldehyde classes of training compounds.



**Figure 3.** Distribution of oxygen-containing functional groups within the training compounds (360) and the  $C_{10}$  validation compounds (50).



**Figure 4.** Log  $K_{PDMS/AIR}$  values for 50 C<sub>10</sub> compounds were predicted at 100 °C and compared with log  $K_{PDMS/AIR}$  values calculated from Kovats retention indices. Solid black line indicates 1:1 correspondence. Solid red line is a regression line. AAE is 0.09 log units and ARE is 3.1 % for all 50 compounds.

**Model Predictions from 20** °C to 40 °C. We first compare model predictions to extrapolated values for the 360 training compounds. This approach assumes a constant slope over a wide temperature range (20 °C to 200 °C) and the error  $(\log K_{PRED} - \log K_{EXTR})$  indicates how well the model replicates single compound extrapolation. We examined histograms of the error at intervals of 20 °C and found that they were normally distributed (Supplementary Figure S1). The mean error is within ± 0.01 log units at each temperature. The standard deviation decreases from 0.14 log units at 20 °C to 0.06 log units at 200 °C, which indicates that predictions at 20 °C to 40 °C have greater absolute errors than predictions at higher temperatures, as we expect based on the temperature distribution of the training compounds (60 °C to 200 °C). However, the standard deviation is less than 3 % of the average log  $K_{EXTR}$  value at 20 °C, demonstrating that predictions at 20 °C to 40 °C do not *a priori* lead to greater uncertainty than predictions at higher temperatures (e.g., 100 °C). We next compare model predictions to measured values, which are limited.

Martos et al.<sup>13</sup> measured  $K_{PDMS/AIR}$  at 25 °C by equilibrating PDMS-coated fibers with a mixture of 29 isoalkanes (100 µm coating) or 33 aromatic hydrocarbons (30 µm coating). We used our model to predict  $K_{PDMS/AIR}$  at 25 °C (Fig. 5). The training data included 10 of the isoalkanes and 26 of the aromatic hydrocarbons. Fig. 5 shows that model predictions can be successfully extended to 25 °C, however, our model tends to overpredict  $K_{PDMS/AIR}$  for both isoalkanes and aromatic hydrocarbons. For the isoalkanes, RE was greater than 5 % for four compounds – 3-methylpentane, 2,4-dimethylpentane, 2-methylhexane, 2,5-dimethylhexane – and in each case the model overpredicted  $K_{PDMS/AIR}$ . This is particularly interesting because the model underpredicts  $K_{PDMS/AIR}$  for several isoalkane training compounds. Here the result may reflect the general observation that the model's RE at 25 °C is smallest for decane (0.0%); the model overpredicts smaller n-alkanes and underpredicts

larger n-alkanes. This is a consequence of the training compounds, whose average carbon number is in the range  $C_9 - C_{10}$ , and the n-alkane structure, which has two CH<sub>3</sub> groups and therefore only differs by the number of CH<sub>2</sub> groups. For the aromatic hydrocarbons, RE was greater than 5 % for one compound – isobutylbenzene.



**Figure 5.** Log  $K_{PDMS/AIR}$  values for 29 isoalkanes, 33 aromatic hydrocarbons, and 10 n-alkanes were predicted at 25 °C and compared to log  $K_{PDMS/AIR}$  values measured by Martos et al.<sup>13</sup> Solid black lines indicate 1:1 correspondence. Solid red lines are regression lines. AAE is 0.08 log units and ARE is 2.3 % for all 72 compounds.

 $K_{PDMS/AIR}$  was directly measured at 20 °C to 30 °C by equilibrating PDMS-coated stir bars with headspace vapors of naphthalene and camphor<sup>19</sup> and at 25 °C by equilibrating PDMS-coated fibers with aliphatic alcohols, aliphatic ketones and monoterpenes.<sup>14,18</sup> Each of these compounds was part of the training data, so several useful comparisons can be made (Fig. 6). Regression lines based on model input (at higher temperatures) indicate that extrapolated log  $K_{PDMS/AIR}$  values were consistently lower than measured log  $K_{PDMS/AIR}$  values for oxygen-containing compounds. On average, extrapolated values were 8 % lower. In contrast, extrapolated and measured log  $K_{PDMS/AIR}$  values for the five monoterpenes were within 2 % of each other. Fig. 6 indicates that the predictive ability of the group contribution model at 25 °C depends on whether extrapolation from higher temperatures is valid. This is true whether  $K_{PDMS/AIR}$  values at higher temperatures are measured or predicted. We also predicted log  $K_{PDMS/AIR}$  values at 25 °C by combining the predictive equation developed by Sprunger et al.<sup>23</sup> with solute descriptors for terpenes.<sup>40</sup> Our predictions were consistently larger than predictions from the Sprunger et al. model; both models had the highest absolute error for  $\alpha$ -pinene. This limited comparison indicates similar performance by the two approaches.



**Figure 6.** Log  $K_{PDMS/AIR}$  values for camphor were predicted at 20 °C, 25 °C, and 30 °C (stars) and compared to log  $K_{PDMS/AIR}$  values measured by De Coensel et al.<sup>19</sup> (open squares) and to log  $K_{PDMS/AIR}$  values extrapolated from higher temperatures (solid regression line). Predicted, measured, and extrapolated log  $K_{PDMS/AIR}$  values for three ketones and five terpenes are similarly plotted. Measurements by Isidorov et al.<sup>14</sup> and Nilsson et al.<sup>18</sup> Inset compares log  $K_{PDMS/AIR}$  measurements at 25 °C (open squares) to Sprunger et al.'s model<sup>23,40</sup> (triangles; AAE is 0.07 log units and ARE is 1.9 %) and our model (stars; AAE is 0.06 log units and ARE is 1.5 %).

 $K_{PDMS/AIR}$  values can be calculated from Kovats retention indices at many temperatures; however, the data at any one temperature is limited. For example, our model incorporates training data for nearly 300 compounds at 140 °C, whereas at 60 °C there are fewer than 200 compounds. By incorporating multiple temperatures, the group contribution method developed here increases model input. We further note that Kovats retention indices have been measured for hundreds of hydrocarbons and oxygen-containing organics at one (or perhaps two) temperatures.<sup>39</sup>  $K_{PDMS/AIR}$  values for such compounds cannot be extrapolated to other temperatures of interest. The quality control check that we implemented in this work only eliminated one compound from the training data. Incorporating such compounds into a future model would further increase model input.

#### Conclusions

We developed a group contribution model to predict  $K_{PDMS/AIR}$  from 20 °C to 200 °C. Our model employs 18 first-order groups to describe the molecular structure of 360 training compounds containing carbon, hydrogen, and oxygen. Predictions at 25 °C for 72 hydrocarbons have an average relative error of 2.3 %. Predictions at 100 °C for 50 C<sub>10</sub> validation compounds, including 6 terpenes and 22 terpenoids, have an average relative error of 3.1 %, which is comparable to average relative errors for the alcohol-phenol and aldehyde-ketone-ether classes of training compounds. We are interested in the partitioning of plant-derived terpenoids; additional training

compounds with oxygen-containing groups are a priority to improve the model. Ideas for acquiring such compounds were discussed. Our results demonstrate that incorporating the van't Hoff equation into a predictive group contribution model is a viable approach to estimating  $K_{PDMS/AIR}$  at multiple temperatures. The value of this modeling approach lies in its ability to incorporate all available data, which is particularly advantageous for properties with limited experimental values at a single temperature.

## Acknowledgments

The authors thank the NIST Special Programs Office (SPO) for its financial support of this work. Funding to E. Holland was provided through the Professional Research Experience Program (PREP) partnership between NIST and the University of Colorado at Boulder.

#### Disclaimer

Certain commercial entities are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that any of the entities identified are necessarily the best available for the purpose.

## **Supporting Information**

Document containing IUPAC InChI Keys, R<sup>2</sup> values from van't Hoff plots of log  $K_{PDMS/AIR}$  vs. 1/*T*, carbon number (#C), and classification for 360 training compounds (Table S1); first-order functional group definitions and comparisons to other group contribution models (Table S2); IUPAC InChI Keys, log  $K_{MEAS}$  values at 100 °C, log  $K_{PRED}$  values at 100 °C, relative error (RE), and classification for 50 validation compounds (Table S3); an example of an error histogram for log  $K_{PRED}$  - log  $K_{EXTR}$  at 100 °C, and mean error, standard deviation, and number of training compounds as a function of temperature.

### References

- 1. Lamplugh, A.; Harries, M.; Xiang, F.; Trinh, J.; Hecobian, A.; Montoya, L. D. Occupational exposure to volatile organic compounds and health risks in Colorado nail salons. *Environ. Pollut.* **2019**, *249*, 518-526.
- 2. Wilford, B. H.; Harner, T.; Zhu, J.; Shoeib, M.; Jones, K. C. Passive sampling survey of polybrominated diphenyl ether flame retardants in indoor and outdoor air in Ottawa, Canada: implications for sources and exposure. *Environ. Sci. Technol.* **2004**, *38*, 5312-5318.

- 3. Wania, F.; Shen, L.; Lei, Y. D.; Teixeira, C.; Muir, D. C. G. Development and calibration of a resin-based passive sampling system for monitoring persistent organic pollutants in the atmosphere. *Environ. Sci. Technol.* **2003**, *37*, 1352-1359.
- 4. Shoeib, M.; Harner, T. Characterization and comparison of three passive air samplers for persistent organic pollutants. *Environ. Sci. Technol.* **2002**, *36*, 4142-4151.
- 5. Okeme, J. O.; Nguyen, L. V.; Lorenzo, M.; Dhal, S.; Pico, Y.; Arrandale, V. H.; Diamond, M. L. Polydimethylsiloxane (silicone rubber) brooch as a personal passive air sampler for semi-volatile organic compounds. *Chemosphere* **2018**, *208*, 1002-1007.
- 6. Kalantzi, O. I.; Alcock, R. E.; Johnston, P. A.; Santillo, D.; Stringer, R. L.; Thomas, G. O.; Jones, K. C. The global distribution of polychlorinated biphenyls and organochlorine pesticides in butter. *Environ. Sci. Technol.* **2001**, *35*, 1013-1018.
- 7. Zhang, Z.; Pawliszyn, J. Headspace solid-phase microextraction. *Anal. Chem.* **1993**, *65*, 1843-1852.
- 8. Steffen, A.; Pawliszyn, J. Determination of liquid accelerants in arson suspected fire debris using headspace solidphase microextraction. *Anal. Comm.* **1996**, *33*, 129-131.
- 9. Guerra-Diaz, P.; Gura, S.; Almirall, J. R. Dynamic planar solid phase microextraction-ion mobility spectrometry for rapid field air sampling and analysis of illicit drugs and explosives. *Anal. Chem.* **2010**, *82*, 2826-2835.
- 10. Fan, W.; Almirall, J. R. High-efficiency headspace sampling of volatile organic compounds in explosives using capillary microextraction of volatiles (CMV) coupled to gas chromatography-mass spectrometry (GC-MS). *Anal. Bioanal. Chem.* **2014**, *406*, 2189-2195.
- 11. Hamblin, D.; Almirall, J. R. Analysis of exhaled breath from cigarette smokers using CMV-GC/MS. *Forensic Chem.* **2017**, *4*, 67-74.
- 12. Okeme, J. O.; Saini, A.; Yang, C.; Zhu, J.; Smedes, F.; Klanova, J.; Diamond, M. L. Calibration of polydimethylsiloxane and XAD-pocket passive air samplers (PAS) for measuring gas- and particle-phase SVOCs. *Atmos. Environ.* **2016**, *143*, 202 208.
- 13. Martos, P. A.; Saraullo, A.; Pawliszyn, J. Estimation of air/coating distribution coefficients for solid phase microextraction using retention indexes from linear temperature-programmed capillary gas chromatography: application to the sampling and analysis of total petroleum hydrocarbons in air. *Anal. Chem.* **1997**, *69*, 402-408.
- 14. Isidorov, V. A.; Vinogorova, V. T. Experimental determination and calculation of distribution coefficients between air and fiber with polydimethylsiloxane coating for some groups of organic compounds. *J. Chromatogr. A* **2005**, *1077*, 195-201.
- 15. Okeme, J. O.; Parnis, J. M.; Poole, J.; Diamond, M. L.; Jantunen, L. M. Polydimethylsiloxane-air partition ratios for semi-volatile organic compounds by GC-based measurement and COSMO-RS estimation: rapid measurements and accurate modelling. *Chemosphere* **2016**, *156*, 204-211.
- 16. Genualdi, S.; Harner, T. Rapidly equilibrating micrometer film sampler for priority pollutants in air. *Environ. Sci. Technol.* **2012**, *46*, 7661-7668.
- 17. Isetun, S.; Nilsson, U.; Colmsjo, A. Evaluation of solid-phase microextraction with PDMS for air sampling of gaseous organophosphate flame retardants and plasticizers. *Anal. Bioanal. Chem.* **2004**, *380*, 319-324.
- 18. Nilsson, T.; Larsen, T. O.; Montanarella, L.; Madsen, J. O. Application of headspace solid-phase microextraction for the analysis of volatile metabolites emitted by *Penicillium* species. *J. Microbiol. Meth.* **1996**, *25*,
- 19. De Coensel, N.; Desmet, K.; Gorecki, T.; Sandra, P. Determination of polydimethylsiloxane-air partition coefficients using headspace sorptive extraction. *J. Chromatogr. A* **2007**, *1150*, 183-189.
- 20. Kloskowski, A.; Chrzanowski, W.; Pilarczyk, M.; Namiesnik, J. Partition coefficients of selected environmentally important volatile organic compounds determined by gas-liquid chromatography with polydimethylsiloxane stationary phase. *J. Chem. Thermodynamics* **2005**, *37*, 21-29.
- 21. Abraham, M. H. Scales of solute hydrogen-bonding: their construction and application to physicochemical and biochemical processes. *Chem. Soc. Rev.* **1993**, *22*, 73-83.
- 22. Abraham, M. H.; Ibrahim, A.; Zissimos, A. M. Determination of sets of solute descriptors from chromatographic measurements. *J. Chromatogr. A* **2004**, *1037*, 29-47.
- 23. Sprunger, L.; Proctor, A.; Acree, W. E.; Abraham, M. H. Characterization of the sorption of gaseous and organic solutes onto polydimethylsiloxane solid-phase microextraction surfaces using the Abraham model. *J. Chromatogr. A* **2007**, *1175*, 162-173.
- 24. Kamprad, I.; Goss, K.-U. Systematic investigation of the sorption properties of polyurethane foams for organic vapors. *Anal. Chem.* **2007**, *79*, 4222-4227.

- 25. Hayward, S. J.; Lei, Y. D.; Wania, F. Sorption of a diverse set of organic chemical vapors on XAD-2 resin: measurement, prediction, and implications for air sampling. *Atmos. Environ.* **2011**, *45*, 296-302.
- 26. Harini, M.; Adhikari, J.; Rani, K. Y. A review on property estimation methods and computational schemes for rational solvent design: a focus on pharmaceuticals. *Ind. Eng. Chem. Res.* **2013**, *52*, 6869-6893.
- 27. Joback, K. G.; Reid, R. C. Estimation of pure-component properties from group-contributions. *Chem. Eng. Comm.* **1987**, *57*, 233-243.
- 28. Constantinou, L.; Gani, R. New group contribution method for estimating properties of pure compounds. *AIChE J.* **1994**, *40*, 1697-1710.
- 29. Kolska, Z.; Ruzicka, V.; Gani, R. Estimation of the enthalpy of vaporization and the entropy of vaporization for pure organic compounds at 298.15 K and at normal boiling temperature by a group contribution method. *Ind. Eng. Chem. Res.* **2005**, *44*, 8436-8454.
- 30. Marrero, J.; Gani, R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibr.* **2001**, *183-184*, 183-208.
- 31. Stefanis, E.; Constantinou, L.; Panayiotou, C. A group-contribution method for predicting pure component properties of biochemical and safety interest. *Ind. Eng. Chem. Res.* **2004**, *43*, 6253-6261.
- 32. Marrero, J.; Gani, R. Group-contribution-based estimation of octanol/water partition coefficient and aqueous solubility. *Ind. Eng. Chem. Res.* **2002**, *41*, 6623-6633.
- 33. Lovestead, T. M.; Bruno, T. J. Determination of cannabinoid vapor pressures to aid in vapor phase detection of intoxication. *Forensic Chem.* **2017**, *5*, 79-85.
- 34. Rice, S.; Koziel, J. A. (2015) Characterizing the smell of marijuana by odor impact of volatile compounds: an application of simultaneous chemical and sensory analysis. *PLOS One*, **10**: e0144160. https://doi.org/10.1371/journal.pone.0144160
- 35. Rice, S.; Koziel, J. A. The relationship between chemical concentration and odor activity value explains the inconsistency in making a comprehensive surrogate scent training tool representative of illicit drugs. *Forens. Sci. Int.* **2015**, *257*, 257-270.
- 36. Wiebelhaus, N.; Hamblin, D.; Kreitals, N. M.; Almirall, J. R. Differentiation of marijuana headspace volatiles from other plants and hemp products using capillary microextraction of volatiles (CMV) coupled to gaschromatography-mass spectrometry (GC-MS). *Forensic Chem.* **2016**, *2*, 1-8.
- 37. Conder, J. R.; Young, C. L. *Physicochemical Measurements by Gas Chromatography*. John Wiley & Sons: Chichester, 1979.
- 38. Kovats, E. Gas chromatographic characterization of organic compounds: retention indices of aliphatic halides, alcohols, aldehydes, and ketones. *Helv. Chim. Acta* **1958**, *41*, 1915-1932.
- 39. NIST Mass Spectrometry Data Center, William E. Wallace, Director, "Retention Indices" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. P.J. Linstrom and W.G. Mallard, National Institute of Standards and Technology, Gaithersburg, MD 20899, https://doi.org/10.18434/T4D303
- 40. Abraham, M. H.; Gola, J. R. M.; Gil-Lostes, J.; Acree, W. E.; Cometto-Muniz, J. E. Determination of solvation descriptors for terpene hydrocarbons from chromatographic measurements. *J. Chromatogr. A* **2013**, *1293*, 133-141.

# Supporting Information for Predicting Sorbent-Air Partition Coefficients for Terpenoids at Multiple Temperatures

Kavita M. Jeerage\* and Elijah N. Holland

Applied Chemicals and Materials Division, Material Measurement Laboratory, National Institute of Standards and Technology (NIST)

325 Broadway, Boulder, CO 80305

#### \*kavita.jeerage@nist.gov

**Table S1.** Training compounds (360) with Standard IUPAC InChI Key.  $R^2$  values reported from van't Hoff plots of log  $K_{PDMS/AIR}$  vs. 1/T. #C = carbon no. AA = Alkane-Alkene, IA = Isoalkane, CA = Cycloalkane, AH = Aromatic Hydrocarbon, AP = Alcohol-Phenol, AKE = Aldehyde-Ketone-Ether, E = Ester, T = Terpenoid

No	Compound	Standard IUPAC InChI Key	R <sup>2</sup>	# C	Class
1	hexane	VLKZOEOYAKHREP-UHFFFAOYSA-N	0.9993	6	AA
2	heptane	IMNFDUFMRHMDMM-UHFFFAOYSA-N	0.9994	7	AA
3	1-heptene	ZGEGCLOFRBLKSE-UHFFFAOYSA-N	0.9980	7	AA
4	octane	TVMXDCGIABBOFY-UHFFFAOYSA-N	0.9993	8	AA
5	1-octene	KWKAKUADMBZCLK-UHFFFAOYSA-N	0.9996	8	AA
6	nonane	BKIMMITUMNQMOS-UHFFFAOYSA-N	0.9995	9	AA
7	1-nonene	JRZJOMJEPLMPRA-UHFFFAOYSA-N	0.9996	9	AA
8	decane	DIOQZVSQGTUSAI-UHFFFAOYSA-N	0.9993	10	AA
9	1-decene	AFFLGGQVNFXPEV-UHFFFAOYSA-N	0.9996	10	AA
10	undecane	RSJKGSCJYJTIGS-UHFFFAOYSA-N	0.9992	11	AA
11	1-undecene	DCTOHCCUXLBQMS-UHFFFAOYSA-N	0.9994	11	AA
12	dodecane	SNRUBQQJIBEYMU-UHFFFAOYSA-N	0.9990	12	AA
13	1-dodecene	CRSBERNSMYQZNG-UHFFFAOYSA-N	0.9993	12	AA
14	tridecane	IIYFAKIEWZDVMP-UHFFFAOYSA-N	0.9997	13	AA
15	tetradecane	BGHCVCJVXZWKCC-UHFFFAOYSA-N	0.9994	14	AA
16	1-tetradecene	HFDVRLIODXPAHB-UHFFFAOYSA-N	0.9997	14	AA
17	pentadecane	YCOZIPAWZNQLMR-UHFFFAOYSA-N	0.9995	15	AA
18	hexadecane	DCAYPVUWAIABOU-UHFFFAOYSA-N	0.9994	16	AA
19	2-methyl-hexane	GXDHCNNESPLIKD-UHFFFAOYSA-N	0.9995	7	IA
20	3-methyl-hexane	VLJXXKKOSFGPHI-UHFFFAOYSA-N	0.9994	7	IA
21	2,2-dimethylpentane	CXOWYJMDMMMMJO-UHFFFAOYSA-N	0.9991	7	IA
22	2,3-dimethyl-pentane	WGECXQBGLLYSFP-UHFFFAOYSA-N	0.9959	7	IA
23	2,4-dimethyl pentane	BZHMBWZPUJHVEE-UHFFFAOYSA-N	0.9994	7	IA
24	3,3-dimethyl-pentane	AEXMKKGTQYQZCS-UHFFFAOYSA-N	0.9994	7	IA
25	2,2,3-trimethyl-butane	ZISSAWUMDACLOM-UHFFFAOYSA-N	0.9990	7	IA
26	3-methyl-heptane	LAIUFBWHERIJIH-UHFFFAOYSA-N	0.9992	8	IA
27	2,2-dimethyl-hexane	FLTJDUOFAQWHDF-UHFFFAOYSA-N	0.9997	8	IA

No	Compound	Standard IUPAC InChI Key	R <sup>2</sup>	# C	Class
28	2,3-dimethyl-hexane	JXPOLSKBTUYKJB-UHFFFAOYSA-N	0.9998	8	IA
29	2,4-dimethyl-hexane	HDGQICNBXPAKLR-UHFFFAOYSA-N	0.9986	8	IA
30	2,5-dimethyl-hexane	UWNADWZGEHDQAB-UHFFFAOYSA-N	0.9993	8	IA
31	3,3-dimethyl-hexane	KUMXLFIBWFCMOJ-UHFFFAOYSA-N	0.9996	8	IA
32	2,3,3-trimethyl-pentane	OKVWYBALHQFVFP-UHFFFAOYSA-N	0.9993	8	IA
33	2,3,4-trimethyl-pentane	RLPGDEORIPLBNF-UHFFFAOYSA-N	0.9996	8	IA
34	2,2,4-trimethyl-pentane	NHTMVDHEPJAVLT-UHFFFAOYSA-N	0.9998	8	IA
35	2,2,3,3-tetramethyl-butane	OMMLUKLXGSRPHK-UHFFFAOYSA-N	0.9998	8	IA
36	2,2,4-trimethyl-hexane	AFTPEBDOGXRMNQ-UHFFFAOYSA-N	0.9999	9	IA
37	2,2,5-trimethyl-hexane	HHOSMYBYIHNXNO-UHFFFAOYSA-N	0.9996	9	IA
38	2-methyl-nonane	SGVYKUFIHHTIFL-UHFFFAOYSA-N	0.9998	10	IA
39	4-methyl-nonane	IALRSQMWHFKJJA-UHFFFAOYSA-N	0.9998	10	IA
40	3-ethyl-octane	OEYGTUAKNZFCDJ-UHFFFAOYSA-N	0.9997	10	IA
41	2-methyl-decane	CNPVJWYWYZMPDS-UHFFFAOYSA-N	0.9995	11	IA
42	2,2-dimethyl-nonane	WDSBVMLUILIJOW-UHFFFAOYSA-N	0.9995	11	IA
43	2,4-dimethyl-nonane	JZUUOAUSQCXSTN-UHFFFAOYSA-N	0.9995	11	IA
44	2-methyl-undecane	GTJOHISYCKPIMT-UHFFFAOYSA-N	0.9993	12	IA
45	3-methyl-undecane	HTZWVZNRDDOFEI-UHFFFAOYSA-N	0.9987	12	IA
46	4-methyl-undecane	KNMXZGDUJVOTOC-UHFFFAOYSA-N	0.9991	12	IA
47	5-methyl-undecane	QULNVKABFWNUCW-UHFFFAOYSA-N	0.9987	12	IA
48	methyl-cyclopentane	GDOPTJXRTPNYNR-UHFFFAOYSA-N	0.9870	6	CA
49	ethyl-cyclopentane	IFTRQJLVEBNKJK-UHFFFAOYSA-N	0.9997	7	CA
50	propyl-cyclopentane	KDIAMAVWIJYWHN-UHFFFAOYSA-N	0.9996	8	CA
51	1-methylethyl-cyclopentane	TVSBRLGQVHJIKT-UHFFFAOYSA-N	1.0000	8	CA
52	1,2,3-trimethyl-cyclopentane	VCWNHOPGKQCXIQ-RNLVFQAGSA-N	0.9997	8	CA
53	butyl-cyclopentane	ZAGHKONXGGSVDV-UHFFFAOYSA-N	0.9997	9	CA
54	cyclohexane	XDTMQSROBMDMFD-UHFFFAOYSA-N	0.9995	6	CA
55	methyl-cyclohexane	UAEPNZWRGJTJPN-UHFFFAOYSA-N	0.9994	7	CA
56	ethyl-cyclohexane	IIEWJVIFRVWJOD-UHFFFAOYSA-N	0.9990	8	CA
57	propyl-cyclohexane	DEDZSLCZHWTGOR-UHFFFAOYSA-N	0.9992	9	CA
58	1-methylethyl-cyclohexane	GWESVXSMPKAFAS-UHFFFAOYSA-N	0.9994	9	CA
59	butyl-cyclohexane	GGBJHURWWWLEQH-UHFFFAOYSA-N	0.9996	10	CA
60	tert-butyl-cyclohexane	XTVMZZBLCLWBPM-UHFFFAOYSA-N	0.9999	10	CA
61	cis-1,2-dimethyl-cyclohexane	KVZJLSYJROEPSQ-OCAPTIKFSA-N	0.9999	8	CA
62	trans-1,2-dimethyl-cyclohexane	KVZJLSYJROEPSQ-YUMQZZPRSA-N	0.9993	8	CA
63	cis-1,3-dimethyl-cyclohexane	SGVUHPSBDNVHKL-OCAPTIKFSA-N	0.9946	8	CA
64	trans-1,3-dimethyl-cyclohexane	SGVUHPSBDNVHKL-YUMQZZPRSA-N	0.9974	8	CA
65	cis-1,4-dimethyl-cyclohexane	QRMPKOFEUHIBNM-OCAPTIKFSA-N	0.9994	8	CA
66	trans-1,4-dimethyl-cyclohexane	QRMPKOFEUHIBNM-ZKCHVHJHSA-N	0.9994	8	CA
67	cis-1,3,5-trimethyl-cyclohexane	ODNRTOSCFYDTKF-AYMMMOKOSA-N	0.9998	9	CA
68	trans-1,2,3,5-tetramethyl-cyclohexane	VWWAILZUSKHANH-ZYUZMQFOSA-N	0.9980	10	CA
69	cis-decalin	NNBZCPXTIHJBJL-AOOOYVTPSA-N	0.9998	10	CA
70	trans-decalin	NNBZCPXTIHJBJL-MGCOHNPYSA-N	0.9997	10	CA

No	Compound	Standard IUPAC InChI Key	R <sup>2</sup>	# C	Class
71	adamantane	ORILYTVJVMAKLC-UHFFFAOYSA-N	0.9953	10	CA
72	1-methyl-adamantane	UZUCFTVAWGRMTQ-UHFFFAOYSA-N	0.9997	11	CA
73	2-methyl-adamantane	VMODAALDMAYACB-UHFFFAOYSA-N	0.9993	11	CA
74	2-ethyl-adamantane	LIAWCKFOFPPVGF-UHFFFAOYSA-N	0.9958	12	CA
75	1,3,5-trimethyl-adamantane	WCACLGXPFTYVEL-UHFFFAOYSA-N	0.9999	13	CA
76	1-n-butyl-adamantane	AZCUIXHZJHFUFI-UHFFFAOYSA-N	0.9827	14	CA
77	2-n-butyl-adamantane	ZOQLCQXGZZFWAF-UHFFFAOYSA-N	0.9982	14	CA
78	benzene	UHOVQNZJYSORNB-UHFFFAOYSA-N	0.9989	6	AH
79	toluene	YXFVVABEGXRONW-UHFFFAOYSA-N	0.9995	7	AH
80	ethylbenzene	YNQLUTRBYVCPMQ-UHFFFAOYSA-N	0.9992	8	AH
81	o-xylene	CTQNGGLPUBDAKN-UHFFFAOYSA-N	0.9986	8	AH
82	m-xylene	IVSZLXZYQVIEFR-UHFFFAOYSA-N	0.9992	8	AH
83	p-xylene	URLKBWYHVLBVBO-UHFFFAOYSA-N	0.9978	8	AH
84	styrene	PPBRXRYQALVLMV-UHFFFAOYSA-N	0.9996	8	AH
85	propylbenzene	ODLMAHJVESYWTB-UHFFFAOYSA-N	0.9993	9	AH
86	cumene	RWGFKTVRMDUZSP-UHFFFAOYSA-N	0.9994	9	AH
87	o-ethyltoluene	HYFLWBNQFMXCPA-UHFFFAOYSA-N	0.9993	9	AH
88	m-ethyltoluene	ZLCSFXXPPANWQY-UHFFFAOYSA-N	0.9995	9	AH
89	p-ethyltoluene	JRLPEMVDPFPYPJ-UHFFFAOYSA-N	0.9992	9	AH
90	1,2,3-trimethylbenzene	FYGHSUNMUKGBRK-UHFFFAOYSA-N	0.9994	9	AH
91	1,2,4-trimethylbenzene	GWHJZXXIDMPWGX-UHFFFAOYSA-N	0.9991	9	AH
92	1,3,5-trimethylbenzene	AUHZEENZYGFFBQ-UHFFFAOYSA-N	0.9982	9	AH
93	allylbenzene	HJWLCRVIBGQPNF-UHFFFAOYSA-N	0.9997	9	AH
94	alpha-methylstyrene	XYLMUPLGERFSHI-UHFFFAOYSA-N	0.9997	9	AH
95	1-ethenyl-2-methylbenzene	NVZWEEGUWXZOKI-UHFFFAOYSA-N	0.9996	9	AH
96	1-ethenyl-3-methylbenzene	JZHGRUMIRATHIU-UHFFFAOYSA-N	0.9996	9	AH
97	indane	PQNFLIBBNBOBRQ-UHFFFAOYSA-N	0.9998	9	AH
98	indene	YBYIRNPNPLQARY-UHFFFAOYSA-N	0.9979	9	AH
99	butylbenzene	OCKPCBLVNKHBMX-UHFFFAOYSA-N	0.9991	10	AH
100	secbutylbenzene	ZJMWRROPUADPEA-UHFFFAOYSA-N	0.9991	10	AH
101	tertbutylbenzene	YTZKOQUCBOVLHL-UHFFFAOYSA-N	0.9992	10	AH
102	o-diethylbenzene	KVNYFPKFSJIPBJ-UHFFFAOYSA-N	0.9994	10	AH
103	m-diethylbenzene	AFZZYIJIWUTJFO-UHFFFAOYSA-N	0.9988	10	AH
104	p-diethylbenzene	DSNHSQKRULAAEI-UHFFFAOYSA-N	0.9997	10	AH
105	o-propyltoluene	YQZBFMJOASEONC-UHFFFAOYSA-N	0.9986	10	AH
106	m-propyltoluene	QUEBYVKXYIKVSO-UHFFFAOYSA-N	0.9989	10	AH
107	p-propyltoluene	JXFVMNFKABWTHD-UHFFFAOYSA-N	0.9995	10	AH
108	o-cymene	WWRCMNKATXZARA-UHFFFAOYSA-N	0.9983	10	AH
109	m-cymene	XCYJPXQACVEIOS-UHFFFAOYSA-N	0.9994	10	AH
110	p-cymene	HFPZCAJZSCWRBC-UHFFFAOYSA-N	0.9985	10	AH
111	3-ethyl-o-xylene	QUBBAXISAHIDNM-UHFFFAOYSA-N	0.9998	10	AH
112	4-ethyl-o-xylene	SBUYFICWQNHBCM-UHFFFAOYSA-N	0.9998	10	AH
113	4-ethyl-m-xylene	MEMBJMDZWKVOTB-UHFFFAOYSA-N	0.9986	10	AH

No	Compound	Standard IUPAC InChI Key	R <sup>2</sup>	# C	Class
114	5-ethyl-m-xylene	LMAUULKNZLEMGN-UHFFFAOYSA-N	0.9988	10	AH
115	2-ethyl-p-xylene	AXIUBBVSOWPLDA-UHFFFAOYSA-N	0.9985	10	AH
116	durene	SQNZJJAZBFDUTD-UHFFFAOYSA-N	0.9984	10	AH
117	isodurene	BFIMMTCNYPIMRN-UHFFFAOYSA-N	0.9970	10	AH
118	prehnitene	UOHMMEJUHBCKEE-UHFFFAOYSA-N	0.9954	10	AH
119	beta-methylindane	MWGYLUXMIMSOTM-UHFFFAOYSA-N	0.9922	10	AH
120	tetraline	CXWXQJXEFPUFDZ-UHFFFAOYSA-N	0.9993	10	AH
121	naphthalene	UFWIBTONFRDIAS-UHFFFAOYSA-N	0.9957	10	AH
122	pentylbenzene	PWATWSYOIIXYMA-UHFFFAOYSA-N	0.9978	11	AH
123	isopentylbenzene	XNXIYYFOYIUJIW-UHFFFAOYSA-N	0.9995	11	AH
124	terpentylbenzene	QHTJSSMHBLGUHV-UHFFFAOYSA-N	0.9996	11	AH
125	1-butyl-2-methylbenzene	NUJILYKLNKQOOX-UHFFFAOYSA-N	0.9994	11	AH
126	1-butyl-3-methylbenzene	OAPCPUDMDJIBOQ-UHFFFAOYSA-N	0.9995	11	AH
127	1-butyl-4-methylbenzene	SBBKUBSYOVDBBC-UHFFFAOYSA-N	0.9995	11	AH
128	4-tertbutyltoluene	QCWXDVFBZVHKLV-UHFFFAOYSA-N	0.9995	11	AH
129	1-ethyl-2-propylbenzene	DMUVQFCRCMDZPW-UHFFFAOYSA-N	0.9993	11	AH
130	1-ethyl-3-propylbenzene	QCYGXOCMWHSXSU-UHFFFAOYSA-N	0.9994	11	AH
131	1-ethyl-4-propylbenzene	ADQDTIAWIXUACV-UHFFFAOYSA-N	0.9995	11	AH
132	1-ethyl-2-isopropylbenzene	ZAJYARZMPOEGLK-UHFFFAOYSA-N	0.9994	11	AH
133	1-ethyl-3-isopropylbenzene	GSLSBTNLESMZTN-UHFFFAOYSA-N	0.9994	11	AH
134	1-ethyl-4-isopropylbenzene	GUUDUUDWUWUTPD-UHFFFAOYSA-N	0.9995	11	AH
135	3,5-diethyltoluene	HILAULICMJUOLK-UHFFFAOYSA-N	0.9999	11	AH
136	pentamethylbenzene	BEZDDPMMPIDMGJ-UHFFFAOYSA-N	0.9995	11	AH
137	1-methylnaphthalene	QPUYECUOLPXSFR-UHFFFAOYSA-N	0.9995	11	AH
138	2-methylnaphthalene	QIMMUPPBPVKWKM-UHFFFAOYSA-N	0.9976	11	AH
139	hexylbenzene	LTEQMZWBSYACLV-UHFFFAOYSA-N	0.9996	12	AH
140	1,2-diisopropylbenzene	OKIRBHVFJGXOIS-UHFFFAOYSA-N	0.9995	12	AH
141	1,3-diisopropylbenzene	UNEATYXSUBPPKP-UHFFFAOYSA-N	0.9996	12	AH
142	1,4-diisopropylbenzene	SPPWGCYEYAMHDT-UHFFFAOYSA-N	0.9996	12	AH
143	1,3,5-triethylbenzene	WJYMPXJVHNDZHD-UHFFFAOYSA-N	0.9996	12	AH
144	2-ethylnaphthalene	RJTJVVYSTUQWNI-UHFFFAOYSA-N	0.9963	12	AH
145	biphenyl	ZUOUZKKEUPVFJK-UHFFFAOYSA-N	0.9994	12	AH
146	diphenylmethane	CZZYITDELCSZES-UHFFFAOYSA-N	0.9995	13	AH
147	fluorene	NIHNNTQXNPWCJQ-UHFFFAOYSA-N	0.9993	13	AH
148	1,3,5-triisopropylbenzene	VUMCUSHVMYIRMB-UHFFFAOYSA-N	0.9994	15	AH
149	1-butanol	LRHPLDYGYMQRHN-UHFFFAOYSA-N	0.9968	4	AP
150	2-methyl-1-propanol	ZXEKIIBDNHEJCQ-UHFFFAOYSA-N	0.9860	4	AP
151	3-buten-1-ol	ZSPTYLOMNJNZNG-UHFFFAOYSA-N	0.9989	4	AP
152	1-pentanol	AMQJEAYHLZJPGS-UHFFFAOYSA-N	0.9944	5	AP
153	2-pentanol	JYVLIDXNZAXMDK-UHFFFAOYSA-N	0.9937	5	AP
154	3-pentanol	AQIXEPGDORPWBJ-UHFFFAOYSA-N	0.9987	5	AP
155	2-methyl-1-butanol	QPRQEDXDYOZYLA-UHFFFAOYSA-N	0.9996	5	AP
156	3-methyl-1-butanol	PHTQWCKDNZKARW-UHFFFAOYSA-N	0.9961	5	AP

No	Compound	Standard IUPAC InChI Key	R <sup>2</sup>	# C	Class
157	3-methyl-2-butanol	MXLMTQWGSQIYOW-UHFFFAOYSA-N	0.9914	5	AP
158	amylene hydrate	MSXVEPNJUHWQHW-UHFFFAOYSA-N	0.9854	5	AP
159	1-penten-3-ol	VHVMXWZXFBOANQ-UHFFFAOYSA-N	0.9995	5	AP
160	4-penten-1-ol	LQAVWYMTUMSFBE-UHFFFAOYSA-N	0.9798	5	AP
161	4-penten-2-ol	ZHZCYWWNFQUZOR-UHFFFAOYSA-N	0.9941	5	AP
162	1-hexanol	ZSIAUFGUXNUGDI-UHFFFAOYSA-N	0.9972	6	AP
163	2-hexanol	QNVRIHYSUZMSGM-UHFFFAOYSA-N	0.9970	6	AP
164	3-hexanol	ZOCHHNOQQHDWHG-UHFFFAOYSA-N	0.9997	6	AP
165	2-methyl-1-pentanol	PFNHSEQQEPMLNI-UHFFFAOYSA-N	0.9981	6	AP
166	4-methyl-1-pentanol	PCWGTDULNUVNBN-UHFFFAOYSA-N	0.9958	6	AP
167	2-methyl-2-pentanol	WFRBDWRZVBPBDO-UHFFFAOYSA-N	0.9965	6	AP
168	(Z)-3-hexen-1-ol	UFLHIIWVXFIJGU-ARJAWSKDSA-N	0.9856	6	AP
169	(E)-3-hexen-1-ol	UFLHIIWVXFIJGU-ONEGZZNKSA-N	0.9830	6	AP
170	cyclohexanol	HPXRVTGHNJAIIH-UHFFFAOYSA-N	0.9803	6	AP
171	1-heptanol	BBMCTIGTTCKYKF-UHFFFAOYSA-N	0.9980	7	AP
172	2-heptanol	CETWDUZRCINIHU-UHFFFAOYSA-N	0.9995	7	AP
173	4-heptanol	YVBCULSIZWMTFY-UHFFFAOYSA-N	0.9994	7	AP
174	1-octanol	KBPLFHHGFOOTCA-UHFFFAOYSA-N	0.9987	8	AP
175	1-octen-3-ol	VSMOENVRRABVKN-UHFFFAOYSA-N	0.9986	8	AP
176	2-ethyl-1-hexanol	YIWUKEYIRIRTPP-UHFFFAOYSA-N	0.9989	8	AP
177	1-nonanol	ZWRUINPWMLAQRD-UHFFFAOYSA-N	0.9977	9	AP
178	2-nonanol	NGDNVOAEIVQRFH-UHFFFAOYSA-N	0.9988	9	AP
179	1-decanol	MWKFXSUHUHTGQN-UHFFFAOYSA-N	0.9979	10	AP
180	3,7-dimethyl-3-octanol	DLHQZZUEERVIGQ-UHFFFAOYSA-N	0.9997	10	AP
181	1-undecanol	KJIOQYGWTQBHNH-UHFFFAOYSA-N	0.9992	11	AP
182	1-dodecanol	LQZZUXJYWNFBMV-UHFFFAOYSA-N	0.9988	12	AP
183	phenol	ISWSIDIOOBJBQZ-UHFFFAOYSA-N	0.9964	6	AP
184	o-cresol	QWVGKYWNOKOFNN-UHFFFAOYSA-N	0.9995	7	AP
185	m-cresol	RLSSMJSEOOYNOY-UHFFFAOYSA-N	0.9988	7	AP
186	p-cresol	IWDCLRJOBJJRNH-UHFFFAOYSA-N	0.9995	7	AP
187	phenylmethanol	WVDDGKGOMKODPV-UHFFFAOYSA-N	0.9985	7	AP
188	phenylethanol	WRMNZCZEMHIOCP-UHFFFAOYSA-N	0.9989	8	AP
189	2-ethyl-phenol	IXQGCWUGDFDQMF-UHFFFAOYSA-N	0.9997	8	AP
190	2,3-dimethyl-phenol	QWBBPBRQALCEIZ-UHFFFAOYSA-N	0.9988	8	AP
191	2,4-dimethyl-phenol	KUFFULVDNCHOFZ-UHFFFAOYSA-N	0.9994	8	AP
192	3,5-dimethyl-phenol	TUAMRELNJMMDMT-UHFFFAOYSA-N	0.9996	8	AP
193	thymol	MGSRCZKZVOBKFT-UHFFFAOYSA-N	0.9995	10	AP
194	2-ethoxyethanol	ZNQVEEAIQZEUHB-UHFFFAOYSA-N	0.9993	4	AP; AKE
195	pentanal	HGBOYTHUEUWSSQ-UHFFFAOYSA-N	0.9995	5	AKE
196	2-pentanone	XNLICIUVMPYHGG-UHFFFAOYSA-N	0.9971	5	AKE
197	3-pentanone	FDPIMTJIUBPUKL-UHFFFAOYSA-N	0.9998	5	AKE
198	hexanal	JARKCYVAAOWBJS-UHFFFAOYSA-N	0.9996	6	AKE
199	2-hexanone	QQZOPKMRPOGIEB-UHFFFAOYSA-N	0.9996	6	AKE

No	Compound	Standard IUPAC InChI Key	R2	# C	Class
200	methyl isobutyl ketone	NTIZESTWPVYFNL-UHFFFAOYSA-N	0.9994	6	AKE
201	2-methoxy-2-methyl-butane	HVZJRWJGKQPSFL-UHFFFAOYSA-N	0.9995	6	AKE
202	cyclohexanone	JHIVVAPYMSGYDF-UHFFFAOYSA-N	0.9997	6	AKE
203	heptanal	FXHGMKSSBGDXIY-UHFFFAOYSA-N	0.9995	7	AKE
204	2-heptanone	CATSNJVOTSVZJV-UHFFFAOYSA-N	0.9998	7	AKE
205	4-heptanone	HCFAJYNVAYBARA-UHFFFAOYSA-N	0.9999	7	AKE
206	octanal	NUJGJRNETVAIRJ-UHFFFAOYSA-N	0.9991	8	AKE
207	2-octanone	ZPVFWPFBNIEHGJ-UHFFFAOYSA-N	0.9999	8	AKE
208	nonanal	GYHFUZHODSMOHU-UHFFFAOYSA-N	1.0000	9	AKE
209	2-nonanone	VKCYHJWLYTUGCC-UHFFFAOYSA-N	0.9998	9	AKE
210	decanal	KSMVZQYAVGTKIV-UHFFFAOYSA-N	0.9999	10	AKE
211	undecanal	KMPQYAYAQWNLME-UHFFFAOYSA-N	0.9988	11	AKE
212	benzaldehyde	HUMNYLRZRPPJDN-UHFFFAOYSA-N	0.9962	7	AKE
213	o-hydroxybenzaldehyde	SMQUZDBALVYZAC-UHFFFAOYSA-N	0.9944	7	AP; AKE
214	p-hydroxybenzaldehyde	RGHHSNMVTDWUBI-UHFFFAOYSA-N	1.0000	7	AP; AKE
215	p-methylbenzaldehyde	FXLOVSHXALFLKQ-UHFFFAOYSA-N	0.9991	8	AKE
216	acetophenone	KWOLFJPFCHCOCG-UHFFFAOYSA-N	0.9969	8	AKE
217	o-hydroxyacetophenone	JECYUBVRTQDVAT-UHFFFAOYSA-N	0.9992	8	AP; AKE
218	o-methylacetophenone	YXWWHNCQZBVZPV-UHFFFAOYSA-N	0.9996	9	AKE
219	m-methylacetophenone	FSPSELPMWGWDRY-UHFFFAOYSA-N	0.9993	9	AKE
220	p-methylacetophenone	GNKZMNRKLCTJAY-UHFFFAOYSA-N	0.9987	9	AKE
221	anisole	RDOXTESZEPMUJZ-UHFFFAOYSA-N	0.9953	7	AKE
222	ethoxybenzene	DLRJIFUOBPOJNS-UHFFFAOYSA-N	0.9998	8	AKE
223	methoxymethylbenzene	GQKZBCPTCWJTAS-UHFFFAOYSA-N	0.9995	8	AKE
224	p-methoxybenzaldehyde	ZRSNZINYAWTAHE-UHFFFAOYSA-N	0.9979	8	AKE
225	vanillin	MWOOGOJBHIARFG-UHFFFAOYSA-N	0.9991	8	AP; AKE
226	p-ethoxybenzaldehyde	JRHHJNMASOIRDS-UHFFFAOYSA-N	0.9966	9	AKE
227	4-acetylanisole	NTPLXRHDUXRPNE-UHFFFAOYSA-N	0.9935	9	AKE
228	cuminaldehyde	WTWBUQJHJGUZCY-UHFFFAOYSA-N	1.0000	10	AKE
229	eugenol	RRAFCDWBNXTKKO-UHFFFAOYSA-N	0.9976	10	AP; AKE
230	acetic acid, propyl ester	YKYONYBAUNKHLG-UHFFFAOYSA-N	0.9736	5	E
231	acetic acid, 2-methoxyethyl ester	XLLIQLLCWZCATF-UHFFFAOYSA-N	0.9994	5	AKE; E
232	acetic acid, butyl ester	DKPFZGUDAPQIHT-UHFFFAOYSA-N	0.9955	6	E
233	acetic acid, 2-methylpropyl ester	GJRQTCIYDGXPES-UHFFFAOYSA-N	0.9997	6	E
234	acetic acid, pentyl ester	PGMYKACGEOXYJE-UHFFFAOYSA-N	0.9951	7	E
235	acetic acid, 3-methylbutyl ester	MLFHJEHSLIIPHL-UHFFFAOYSA-N	0.9993	7	E
236	acetic acid, hexyl ester	AOGQPLXWSUTHQB-UHFFFAOYSA-N	0.9989	8	E
237	acetic acid, heptyl ester	ZCZSIDMEHXZRLG-UHFFFAOYSA-N	0.9957	9	E
238	acetic acid, octyl ester	YLYBTZIQSIBWLI-UHFFFAOYSA-N	0.9996	10	E
239	1-octenyl-3-acetate	DOJDQRFOTHOBEK-UHFFFAOYSA-N	0.9999	10	E
240	acetic acid, nonyl ester	GJQIMXVRFNLMTB-UHFFFAOYSA-N	0.9997	11	E
241	acetic acid, decyl ester	NUPSHWCALHZGOV-UHFFFAOYSA-N	0.9998	12	E
242	acetic acid, dodecyl ester	VZWGRQBCURJOMT-UHFFFAOYSA-N	0.9982	14	E

No	Compound	Standard IUPAC InChI Key	R <sup>2</sup>	# C	Class
243	acetic acid, phenylmethyl ester	QUKGYYKBILRGFE-UHFFFAOYSA-N	0.9932	9	E
244	propanoic acid, methyl ester	RJUFJBKOKNCXHH-UHFFFAOYSA-N	0.9935	4	E
245	propanoic acid, ethyl ester	FKRCODPIKNYEAC-UHFFFAOYSA-N	0.9917	5	E
246	propanoic acid, propyl ester	MCSINKKTEDDPNK-UHFFFAOYSA-N	0.9822	6	E
247	propanoic acid, 1-methylethyl ester	IJMWOMHMDSDKGK-UHFFFAOYSA-N	0.9974	6	E
248	propanoic acid, 2-methyl-, ethyl ester	WDAXFOBOLVPGLV-UHFFFAOYSA-N	0.9998	6	E
249	propanoic acid, 2-propenyl ester	XRFWKHVQMACVTA-UHFFFAOYSA-N	0.9805	6	E
250	propanoic acid, butyl ester	BTMVHUNTONAYDX-UHFFFAOYSA-N	0.9987	7	E
251	propanoic acid, 1-methylpropyl ester	VPSLGSSVPWVZFG-UHFFFAOYSA-N	0.9949	7	E
252	propanoic acid, 2-methyl, propyl ester	AZFUASHXSOTBNU-UHFFFAOYSA-N	0.9943	7	E
253	propanoic acid, 1,1-dimethylethyl ester	JAELLLITIZHOGQ-UHFFFAOYSA-N	0.9909	7	E
254	propanoic acid, 3-butenyl ester	IVOGAUVYWHQIBD-UHFFFAOYSA-N	0.9868	7	E
255	propanoic acid, pentyl ester	TWSRVQVEYJNFKQ-UHFFFAOYSA-N	0.9982	8	E
256	propanoic acid, isopentyl ester	XAOGXQMKWQFZEM-UHFFFAOYSA-N	0.9966	8	E
257	propanoic acid, 1-methylbutyl ester	IPVKBEOJURLVER-UHFFFAOYSA-N	0.9958	8	E
258	propanoic acid, 1,1-dimethylpropyl ester	RHSLWHLBYFOBAF-UHFFFAOYSA-N	0.9866	8	E
259	propanoic acid, 1,2 dimethylpropyl ester	RYYMZRUXQWULCT-UHFFFAOYSA-N	0.9960	8	E
260	propanoic acid, 4-pentenyl ester	ROJSQZUHNAFOPE-UHFFFAOYSA-N	0.9952	8	E
261	propanoic acid, 1-methyl-3-butenyl ester	QTJLXWLLLXXECC-UHFFFAOYSA-N	0.9929	8	E
262	propanoic acid, hexyl ester	ester GOKKOFHHJFGZHW-UHFFFAOYSA-N		9	E
263	propanoic acid, Z-3-hexenyl ester	xenyl ester LGTLDEUQCOJGFP-WAYWQWQTSA-N		9	E
264	propanoic acid, E-3-hexenyl ester	-hexenyl ester LGTLDEUQCOJGFP-AATRIKPKSA-N		9	Е
265	propanoic acid, heptyl ester	BGYICJVBGZQOCY-UHFFFAOYSA-N	0.9990	10	E
266	propanoic acid, octyl ester	CEQGYPPMTKWBIU-UHFFFAOYSA-N	0.9991	11	Е
267	propanoic acid, nonyl ester	MPSVBCFDONBQFM-UHFFFAOYSA-N	0.9993	12	E
268	propanoic acid, decyl ester	HUOYUOXEIKDMFT-UHFFFAOYSA-N	0.9989	13	E
269	propanoic acid, undecyl ester	YYOMLCJPYHLLRY-UHFFFAOYSA-N	0.9995	14	E
270	propanoic acid, phenylmethyl ester	VHOMAPWVLKRQAZ-UHFFFAOYSA-N	0.9996	10	E
271	butanoic acid, methyl ester	UUIQMZJEGPQKFD-UHFFFAOYSA-N	0.9976	5	E
272	butanoic acid, ethyl ester	OBNCKNCVKJNDBV-UHFFFAOYSA-N	0.9958	6	E
273	butanoic acid, propyl ester	HUAZGNHGCJGYNP-UHFFFAOYSA-N	0.9989	7	E
274	butanoic acid, 1-methylethyl ester	FFOPEPMHKILNIT-UHFFFAOYSA-N	0.9980	7	E
275	butanoic acid, 2-propenyl ester	RMZIOVJHUJAAEY-UHFFFAOYSA-N	0.9778	7	E
276	butanoic acid, butyl ester	XUPYJHCZDLZNFP-UHFFFAOYSA-N	0.9993	8	E
277	butanoic acid, 1-methylpropyl ester	QJHDFBAAFGELLO-UHFFFAOYSA-N	0.9910	8	E
278	butanoic acid, 2-methylpropyl ester	RGFNRWTWDWVHDD-UHFFFAOYSA-N	0.9970	8	E
279	butanoic acid, 1,1-dimethylethyl ester	TWBUVVYSQBFVGZ-UHFFFAOYSA-N	0.9956	8	E
280	butanoic acid, 3-butenyl ester	VDMZXZOJKWISSE-UHFFFAOYSA-N	0.9913	8	E
281	butanoic acid, pentyl ester	CFNJLPHOBMVMNS-UHFFFAOYSA-N	0.9988	9	E
282	butanoic acid, 1-methylbutyl ester	DJOCFLQKCMWABC-UHFFFAOYSA-N	0.9957	9	E
283	butanoic acid, 3-methyl-, butyl ester	AYWJSCLAAPJZEF-UHFFFAOYSA-N	0.9964	9	E
284	butanoic acid, 1,1-dimethylpropyl ester	VTZSXMMBJHMLEE-UHFFFAOYSA-N	0.9957	9	E
285	butanoic acid, 1,2-dimethylpropyl ester	YBSMWGNIGFSKPE-UHFFFAOYSA-N	0.9955	9	E

No	Compound	Standard IUPAC InChI Key	R <sup>2</sup>	# C	Class
286	butanoic acid, 4-pentenyl ester	DTGXDHGNUDOWJS-UHFFFAOYSA-N	0.9918	9	E
287	butanoic acid, 1-methyl-3-butenyl ester	GBDHZZRHBRVVII-UHFFFAOYSA-N	0.9891	9	E
288	butanoic acid, hexyl ester	XAPCMTMQBXLDBB-UHFFFAOYSA-N	0.9980	10	E
289	butanoic acid, Z-3-hexenyl ester	ZCHOPXVYTWUHDS-WAYWQWQTSA-N	0.9932	10	E
290	butanoic acid, E-3-hexenyl ester	ZCHOPXVYTWUHDS-AATRIKPKSA-N	0.9924	10	Е
291	butanoic acid, heptyl ester	JPQHLIYIQARLQM-UHFFFAOYSA-N	0.9970	11	E
292	butanoic acid, octyl ester	PWLNAUNEAKQYLH-UHFFFAOYSA-N	0.9998	12	Е
293	butanoic acid, nonyl ester	RVNCBAPCNVAWOY-UHFFFAOYSA-N	0.9990	13	Е
294	butanoic acid, decyl ester	PUCQHFICPFUPKW-UHFFFAOYSA-N	0.9990	14	Е
295	pentanoic acid, ethyl ester	ICMAFTSLXCXHRK-UHFFFAOYSA-N	0.9960	7	Е
296	hexanoic acid, ethyl ester	SHZIWNPUGXLXDT-UHFFFAOYSA-N	0.9976	8	E
297	heptanoic acid, ethyl ester	TVQGDYNRXLTQAP-UHFFFAOYSA-N	0.9997	9	Е
298	octanoic acid, ethyl ester	YYZUSRORWSJGET-UHFFFAOYSA-N	0.9997	10	Е
299	nonanoic acid, ethyl ester	BYEVBITUADOIGY-UHFFFAOYSA-N	0.9999	11	Е
300	decanoic acid, ethyl ester	RGXWDWUGBIJHDO-UHFFFAOYSA-N	1.0000	12	Е
301	benzoic acid, methyl ester	QPJVMBTYPHYUOC-UHFFFAOYSA-N	0.9969	8	E
302	methyl salicylate	OSWPMRLSEDHDFF-UHFFFAOYSA-N	0.9988	8	AP; E
303	benzoic acid, ethyl ester	MTZQAGJQAFMTAQ-UHFFFAOYSA-N	0.9965	9	E
304	benzoic acid, propyl ester	UDEWPOVQBGFNGE-UHFFFAOYSA-N	0.9988	10	Е
305	benzoic acid, 1-methylethyl ester	FEXQDZTYJVXMOS-UHFFFAOYSA-N	0.9817	10	E
306	benzoic acid, 2-propenyl ester	LYJHVEDILOKZCG-UHFFFAOYSA-N	0.9830	10	Е
307	benzoic acid, butyl ester	XSIFPSYPOVKYCO-UHFFFAOYSA-N	0.9997	11	Е
308	benzoic acid, 1-methylpropyl ester	LSLWNAOQPPLHSW-UHFFFAOYSA-N	0.9928	11	E
309	benzoic acid, 2-methylpropyl ester	KYZHGEFMXZOSJN-UHFFFAOYSA-N		11	Е
310	benzoic acid, E-2-butenyl ester	QYOVCBKZFRCMCG-NSCUHMNNSA-N	0.9958	11	E
311	benzoic acid, 3-butenyl ester	HCAQKYMCZZDLGU-UHFFFAOYSA-N	0.9950	11	E
312	benzoic acid, pentyl ester	QKNZNUNCDJZTCH-UHFFFAOYSA-N	0.9982	12	E
313	benzoic acid, isopentyl ester	MLLAPOCBLWUFAP-UHFFFAOYSA-N	0.9984	12	E
314	benzoic acid, 1,2-dimethylpropyl ester	LTUYGFZKXSBLDW-UHFFFAOYSA-N	0.9962	12	E
315	benzoic acid, 4-pentenyl ester	MFUGAJWSACDNNH-UHFFFAOYSA-N	0.9989	12	E
316	benzoic acid, 1-methyl-3-butenyl ester	QLQPHDBTEQSJKE-UHFFFAOYSA-N	0.9954	12	E
317	benzoic acid, hexyl ester	UUGLJVMIFJNVFH-UHFFFAOYSA-N	0.9984	13	E
318	benzoic acid, cis-3-hexenyl ester	BCOXBEHFBZOJJZ-ARJAWSKDSA-N	0.9994	13	E
319	benzoic acid, trans-3-hexenyl ester	BCOXBEHFBZOJJZ-ARJAWSKDSA-N	0.9992	13	E
320	benzeneacetic acid, methyl ester	CRZQGDNQQAALAY-UHFFFAOYSA-N	0.9992	9	E
321	benzeneacetic acid, ethyl ester	DULCUDSUACXJJC-UHFFFAOYSA-N	0.9983	10	E
322	adamantane-1-carboxylic acid, methyl ester	CLYOOVNORYNXMD-UHFFFAOYSA-N	0.9993	12	E
323	alpha-terpinene	YHQGMYUVUMAZJR-UHFFFAOYSA-N	0.9973	10	Т
324	gamma-terpinene	YKFLAYDHMOASIY-UHFFFAOYSA-N	0.9991	10	Т
325	terpinolene	MOYAFQVGZZPNRA-UHFFFAOYSA-N	0.9979	10	Т
326	limonene	XMGQYMWWDOXHJM-UHFFFAOYSA-N	0.9986	10	Т
327	alpha-sabinene	NDVASEGYNIMXJL-UHFFFAOYSA-N	0.9986	10	Т
328	alpha-pinene	GRWFGVWFFZKLTI-UHFFFAOYSA-N	0.9968	10	Т

No	Compound	Standard IUPAC InChI Key	R2	# C	Class
329	L-alpha-pinene	GRWFGVWFFZKLTI-VEDVMXKPSA-N	0.9996	10	Т
330	beta-pinene	WTARULDDTDQWMU-UHFFFAOYSA-N	0.9978	10	Т
331	2-carene	IBVJWOMJGCHRRW-UHFFFAOYSA-N	0.9993	10	Т
332	alpha-thujene	KQAZVFVOEIRWHN-UHFFFAOYSA-N	0.9987	10	Т
333	camphene	CRPUJAZIXJMDBK-UHFFFAOYSA-N	0.9984	10	Т
334	beta-myrcene	UAHWPYUMFXYFJY-UHFFFAOYSA-N	0.9998	10	Т
335	cis-beta-ocimene	IHPKGUQCSIINRJ-NTMALXAHSA-N	0.9997	10	Т
336	trans-beta-ocimene	IHPKGUQCSIINRJ-CSKARUKUSA-N	0.9993	10	Т
337	myrtenal	KMRMUZKLFIEVAO-UHFFFAOYSA-N	0.9992	10	AKE; T
338	D-carvone	ULDHMXUKGWMISQ-SECBINFHSA-N	0.9992	10	AKE; T
339	pulegone	NZGWDASTMWDZIW-QMMMGPOBSA-N	0.9994	10	AKE; T
340	camphor	DSSYKIVIOFKYAU-UHFFFAOYSA-N	0.9983	10	AKE; T
341	beta-citral	WTEVQBCEXWBHNA-YFHOEESVSA-N	0.9995	10	AKE; T
342	myrtenol	RXBQNMWIQKOSCS-UHFFFAOYSA-N	0.9998	10	AP; T
343	cis-verbenol	WONIGEXYPVIKFS-UHFFFAOYSA-N	0.9990	10	AP; T
344	trans-verbenol	WONIGEXYPVIKFS-VGMNWLOBSA-N	0.9992	10	AP; T
345	alpha-terpineol	WUOACPNHFRMFPN-SECBINFHSA-N	0.9973	10	AP; T
346	gamma-terpineol	NNRLDGQZIVUQTE-UHFFFAOYSA-N	0.9980	10	AP; T
347	terpinen-4-ol	WRYLYDPHFGVWKC-UHFFFAOYSA-N	0.9988	10	AP; T
348	isopulegol	ZYTMANIQRDEHIO-AEJSXWLSSA-N	0.9990	10	AP; T
349	eucalyptol	WEEGYLXZBRQIMU-UHFFFAOYSA-N	0.9983	10	AKE; T
350	trans-4-thujanol	KXSDPILWMGFJMM-UHFFFAOYSA-N	0.9767	10	AP; T
351	borneol	DTGKSKDOIYIVQL-CCNFQMFXSA-N	0.9993	10	AP; T
352	linalool	CDOSHBSSFJOMGT-UHFFFAOYSA-N	0.9989	10	AP; T
353	lavandulol	CZVXBFUKBZRMKR-SNVBAGLBSA-N	0.9995	10	AP; T
354	cis-geraniol	GLZPCOQZEFWAFX-YFHOEESVSA-N	0.9986	10	AP; T
355	levomenthol	NOOLISFMXDJSKH-AEJSXWLSSA-N	0.9992	10	AP; T
356	citronellol	QMVPMAAFGQKVCJ-UHFFFAOYSA-N	0.9992	10	AP; T
357	linalyl acetate	UWKAYLJWKGQEPM-UHFFFAOYSA-N	0.9994	12	Е; Т
358	lavandulyl acetate	HYNGAVZPWWXQIU-UHFFFAOYSA-N	0.9990	12	Е; Т
359	caryophyllene	NPNUFJAVOOONJE-IOMPXFEGSA-N	0.9963	15	Т
360	beta-farnesene	JSNRRGGBADWTMC-NTCAYCPXSA-N	0.9995	15	Т

**Table S2.** First-order functional groups from this work compared to a subset of groups containing carbon, hydrogen,and oxygen defined by Joback & Reid (25 total), Marrero & Gani (64 total), and Stefanis et al. (37 total).

Group No	This Work	Joback & Reid (1987)	Marrero & Gani (2002)	Stefanis et al. (2004)
1	CH <sub>3</sub>	CH <sub>3</sub>	CH <sub>3</sub>	CH <sub>3</sub>
2	$CH_2$	$CH_2$	$CH_2$	$CH_2$
3	СН	СН	СН	СН
4	С	С	С	С

Group No	This Work	Joback & Reid (1987)	Marrero & Gani (2002)	Stefanis et al. (2004)
5	CU	CII	CH=CH <sub>2</sub>	CH=CH <sub>2</sub>
5	$=CH_2$	$=CH_2$	C=CH <sub>2</sub>	C=CH <sub>2</sub>
6	-CH	-CH	CH=CH	CH=CH
0	-СП	-сп	C=CH	C=CH
7	=C	=C	C=C	C=C
8	CH <sub>2</sub> (cyclic)	CH <sub>2</sub> (cyclic)	CH <sub>2</sub> (cyclic)	
9	CH (cyclic)	CH (cyclic)	CH (cyclic)	
10	C (cyclic)	C (cyclic)	C (cyclic)	
11	=CH (cyclic)	=CH (cyclic)	CH=CH (cyclic) C=CH (cyclic)	
12	=C (cyclic)	=C (cyclic)	C=C (cyclic) CH <sub>2</sub> =C (cyclic)	
13	OH (alcohol)	OH (alcohol)	ОН	OH
14	OH (phenol)	OH (phenol)	aC-OH	aC-OH
15	-0-	-O- -O- (cyclic)	CH <sub>3</sub> -O CH <sub>2</sub> -O CH-O C-O -O- (cyclic) aC-O	CH3-O CH2-O CH-O -O-
16	C=O	C=O C=O (cyclic)	CH <sub>3</sub> C=O CH <sub>2</sub> C=O CHC=O CC=O C=O (cyclic) aC-CO	CH <sub>3</sub> C=O CH <sub>2</sub> C=O CHC=O C=O
17	СНО	СНО	CHO aC-CHO	СНО
		СООН	COOH aC-COOH	СООН
18	COO	COO	$\begin{array}{c} CH_{3}COO\\ CH_{2}COO\\ CHCOO\\ CCOO\\ HCOO\\ aC-COO\\ aC-OOCH\\ aC-OOC\\ COO\\ COO\\ \end{array}$	CH3COO CH2COO HCOO COO
			aCH aC (fused aromatic ring) aC (fused non- aromatic subring) aC aC-CH <sub>3</sub> aC-CH <sub>2</sub> aC-CH	aCH aC aC-CH <sub>3</sub> aC-CH <sub>2</sub>
			aC-CH aC-CH=CH2 aC-CH=CH2 aC-CH=CH aC-C=CH2	aC-CH

**Table S3.** C<sub>10</sub> validation compounds (50) with standard IUPAC InChI Key.  $K_{MEAS}$  was calculated at 100 °C by combining IGC with literature  $I_x$  values.  $K_{PRED}$  was calculated at 100 °C with group contribution model. RE = relative error. AH = Aromatic Hydrocarbon, AP = Alcohol-Phenol, AKE = Aldehyde-Ketone-Ether, E = Ester, T = Terpenoid

No	Compound	Standard IUPAC InChI Key	log Kmeas	log K <sub>PRED</sub>	RE [%]	Class
1	isobutylbenzene	KXUHSQYYJYAXGZ-UHFFFAOYSA-N	2.47	2.53	2.4	AH
2	m-ethylstyrene	XHUZSRRCICJJCN-UHFFFAOYSA-N	2.67	2.62	1.9	AH
3	2,5-dimethylstyrene	DBWWINQJTZYDFK-UHFFFAOYSA-N	2.70	2.64	2.3	AH
4	2-methyl-1-phenyl-2-propanol	RIWRBSMFKVOJMN-UHFFFAOYSA-N	2.85	3.01	5.6	AP
5	carvacrol	RECUKUPTGUEGMW-UHFFFAOYSA-N	3.22	3.18	1.2	AP; T
6	benzilideneacetone	BWHOZHOGCMHOBV-UHFFFAOYSA-N	3.41	3.06	10.3	AKE
7	anethole	RUVINXPYWBROJD-ONEGZZNKSA-N	3.19	2.95	7.5	AKE
8	estragole	ZFMSMUAANRJZFM-UHFFFAOYSA-N	2.96	2.92	1.3	AKE
9	acetic acid, 1-phenylethyl ester	QUMXDOLUJCHOAY-UHFFFAOYSA-N	2.92	2.99	2.4	E
10	acetic acid, 2-phenylethyl ester	MDHYEMXUFSJLGV-UHFFFAOYSA-N	3.09	3.09	0.1	E
11	anisyl acetate	HFNGYHHRRMSKEU-UHFFFAOYSA-N	3.51	3.42	2.6	AKE; E
12	butanoic acid, 2-methyl-,3- methylbutyl ester	VGIRHYHLQKDEPP-UHFFFAOYSA-N	2.71	2.72	0.2	E
13	butanoic acid, 3-methyl-,3- methylbutyl ester	XINCECQTMHSORG-UHFFFAOYSA-N	2.71	2.72	0.1	E
14	cis-alloocimene	GQVMHMFBVWSSPF-SOYUKNQTSA-N	2.83	2.48	12.2	Т
15	trans-alloocimene	GQVMHMFBVWSSPF-SOYUKNQTSA-N	2.79	2.48	11.0	Т
16	dihydrocitronellol	PRNCMAKCNVRZFX-UHFFFAOYSA-N	2.96	2.86	3.5	AP
17	tetrahydromyrcenol	WRFXXJKURVTLSY-UHFFFAOYSA-N	2.71	2.75	1.6	AP
18	tetrahydrolavandulol	SFIQHFBITUEIBP-SNVBAGLBSA-N	2.87	2.75	4.2	AP
19	9-decen-1-ol	QGFSQVPRCWJZQK-UHFFFAOYSA-N	3.14	3.04	3.2	AP
20	6,7-dihydrolinalool	IUDWWFNDSJRYRV-UHFFFAOYSA-N	2.79	2.70	3.3	AP
21	dihydromyrcenol	XSNQECSCDATQEL-UHFFFAOYSA-N	2.61	2.70	3.6	AP; T
22	trans-geraniol	GLZPCOQZEFWAFX-JXMROGBWSA-N	3.11	3.09	0.4	AP; T
23	citronellal	NEHNMFOYXAPHSD-UHFFFAOYSA-N	2.84	2.86	0.6	AKE; T
24	hydroxycitronellal	WPFVBOQKRVRMJB-UHFFFAOYSA-N	3.17	3.22	1.7	AP; AKE
25	beta-phellandrene	LFJQCDVYDGGFCH-UHFFFAOYSA-N	2.54	2.42	4.4	Т
26	cis-carveol	BAVONGHXFVOKBV-UWVGGRQHSA-N	3.05	3.06	0.3	AP; T
27	trans-carveol	BAVONGHXFVOKBV-VHSXEESVSA-N	3.02	3.06	1.3	AP; T
28	dihydrocarveol	KRCZYMFUWVJCLI-UHFFFAOYSA-N	2.96	2.96	0.0	AP; T
29	carvone	ULDHMXUKGWMISQ-UHFFFAOYSA-N	3.09	3.00	2.7	AKE; T
30	carvenone	RLYSXAZAJUMULG-UHFFFAOYSA-N	3.11	2.91	6.4	AKE; T
31	piperitone	YSTPAHQEHQSRJD-UHFFFAOYSA-N	3.11	2.91	6.2	AKE; T
32	tetrahydrocarvone	GCRTVIUGJCJVDD-RKDXNWHRSA-N	2.97	2.82	5.0	AKE; T
33	isocineole	RFFOTVCVTJUTAD-UHFFFAOYSA-N	2.52	2.51	0.2	AKE; T
34	menthol	NOOLISFMXDJSKH-KXUCPTDWSA-N	2.92	2.88	1.7	AP; T
35	neo-menthol	NOOLISFMXDJSKH-UHFFFAOYSA-N	2.89	2.88	0.4	AP; T
36	cis-menthone	NFLGAXVYCFJBMK-RKDXNWHRSA-N	2.89	2.82	2.4	AKE; T
37	trans-menthone	NFLGAXVYCFJBMK-DTWKUNHWSA-N	2.84	2.82	0.9	AKE; T
38	methofurane	YGWKXXYGDYYFJU-UHFFFAOYSA-N	2.90	3.11	7.2	AKE; T

No	Compound	Standard IUPAC InChI Key	log К <sub>МЕАS</sub>	log K <sub>PRED</sub>	RE [%]	Class
39	terpin	RBNWAMSGVWEHFP-UHFFFAOYSA-N	3.23	3.27	1.2	AP
40	isocyclocitral	DEMWVPUIZCCHPT-UHFFFAOYSA-N	2.79	2.87	2.8	AKE
41	isocyclocitral (isomer)	YJSUCBQWLKRPDL-UHFFFAOYSA-N	2.86	2.87	0.6	AKE
42	tetrahydrocitral	UCSIFMPORANABL-UHFFFAOYSA-N	2.89	2.74	5.4	AKE
43	delta-3-carene	BQOFWKZOCNGFEC-DTWKUNHWSA-N	2.49	2.41	3.3	Т
44	thujone	USMNOWBWPHYOEA-MRTMQBJTSA-N	2.72	2.74	0.7	AKE; T
45	alpha-fenchene	XCPQUQHBVVXMRQ-UHFFFAOYSA-N	2.34	2.38	1.6	Т
46	beta-fenchene	FUIDRYCKEXJNOK-SCZZXKLOSA-N	2.32	2.21	5.0	Т
47	alpha-fenchol	IAIHUHQCLTYTSF-ZCUBBSJVSA-N	2.77	2.73	1.7	AP; T
48	fenchone	LHXDLQBQYFFVNW-UHFFFAOYSA-N	2.70	2.67	1.1	AKE; T
49	isoborneol	DTGKSKDOIYIVQL-CCNFQMFXSA-N	2.89	2.73	5.7	AP; T
50	verbenone	DCSCXTJOXBUFGB-UHFFFAOYSA-N	2.99	2.86	4.5	AKE; T

**Figure S1.** Histograms were created by binning errors calculated from  $\log K_{PRED} - \log K_{EXTR}$  for each compound.  $K_{PRED}$  was calculated with the group contribution model and  $K_{EXTR}$  was extrapolated to 20 °C, 40 °C, 60 °C, 80 °C, 100 °C, 120 °C, 140 °C, 160 °C, 180 °C, and 200 °C by the van't Hoff equation. The histogram at 100 °C is provided as an example (a). The mean and the standard deviation of the error (b) and the number of training compounds (c) are plotted as a function of temperature.

