The determination of solvation descriptors for terpenes, and the prediction of nasal pungency thresholds

Michael H. Abraham,[a] **Rachel Kumarsingh,**[a] **J. Enrique Cometto-Muñiz,**[b] **William S. Cain,**[b] **Martí Rosés,**[c] **Elizabeth Bosch**[c] **and Maria Luisa Díaz**[c]

[a] *Department of Chemistry, University College London, 20 Gordon Street, London, UK WC1H 0AJ*

[b] *Chemosensory Perception Laboratory, Department of Surgery (Otolaryngology), University of California, San Diego, Mail Code 0957, La Jolla, CA 92093-0957, USA*

[c] *Departament de Química Analítica, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain*

We have previously put forward two solvation equations for the correlation of chemical and biochemical phenomena, including nasal pungency thresholds (NPT) in man. In order to make use of these equations, a number of descriptors need to be assigned to any given solute, *viz.* π_2^H the dipolarity/polarizability, Σa_2^H and $\Sigma \beta_2^H$ the overall hydrogen-bond acidity and basicity, and log L^{16} where L^{16} is the solute gas–hexadecane partition coefficient. We show that these descriptors can be obtained from data from a number of processes, including gas-liquid chromatography, high performance liquid chromatography, gas–water partitions, gas–solvent partitions and water–solvent partitions. In this way, the above descriptors have been obtained for eleven terpenes, including hydrocarbons, aldehydes, ketones, alcohols and an ether. The π_2^H, Σa_2^H and log L^{16} descriptors have also been obtained for a further 21 terpenes, but $\Sigma \beta_2^H$ now has to be estimated. Through our previously reported equation, we have predicted NPT values for the total 32 terpenes. Small hydrocarbons such as α-pinene are predicted to have large NPT values, but aldehydes (*e.g.* geranial) and ketones (*e.g.* camphor) are predicted to have moderate values, and alcohols (*e.g.* menthol and borneol) and their acetates are predicted to have very small NPT values.

There is increasing concern over health impacts of indoor air pollution.[1] Important pollutants are not only compounds such as formaldehyde, but volatile organic compounds (VOCs) in general. Most people spend 70 to 90% of their time indoors, where the concentration of VOCs in the atmosphere is typically from 2 to 20 times greater than concentrations found outdoors.[1,2] The terpenes are one class of VOCs that is prevalent in indoor air, as constituents of varnishes, lacquers, *etc.*, and as emissions from flooring materials.[3] Limonene and α-pinene are the most abundant,[4] together with[3] β-pinene, camphene, 3-carene, longifolene, *trans*-caryophyllene and α-cedrene. Household products can also emit VOCs; these include terpene alcohols and terpene acetates.[4,5]

The perceived effect of VOCs can be divided into odor and sensory irritation, the latter being so important that 40% of the workplace threshold limit values (TLVs) of the American Conference of Governmental Industrial Hygienists are based on this effect.[6] Sensory irritation includes both nasal pungency and eye irritation. Bio-assays for nasal pungency are thus very relevant to the assessment of indoor air quality. In an industrial environment, the number of VOCs that could be encountered is several thousand, and even in non-industrial buildings several hundred VOCs have been identified.[7] It is quite impractical for bio-assays on sensory irritation to be conducted on this number of VOCs, using panels of human subjects.

In order to overcome this problem, we have devised[8] a predictive equation for nasal pungency thresholds (NPT) in man, using the general solvation equation (1),[9] where the dependent variable log SP is 'some property' of a series of VOCs (solutes) in a given phase system. For nasal pungency thresholds log SP

$$\log SP = c + rR_2 + s\pi_2^H + a\Sigma a_2^H + b\Sigma \beta_2^H + l \log L^{16} \quad (1)$$

is log (1/NPT), where NPT is the nasal pungency threshold in ppm. For the solubility of VOCs in solvent phases, log SP is taken as log L where L is the Ostwald solubility coefficient defined by eqn. (2). The independent variables in eqn. (1) are as

$$L = [\text{conc. of VOC in phase}]/[\text{conc. of VOC in gas}] \quad (2)$$

follows.[9] R_2 is an excess molar refraction, π_2^H is the dipolarity/polarizability, Σa_2^H and $\Sigma \beta_2^H$ are the overall hydrogen-bond acidity and basicity, and log L^{16} is a descriptor where L^{16} is the solute L-coefficient on hexadecane at 298 K.

For 43 values of NPT, eqn. (1) can be stated as eqn. (3),[8]

$$\log (1/NPT) = -8.519 + 2.154 \, \pi_2^H + 3.522 \, \Sigma a_2^H + \\ 1.397 \, \Sigma \beta_2^H + 0.860 \log L^{16} \quad (3)$$

where $n = 43$, $\rho = 0.977$, sd $= 0.27$, $F = 201$. Here, and elsewhere, n is the number of data points, ρ is the correlation coefficient, sd is the standard deviation, and F is the F-statistic. Note that the rR_2 term in eqn. (1) was statistically not significant and has been omitted in eqn. (3). In order to predict NPT values for terpenes, the four descriptors in eqn. (3) need to be determined for each terpene individually; this is the aim of the present work.

Methodology

As we have shown before,[10,11] values of log L^{16} can be obtained from gas-liquid chromatographic (GLC) data using a non-polar stationary phase. Eqn. (1) then reduces to eqn. (4). Once the

$$\log SP = c + rR_2 + l \log L^{16} \quad (4)$$

system has been calibrated with compounds of known log L^{16}

Table 1 Coefficients in eqn. (1) and eqn. (6) for various processes

No.	Process	c	r	s	a	b	$1/v$
A	**Gas to solvent, eqn. (1)**						
1	OV1, GLC[a]	0.705	0.000	0.746	0.683	0.000	2.001
2	SE54, GLC[a]	0.578	−0.419	1.430	0.983	0.000	2.010
3	PEG, GLC[a]	−5.496	0.960	5.629	8.841	0.000	2.037
4	OV101, GLC[b]	0.719	−0.180	0.854	0.768	0.000	1.983
5	OV101, GLC[b]	0.715	−0.178	0.835	0.417	0.000	1.982
6	PEG, GLC[c]	0.560	0.980	5.686	9.450	0.000	2.001
7	TEG, GLC[d]	−2.731	0.208	1.882	4.636	0.313	0.584
8	DBS, GLC[d]	−2.034	0.000	0.968	1.911	0.045	0.765
9	Water[e]	−1.271	0.822	2.743	3.904	4.814	−0.213
10	Octanol[f]	−0.222	0.088	0.701	3.478	1.477	0.851
11	Hexadecane[g]	0	0	0	0	0	1.000
B	**Water to solvent, eqn. (6)**						
12	Gas phase[h]	−0.994	0.577	2.549	3.813	4.841	−0.869
13	Octanol[i]	0.088	0.562	−1.054	0.034	−3.460	3.814
14	Hexadecane[i]	0.087	0.667	−1.617	−3.587	−4.869	4.433
C	**HPLC, eqn. (6)**						
15	Aq. MeOH[j]	−0.308	0.317	−0.543	−0.220	−1.251	1.371
16	Aq. MeCN[j]	−0.174	0.179	−0.329	−0.415	−1.107	1.056

[a] Equations for $I/100$, this work. [b] Equation for $I/100$ from ref. 12. [c] Equation for $I/100$ from ref. 13. [d] Equations for log t_{rel}, this work. [e] Log L^W at 298 K, ref. 15. [f] Equation for log L into wet octanol at 298 K, ref. 19. [g] Equation for log L into hexadecane at 298 K, by definition. [h] This is an equation for log L^W at 298 K, ref. 15, with Vx as the final descriptor. [i] Equation for log P, ref. 16. [j] Equations for log k', this work.

values, it is possible to derive log L^{16} values for new compounds, because the R_2 descriptor can be obtained either from the liquid refractive index or by simple summation of fragment values. Then if a stationary phase that is polar (and hence almost certainly basic) is selected, eqn. (1) will reduce to eqn. (5), so that

$$\log SP = c + rR_2 + s\pi_2^H + a\Sigma a_2^H + l \log L^{16} \quad (5)$$

for any new compound that has no hydrogen-bond acidity, the descriptor π_2^H can now be found,[10,11] knowing R_2 and log L^{16}. If both π_2^H and Σa_2^H have to be obtained, then other equations are needed—either eqn. (5) for another stationary phase that gives rise to very different coefficients, or an equation for another process altogether. We have used a large selection of compounds that have been chromatographed on three phases, OV1, SE54 and PEG, to construct equations based on eqn. (1) where log SP is the Kovats retention index, I; details are in Table 1. We also used equations we have calculated from GLC retention indices, I, given by Wang *et al.*[12,13] on two OV-101 stationary phases (A and B), and on a polar PEG phase (C), see Table 1. In order to effect a suitable weighting of equations with I as the dependent variable in eqn. (1), with equations in which log L is the dependent variable, we have scaled all retention indices by a factor of 100. We also measured retention data for a number of compounds on two new stationary phases, tetraethylene glycol (TEG) and dibutyl suberate (DBS), which are intermediate in dipolarity and hydrogen-bond basicity between OV1/SE54 and PEG. The data on TEG and DBS are in terms of log (t_{rel}) where t_{rel} is the relative retention time; no weighting is necessary.

It is useful to examine the coefficients in eqn. (1) for the GLC phases, to check that they are in accord with general chemical principles. Coefficients obtained with the Kovats index as the dependent variable are not scaled in the same way as those obtained using log (t_{rel}), and must be multiplied by the B-coefficient in order to be correctly scaled.[14] The latter is the factor for variation of log (t_{rel}) or log Vg with carbon number for alkanes; we take 0.2 as a rough approximation for the B-coefficient.[14] OV1, SE54 and OV101 are generally regarded as rather non-polar GLC phases, but they contain Si–O–Si groups that are somewhat dipolar and weakly basic. Hence the solvation equations contain small s-coefficients (0.15 to 0.29 after the scale change) and small a-coefficients (0.14 to 0.20). PEG is a much more dipolar and basic phase with s-coefficients near

to 1.1 and b-coefficients around 1.8, very close to those found before for other samples of PEG.[15] Note that the c-constant for process 3 differs from that for process 5 because methyl esters were used as the standards in process 3. TEG and DBS are not standard GLC phases at all, and we hope to report on these in detail later. The coefficients for TEG show that it is quite dipolar and basic, as expected, but that it is only weakly acid with a b-coefficient of only 0.31. Presumably, the terminal -CH$_2$OH groups are involved in intra- or inter-molecular hydrogen bonding that greatly reduces their ability to act as hydrogen-bond acids to external hydrogen-bond bases. Exactly the same effect has been observed for docosanol and diglycerol stationary phases.[16] The coefficients for DBS are as expected for a long-chain ester which is somewhat dipolar and basic.

A very useful property is the solute L-value in water, L^W, obtained either directly or from the solubility of the solute in water, C_W, combined with the solute saturated vapor pressure (that is equivalent to the solute concentration in the gas phase, C_G), through $L^W = C_W/C_G$. Solute descriptors can also be obtained from water–solvent partition coefficients, as log P; the latter are correlated through our second general solvation equation (6),[9] where the McGowan volume, Vx,[17] is used

$$\log SP = c + rR_2 + s\pi_2^H + a\Sigma a_2^H + b\Sigma\beta_2^H + vVx \quad (6)$$

instead of log L^{16}. Since Vx can be calculated from structure,[9,17] the use of eqn. (6) introduces no extra descriptor that has to be determined. If log P for a solute in a given water–solvent system and log L^W for that solute are known, then log L for the solute in the solvent can be deduced from log P = log L − log L^W.

The reason for setting out the above connections is that both eqn. (1) and eqn. (6) have been used to correlate log L^W values,[18] eqn. (6) has been applied to various log P values,[19] and we have recently set out eqn. (1) for log L values in several solvents.[20–22] In Table 1 are given the coefficients in eqn. (1) and eqn. (6) for a number of these processes.

The final type of system that we used is high performance liquid chromatography, HPLC, in the usual reverse phase mode. We,[23,24] and others,[25,26] have shown that eqn. (6) satisfactorily correlates HPLC capacity factors, as log k', and so it is possible to use such correlative equations to determine descriptors. We used a C-18 stationary phase with 70% aqueous methanol and

Table 2 Statistics for the correlation equations determined in this work (Table 1)

Correlation	n	r	sd	F
A Gas to solvent, eqn. (1)				
1 OV1, GLC	147	0.9982	0.214	13224
2 SE54, GLC	150	0.9991	0.154	20359
3 PEG, GLC	142	0.9983	0.221	9905
4 OV101, GLC	18	0.9999	0.067	13244
5 OV101, GLC	16	0.9999	0.047	22725
6 PEG, GLC	14	0.9990	0.204	1083
7 TEG, GLC	56	0.9953	0.065	1062
8 DBS, GLC	73	0.9969	0.050	2976
9 Water	392	0.9962	0.185	10229
10 Octanol	468	0.9946	0.201	8520
B Water to solvent, eqn. (6)				
12 Gas phase	408	0.9976	0.151	16810
13 Octanol	613	0.9974	0.116	23162
14 Hexadecane	370	0.9982	0.124	20236
C HPLC, eqn. (6)				
15 Aq. MeOH[a]	33	0.9869	0.087	248
16 Aq. MeCN[a]	36	0.9844	0.075	226

[a] Equations for log k', this work.

60% aqueous acetonitrile as eluents. The correlation equations are summarized in Tables 1 and 2.

Experimental

HPLC

The HPLC capacity factors, k', were measured on a 25 cm × 4.0 mm id Merck Lichrospher 100 RP-18 (5 µm) column, thermostatted at 298 K with a water jacket. The chromatographic equipment consisted of an ISCO (Lincoln, NE) model 2350 dual-pump system with a 20 ml loop valve and either a variable wavelength absorbance detector (ISCO) or a refractive index detector (Shimadzu). The retention times of the solutes were obtained as the average of three independent injections. Potassium bromide was used as the holdup marker, and the mobile phase flow rate was 1 ml min^{-1}.

Mobile phases were prepared from methanol or acetonitrile (HPLC Ultra Gradient Grades from Baker) and water purified by the Milli-Q plus system (Millipore). The aqueous phase for basic solutes used as standards was 0.005 M KH$_2$PO$_4$–0.005 M Na$_2$HPO$_4$ buffered to pH 7. For acidic solutes used as standards the aqueous phase was 0.1 M acetic acid at pH 2.87; the mobile phases were filtered and degassed under helium immediately prior to use.

GLC

The GLC experiments were all carried out by Quest International, Ashford, Kent. Temperature programmed retention indices were obtained on the OV1 and SE54 phases with n-alkane standards in the usual way. For the PEG phase, temperature programmed retention indices were calculated with methyl esters as the standards, taking methyl decanoate as 1000, methyl dodecanoate as 1200, *etc.*

Results and discussion

As pointed out above, in eqn. (1) and eqn. (6) there are two descriptors, R_2 and Vx, that can be calculated for any compound, and four descriptors that remain to be determined, *viz.* π_2^H, Σa_2^H, $\Sigma \beta_2^H$ and log L^{16}. In theory, if the dependent variable for a given compound is known for four different processes that yield four different equations, then it should be possible to solve the set of equations for the four unknowns, *i.e.* the four descriptors. In practice, this is only possible if the coefficients in the

Table 3 The dependent variables for processes in Table 1 for the solute α-pinene; calculation of descriptors

No.	log SP(obs)	log SP(calc)[a]
1	9.31	9.43
2	9.38	9.25
3	4.47	4.50
7	0.17	0.18
8	1.39	1.40
9	−0.84	−0.86
10	3.83	3.77
12	−0.84	−0.89
13	4.67	4.57
14	5.15	5.15
n:	10	
sd:	0.073	

[a] With fixed descriptors $R_2 = 0.446$, $\Sigma a_2^H = 0$ and $Vx = 1.2574$, and calculated descriptors $\pi_2^H = 0.14$, $\Sigma \beta_2^H = 0.12$ and log $L^{16} = 4.3078$.

four equations differ considerably. Our method is to set up as many equations as we can, and to determine the values of π_2^H, Σa_2^H, $\Sigma \beta_2^H$ and log L^{16} that best reproduce the dependent variables.

There are enough data for eleven terpenes to carry out a full analysis on these lines. We use the case of α-pinene as an example of a compound where Σa_2^H is zero and so only three descriptors need to be determined. We have our own GLC data for α-pinene in systems 1, 2, 3, 6 and 7, as shown in Table 3. The solubility in water has been determined by Lebosse *et al.*[27] as log $C_W = -4.45$ with C_W in mol dm^{-1}; log C_G is -3.61 in the same units,[28] thus giving log L^W as -0.84. No experimental log P values are available, but we can calculate a water–octanol log P_{oct} value as 4.67 using the CLOGP programme of Leo.[29] Then from log $L^W = -0.84$ and log $P_{oct} = 4.67$, a value of 3.83 is deduced for log L^{oct}. Finally, we can use the GLC data on non-polar phases to obtain log L^{16}, and then from log $P_{16} =$ log $L^{16} -$ log L^W we can obtain the water–hexadecane partition coefficient as 5.15 for log P_{16}. All these values are in Table 3. We therefore have ten equations for which all the coefficients are known, and for which the dependent variables, as log SP, are all known. The only unknowns are the three descriptors π_2^H, $\Sigma \beta_2^H$ and log L^{16} (Σa_2^H is zero). We can then solve the set of ten equations for the best fit values of π_2^H, $\Sigma \beta_2^H$ and log L^{16}, that is the values that lead to the smallest standard deviation between log SP(calc.) and log SP(obs.). With fixed values of $R_2 = 0.446$, $\Sigma a_2^H = 0$ and $Vx = 1.2574$ in units of (mol cm^{-1})/100 we find that values of 0.14 for π_2^H and 0.12 for $\Sigma \beta_2^H$ reproduce the ten observed dependent variables with an sd value of 0.074 log units, see Table 3.

It might be suggested that we have used an estimated value of log P_{oct} and hence an estimated value of log L^{oct}. However, with a set of ten dependent variables, it is reasonably easy to see if the set is self-consistent.

Our second worked example is that of menthol, for which four descriptors now have to be determined. As before, we use the three systems (1, 2 and 3) studied by Quest International to obtain a value of 5.177 for log L^{16}. We can combine the solubility in water (log $C_W = -2.57$)[30,31] with the saturated vapor pressure at 298 K (log $C_G = -5.46$)[32] to give log L^W as 2.89, and hence log P_{16} as 2.29 (5.18 − 2.89). An experimental value of 3.31 is available[29] for log P_{oct}, so that log L^{oct} is 6.20 (3.31 + 2.89). A summary of these data is in Table 4, together with calculated values using the 'best-fit' descriptors, *viz.* $\pi_2^H = 0.50$, $\Sigma a_2^H = 0.23$, $\Sigma \beta_2^H = 0.58$ and log $L^{16} = 5.177$ units. Calculations on the above lines can be carried out for nine other terpenes, making eleven in all; results are summarized in Table 5.

If only a few equations have been used in the determination of descriptors, it is difficult to assess the individual errors in the descriptors. This is partly because the descriptors are not calculated independently and partly because the errors in the input

Table 4 The dependent variables for processes in Table 1 for the solute menthol; calculation of descriptors

No.	log SP(obs)	log SP(calc)[a]
1	11.55	11.59
2	11.79	11.76
3	10.30	10.28
9	2.89	3.02
10	6.20	6.24
12	2.89	2.91
13	3.31	3.38
14	2.29	2.40
n:		8
sd:		0.075

[a] The fixed descriptors were $R_2 = 0.400$ and $Vx = 1.4677$, calculated descriptors were $\pi_2^H = 0.50$, $\Sigma a_2^H = 0.23$, $\Sigma \beta_2^H = 0.58$ and $\log L^{16} = 5.1774$.

Table 5 Calculated descriptors for terpenes

Terpene	R_2	π_2^H	Σa_2^H	$\Sigma \beta_2^H$	$\log L^{16}$	Vx
α-Pinene[a]	0.446	0.14	0.00	0.12	4.308	1.2574
β-Pinene[b]	0.530	0.24	0.00	0.19	4.394	1.2574
Limonene[c]	0.488	0.28	0.00	0.21	4.725	1.3230
Carvone[d]	0.674	0.86	0.00	0.57	5.445	1.3390
Menthone[e]	0.322	0.61	0.00	0.62	5.102	1.4247
Camphor[e]	0.500	0.69	0.00	0.71	5.084	1.3161
Neral[e]	0.589	0.80	0.00	0.52	5.441	1.4473
Geranial[e]	0.610	0.83	0.00	0.52	5.588	1.4473
Menthol[a]	0.400	0.50	0.23	0.58	5.177	1.4677
Borneol[f]	0.510	0.52	0.28	0.68	5.129	1.3591
1,8-Cineole[g]	0.383	0.33	0.00	0.76	4.688	1.4250
p-Cymene[h]	0.607	0.49	0.00	0.19	4.590	1.2800

[a] Tables 3, 4 and 5. [b] Using systems 1, 3, 7–10, 12–14. [c] Using systems 2–7, 9, 10, 12–14. [d] Using systems 1–3, 9, 12. [e] Using systems 1–3, 9, 12, 14. [f] Using systems 1–3, 9, 10, 12–14. [g] Using systems 1, 3, 7–10, 12, 13. [h] 4-Isopropyltoluene; data from ref. 15.

data (*i.e.* the various log *L*, log *P* and GLC data) are not known. However, when a reasonable number of equations are available for any given terpene, we can apply the method known as 'leave-one-out',[33] which is a particular case of cross-validation.[34] We take α-pinene, Table 3, and menthol, Table 4, as examples. For α-pinene we have 10 equations that are solved to yield the best-fit descriptors shown in Table 3. We can repeat the calculations leaving out the first equation (no. 1 in Table 3) and obtain best-fit descriptors from the remaining 9 processes. We then leave out the second equation (no. 2 in Table 3) and obtain another set of descriptors. This is repeated, leaving out each equation in turn, until we have calculated ten 'leave-one-out' sets of descriptors. These sets can be analyzed to obtain the mean value, the standard deviation, sd, and the 95% confidence limits, CI, for each descriptor. Results for α-pinene and menthol are in Table 6. For α-pinene the mean values for π_2^H, $\Sigma \beta_2^H$ and $\log L^{16}$ are exactly the same as those from the entire set of equations, and the sd values are 0.013, 0.007 and 0.026 respectively. In the case of menthol, the mean values differ only marginally from those calculated using the entire set, or the rounded-off values from the entire set. The sd values for the descriptors are somewhat higher than those for α-pinene; 0.039 (π_2^H), 0.013 (Σa_2^H), 0.020 ($\Sigma \beta_2^H$) and 0.022 (log L^{16}). If we take α-pinene and menthol as representative solutes, then an estimate of 0.03 can be suggested as the general error in π_2^H, Σa_2^H and $\Sigma \beta_2^H$, and 0.025 as the general error in log L^{16} when the descriptors have been obtained by the present method.

The determined values for descriptors seem reasonable by comparison to those for solutes with the same functionalities. The cyclic mono-alkene, α-pinene, has π_2^H and $\Sigma \beta_2^H$ values quite close to those for 1-methylcyclohexene (0.18 and 0.10 respectively). β-Pinene with a more open C=C bond has rather higher values, and limonene with two C=C bonds has π_2^H and

Table 6 Statistical analysis of the determination of descriptors for α-pinene and menthol, using the leave-one-out method

Descriptor	Leave-one-out				All data used	
	Mean	sd	95% CL		Mean[a]	Mean[b]
A α-Pinene						
π_2^H	0.142	0.013	0.132	0.152	0.140	0.14
$\Sigma \beta_2^H$	0.120	0.007	0.115	0.125	0.120	0.12
$\log L^{16}$	4.308	0.026	4.289	4.326	4.308	4.308
B Menthol						
π_2^H	0.489	0.039	0.457	0.522	0.502	0.50
Σa_2^H	0.224	0.013	0.213	0.235	0.225	0.23
$\Sigma \beta_2^H$	0.589	0.020	0.572	0.605	0.583	0.58
$\log L^{16}$	5.190	0.022	5.172	5.209	5.184	5.177[c]

[a] Using all the 10 data points in Table 3, with no rounding-off. [b] Using all the 10 data points in Table 3, with values of π_2^H, Σa_2^H and $\Sigma \beta_2^H$ rounded-off to two decimal places. [c] This differs from 5.184 because when π_2^H, Σa_2^H and $\Sigma \beta_2^H$ are fixed at their rounded-off values, 5.177 leads to the minimum overall sd.

$\Sigma \beta_2^H$ about twice as large as those for for α-pinene. The three cyclic ketones carvone, menthone and camphor all have quite large π_2^H and $\Sigma \beta_2^H$ values, compare cyclohexanone (0.86 and 0.56 respectively), although we would have expected carvone to have the largest $\Sigma \beta_2^H$ value. Geranial and neral are acyclic aldehydes with two C=C bonds and have π_2^H and $\Sigma \beta_2^H$ values close to those calculated using simple aldehydes and alkenes as models. Both menthol and borneol are derivatives of cyclohexanol, although borneol is bicyclic. For both terpenes, the determined descriptors are not far from those for cyclohexanol ($\pi_2^H = 0.54$, $\Sigma a_2^H = 0.32$ and $\Sigma \beta_2^H = 0.57$), although Σa_2^H is somewhat smaller than the cyclohexanol value, being 0.23 for menthol and 0.28 for borneol. For α-pinene, 1,8-cineole, carvone and α-terpineol additional GLC data are available from Laffort *et al.*[35] We did not use these data in the determination of descriptors, but checked that the descriptors we obtained were compatible with Laffort's retention data.

In the determination of descriptors for solutes in general, values of log P_{oct} are important, in that they can be used to determine $\Sigma \beta_2^H$ values; note that in any equation the larger the coefficient of a given term, the smaller will be the error in the determined descriptor. Since there are now available several methods for the calculation of log P_{oct}, we thought it useful to ascertain if the calculated values could be used to assign $\Sigma \beta_2^H$ values for terpenes. In addition, we examined a number of terpenes by our HPLC procedure, again to test if equations for systems 15 and 16 could be used in the assignment of $\Sigma \beta_2^H$ values. We used the CLOGP programme, as above, and also the PROLOGP method[36] to calculate log P_{oct}. Then knowing the other descriptors, we used these log P_{oct} values to estimate $\Sigma \beta_2^H$. Log k' values obtained in our HPLC systems were used similarly. Results are in Table 7.

There are six terpenes for which we have assigned $\Sigma \beta_2^H$ values (Table 5) and for which the CLOGP and the PROLOGP algorithms were not used in the determination of descriptors. By comparison with the results in Table 5, the CLOGP algorithm often leads to overestimates of $\Sigma \beta_2^H$ for these six compounds. The PROLOGP and the HPLC method lead to reasonable estimates of $\Sigma \beta_2^H$ in most cases. For the nine terpenes not listed in Table 5, the CLOGP algorithm again leads to a higher calculated value for $\Sigma \beta_2^H$ than do the PROLOGP and the HPLC methods. Our conclusion is that it is unwise to rely on one method, and that as many methods as possible should be used in the determination of $\Sigma \beta_2^H$, as above, and in the determination of descriptors generally.

As well as the eleven terpenes in Table 5, GLC retention data were obtained by Quest International for a large number of other terpenes. From these data, it was possible to obtain values

of $\pi_2{}^{\mathrm{H}}$ and $\Sigma a_2{}^{\mathrm{H}}$ from systems 1–3, but it was not possible to obtain $\Sigma\beta_2{}^{\mathrm{H}}$ values. In the absence of other experimental data, and taking into account our estimates of $\Sigma\beta_2{}^{\mathrm{H}}$ through the CLOGP and PROLOGP algorithms, Table 7, we estimated $\Sigma\beta_2{}^{\mathrm{H}}$ by comparison to values for the terpenes in Table 5, and by comparison to values for other compounds. The assigned descriptors for the additional terpenes are in Table 8.

The descriptors we have assigned to the terpenes, Tables 5 and 8, are not unusual, and fall within the range of values for the solutes used to set up the correlation equations given in Table 1. Hence the present method amounts to an interpolation procedure and not to an extrapolation procedure. This is one of the reasons why the sd values for the descriptors found by the 'leave-one-out' method are so small.

It should be noted that all the processes used to assign descriptors, see Table 1, involve non-chiral systems. Hence any determined value of log (t_{rel}) in one of the GLC systems, or any determined value of log P in one of the water–solvent systems will be the same for enantiomers of a given terpene. Our descriptors for α-pinene, as an example, will be exactly the same for D-α-pinene and L-α-pinene. There are no reports of nasal pungency thresholds in man for enantiomers of a given compound, but Kasanen et al.[37] have determined concentrations in ppm that produce a 50% decrease in respiratory rate in mice, RD_{50}, for enantiomers of α-pinene and β-pinene. The RD_{50} values are D-α-pinene (1053), L-α-pinene (inactive), D-β-pinene (1279) and L-β-pinene (4663). The values for α-pinene are not very useful in the present context. However, the values for β-pinene expressed as log units, 3.11 and 3.67, differ only by 0.56 log units so that enantiomeric differences in this particular case are not very large.

Once descriptors have been assigned to any given terpene, then it is trivial to obtain missing observations for the processes given in Table 1. Thus log P for the water–hexadecane partition of 1,8-cineole is calculated as 2.51 by combining the descriptor values with the coefficients for process 14. In addition to the processes set out in Table 1, coefficients are known for many more water–solvent partitions,[38] for various gas–solvent partitions,[20,21,39] and for a number of biological processes.[38] Values for all these can be predicted with the sets of descriptors given in Tables 5 and 8. We deal with nasal pungency thresholds in more detail, below.

Estimation of nasal pungency thresholds

We have all the descriptors in eqn. (3) for a variety of terpenes, so that it is then possible to calculate log (1/NPT) values. The corresponding NPT values themselves are in Table 9 for terpenes that are commonly found in indoor air. To give an idea of the range of NPT values observed, we include also some determined values[8] for a number of other compounds. The smaller is the NPT value, the more potent is the VOC; note that we use log (1/NPT) in eqn. (3) in order to derive a measure that becomes larger as the VOC becomes more potent.

The relative significance of the descriptors in eqn. (3) can be calculated from the product of the coefficient and the mean value of the corresponding descriptor. The percentage weights are then $\pi_2{}^{\mathrm{H}}$ (23), $\Sigma a_2{}^{\mathrm{H}}$ (10), $\Sigma\beta_2{}^{\mathrm{H}}$ (12) and log L^{16} (55) so that the l log L^{16} term is by far the most important. The latter is related to the size of the solute, and very roughly to the molecular weight. Hence the lower molecular weight terpene hydrocarbons are predicted to have small values of log (1/NPT) and hence large NPT values. The large and very lipophilic terpene

Table 7 Estimation of $\Sigma\beta_2{}^{\mathrm{H}}$ values for terpenes, using calculated log P_{oct} values, and HPLC capacity factors

Terpene	$\Sigma\beta_2{}^{\mathrm{H}}$			
	Table 5	CLOGP	PROLOGP	HPLC[a]
α-Pinene	0.12	0.09[b]	0.15	<0.00
β-Pinene	0.19	0.22[b]	0.14	<0.00
Limonene	0.21	0.22[b]	0.19	<0.00
α-Terpinene	—	0.21	0.11	
γ-Terpinene	—	0.21	0.17	
α-Phellandrene	—	0.21	0.11	
Camphene	—	0.21	0.13	
Δ-3-Carene	—	0.22	0.00	
Carvone	0.57	0.76	0.61	
Menthone	0.62	0.64	0.60	0.42
Camphor	0.71	0.71[b]	0.58	0.57
Neral	0.52	0.76	0.53	0.60
Geranial	0.52	0.77	0.53	0.58
Menthol	0.58	0.62	0.54	0.50
Borneol	0.68	0.78	0.54	0.55
Linalool	—	0.99	0.66	0.64
Geraniol	—	0.92	0.65	0.63
Nerol	—	0.77	0.53	0.60
α-Terpineol	—	0.68	0.54	0.58

[a] Average from systems 15 and 16, Table 1. [b] Used in the determination of the descriptors.

Table 8 Descriptors for terpenes from GLC data[a]

Terpene	R_2	$\pi_2{}^{\mathrm{H}}$	$\Sigma a_2{}^{\mathrm{H}}$	$\Sigma\beta_2{}^{\mathrm{H}}$	log L^{16}	Phases[b]
α-Terpinene	0.526	0.25	0.00	(0.15)	4.715	5
γ-Terpinene	0.497	0.32	0.00	(0.20)	4.815	5
α-Phellandrene	0.522	0.24	0.00	(0.15)	4.607	4
β-Phellandrene	0.529	0.25	0.00	(0.15)	4.703	4
Terpinolene	0.593	0.31	0.00	(0.20)	4.998	4
Camphene	0.424	0.22	0.00	(0.15)	4.326	3
Δ-3-Carene	0.511	0.22	0.00	(0.10)	4.649	4
Myrcene	0.483	0.29	0.00	(0.21)	4.509	3
Longifolene	0.765	0.20	0.00	(0.15)	6.575	2[c]
α-Cedrene	0.759	0.23	0.00	(0.12)	6.688	2[c]
trans-β-Caryophyllene	0.724	0.15	0.00	(0.25)	6.874	5[c]
Linalool	0.398	0.55	0.20	(0.67)	4.794	3
Geraniol	0.513	0.63	0.29	(0.66)	5.479	3
Nerol	0.498	0.61	0.27	(0.66)	5.370	3
α-Terpineol	0.553	0.61	0.20	(0.70)	5.229	3
Menthyl acetate	0.234	0.54	0.00	(0.57)	5.830	3
Bornyl acetate	0.397	0.54	0.00	(0.58)	5.812	3
Linalyl acetate	0.311	0.60	0.00	(0.55)	5.613	3
Geranyl acetate	0.368	0.72	0.00	(0.65)	6.167	3
Neryl acetate	0.350	0.70	0.00	(0.65)	6.088	3
α-Terpinyl acetate	0.447	0.62	0.00	(0.64)	6.099	3

[a] Values for $\Sigma\beta_2{}^{\mathrm{H}}$ estimated, see text. [b] Number of GLC phases used; systems 1–3 are always included. [c] System 2 not used.

Table 9 Predicted values of nasal pungency thresholds for terpenes

Terpene	NPT (ppm)	Terpene	NPT (ppm)
α-Pinene	22 000	Carvone	15
β-Pinene	9000	Menthone	89
Limonene	3600	Camphor	47
α-Terpinene	5200	Neral	24
γ-Terpinene	2600	Geranial	16
α-Phellandrene	6800	1,8-Cineol	510
β-Phellandrene	5300	p-Cymene	1800
Terpinolene	1880	Menthyl acetate	35
Camphene	13 000	Bornyl acetate	35
Δ-3-Carene	8000	Linalyl acetate	43
Myrcene	5300	Geranyl acetate	5.7
Longifolene	170	Neryl acetate	7.4
α-Cedrene	130	α-Terpinyl acetate	11
trans-β-Caryophyllene	85	Toluene	29 500[a]
Menthol	23	Pyridine	1300[a]
Borneol	11	Heptanal	1300[a]
Linalool	37	Heptan-2-one	810[a]
Geraniol	3.2	Formic acid	320
Nerol	5.2	Heptan-1-ol	210[a]
α-Terpineol	11	Octanoic acid	2.0[a]

[a] Observed value, ref. 8.

Table 10 The effect of size on NPT values for hydrocarbons

Terpene	NPT	Log L^{16}	MW[a]
α-Pinene	22 000	4.308	136.2
Camphene	13 000	4.326	136.2
β-Pinene	9000	4.394	136.2
Δ-3-Carene	8000	4.649	136.2
α-Phellandrene	6800	4.607	136.2
Myrcene	5300	4.509	136.2
β-Phellandrene	5300	4.703	136.2
α-Terpinene	5200	4.715	136.2
Limonene	3600	4.725	136.2
γ-Terpinene	2600	4.815	136.2
Terpinolene	1880	4.998	136.2
Longifolene	170	6.575	204.4
α-Cedrene	130	6.688	204.4
trans-β-Caryophyllene	85	6.874	204.4

[a] Terpene molecular weight.

hydrocarbons such as longifolene, α-cedrene and *trans*-β-caryophyllene are predicted to have larger values of log (1/NPT) and hence moderate NPT values. Our analysis shows that the greater the value of log L^{16} for a given functionally substituted compound, the larger will be log (1/NPT), the smaller will be the NPT value and the more potent will be the VOC. This is shown in Table 10, where the terpene hydrocarbons are listed in order of increasing potency. There is a very good correspondence with log L^{16}, illustrating the importance of this descriptor in eqn. (3). The terpene aldehydes and ketones are quite potent, and the terpene alcohols and acetates are predicted to have very small nasal pungency thresholds.

Quite recently, we have determined NPT values for a few terpenes,[40] as given in Table 11. We were unable to assign error limits for α-terpinene because NPT values were obtained at a vapor concentration that corresponded to the saturated vapor concentration. In Table 1 and in Fig. 1 we have assigned an sd value of 0.22 units to log (NPT) for α-terpinene, as an average of the sd values for the other five terpenes in Table 11. For the sd values of the predicted log (NPT) values, we have taken the regression sd value in eqn. (3). In Fig. 1 are shown the observed and predicted values of log (NPT). Within the error limits of the log (NPT) determinations and the error in our predictions there is reasonable agreement. This is an important result. The number of terpenes that could be present in indoor air is so large that it is simply not possible to determine NPT values on human subjects for more than a small fraction. Some

Table 11 Observed and predicted values of log NPT

Terpene	Observed log NPT	sd	Predicted log NPT	sd[a]
Menthol	1.71	0.62	1.37	0.27
1,8-Cineole	2.37	0.30	2.71	0.27
Linalool	2.55	0.06	1.57	0.27
p-Cymene	3.05	0.08	3.25	0.27
Δ-3-Carene	3.21	0.02	3.91	0.27
α-Terpinene	3.30	0.22[b]	3.71	0.27

[a] Taken as the regression sd value in eqn. (3). [b] No error limits were assigned;[40] we have taken an sd value of 0.22 as the average of the sd values for the other five terpenes.
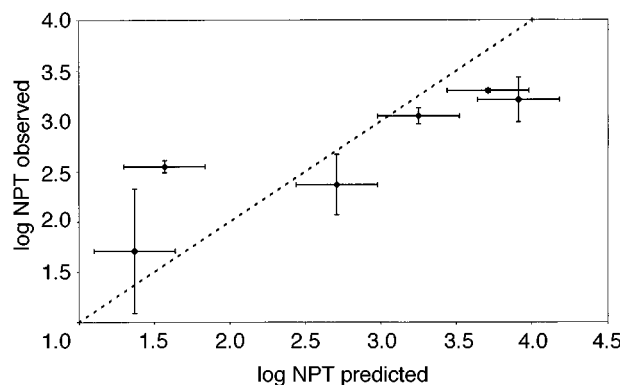
Fig. 1 Observed[40] and estimated values of human nasal pungency thresholds log NPT for the terpenes: menthol, linalool, 1,8-cineole, p-cymene, α-terpinene and Δ-3-carene. The dotted line represents the line of identity; the correlation coefficient is 0.88. Bars indicate standard deviations.

method of prediction of NPT values must be devized; we believe that eqn. (3) can be used to obtain useful predictions.

The difficulty of eqn. (3) is that descriptors have to be assigned to the various terpenes. However, we have now shown that the determination of descriptors is aided through the potential use of a large number of different physicochemical processes. Not all the processes need to be used in any given case, but their availability means that it is now possible to assign descriptors to terpenes, and other compounds, much more easily than hitherto. It must be pointed out that the present method is feasible only because the same four descriptors (R_2, π_2^H, $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$) are used in eqn. (1) and eqn. (6). If a different set of descriptors is used for gas-to-condensed phase processes than for condensed phase-to-condensed phase processes, each type of process has to be dealt with separately, and the number of descriptors that have to be assigned becomes too large to be manageable.

Acknowledgements

References

1 P. T. C. Harrison, *Chem. Ind.*, 1997, 677.
2 S. K. Brown, M. R. Sim, M. J. Abramson and C. N. Gray, *Indoor Air*, 1994, **4**, 123.
3 Evaluation of VOC Emissions from Building Products, Report No. 18, European Collaborative Action, Indoor Air Quality and its Impact on Man, ECSC-EC-EAEC, Brussels, 1997.

4 H. Rothweiler and C. Schlatter, *Toxicol. Environ. Chem.*, 1993, **40**, 93.
5 H. Knoppel and H. Schauenburg, *Environ. Inst.*, 1989, **15**, 413.
6 Y. Alarie, *Environ. Health Perspect.*, 1981, **42**, 9.
7 B. Berglund, U. Berglund and R. Lindvall, Proceedings of the ASHRAE Conference IAQ '86, Atlanta, Georgia, USA, pp. 138–149, 1986.
8 M. H. Abraham, R. Kumarsingh, J. E. Cometto-Muniz and W. S. Cain, *Arch. Toxicol.*, 1998, **72**, 227.
9 M. H. Abraham, *Chem. Soc. Rev.*, 1993, **22**, 73.
10 M. H. Abraham and G. S. Whiting, *J. Chromatogr.*, 1992, **594**, 229.
11 M. H. Abraham, *J. Chromatogr.*, 1993, **644**, 95.
12 T. Wang, B. Zhong, M. Chen and Y. Sun, *J. Chromatogr.*, 1987, **390**, 275.
13 T. Wang and Y. Sun, *J. High Resolut. Chromatogr.*, 1987, **10**, 603.
14 M. H. Abraham, G. S. Whiting, R. M. Doherty and W. J. Shuely, *J. Chem. Soc., Perkin Trans. 2*, 1990, 1451.
15 M. H. Abraham, G. S. Whiting, R. M. Doherty and W. J. Shuely, *J. Chromatogr.*, 1991, **587**, 213.
16 M. H. Abraham, G. S. Whiting, R. M. Doherty and W. J. Shuely, *J. Chromatogr.*, 1990, **518**, 329.
17 M. H. Abraham and J. C. McGowan, *Chromatographia*, 1987, **23**, 243.
18 M. H. Abraham, J. Andonian-Haftvan, G. S. Whiting, A. Leo and R. W. Taft, *J. Chem. Soc., Perkin Trans. 2*, 1994, 1777.
19 M. H. Abraham, H. S. Chadha, G. S. Whiting and R. C. Mitchell, *J. Pharm. Sci.*, 1994, **83**, 1085.
20 M. H. Abraham, G. S. Whiting, P. W. Carr and H. Ouyang, *J. Chem. Soc., Perkin Trans. 2*, 1998, 1385.
21 M. H. Abraham, G. S. Whiting, W. J. Shuely and R. M. Doherty, *Can. J. Chem.*, in the press.
22 M. H. Abraham, unpublished work.
23 M. H. Abraham and M. Roses, *J. Phys. Org. Chem.*, 1994, **7**, 672.
24 M. H. Abraham, M. Roses, C. F. Poole and S. K. Poole, *J. Phys. Org. Chem.*, 1997, **10**, 358.
25 D. S. Seibert and C. F. Poole, *J. High Resolut. Chromatogr.*, 1995, **18**, 26.
26 P. T. Jackson, M. R. Schure, T. P. Weber and P. W. Carr, *Anal. Chem.*, 1997, **69**, 416.
27 R. Lebosse, A. Feigenbaum and V. Ducret, *J. High Resolut. Chromatogr.*, 1996, **19**, 413.
28 M. F. Rodrigues and M. G. Bernardo-Gil, *J. Chem. Eng. Data*, 1995, **40**, 1193.
29 A. Leo, The Medicinal Chemistry Project, Pomona College, Claremont, CA 91711; *Chem. Rev.*, 1993, **93**, 1281.
30 J. E. Amoore and R. G. Buttery, *Chem. Senses Flavour*, 1978, **3**, 57.
31 A. Seidell, *Solubilities of Organic Compounds*, 3rd edn., Van Nostrand, New York, 1941; A. Seidell and W. F. Linke, *ibid.*, 3rd edn. supplement, Van Nostrand, New York, 1952.
32 R. M. Stephenson and S. Malinowski, *Handbook of the Thermodynamics of Organic Compounds*, Elsevier, New York, 1987.
33 H. van de Waterbeemd, *Drug Design Discovery*, 1993, **9**, 277.
34 R. D. Cramer III, J. D. Bunce and D. E. Patterson, *Quant. Struct. Act. Relat.*, 1988, **7**, 18.
35 F. Patte, M. Etcheto and P. Laffort, *Anal. Chem.*, 1982, **54**, 2239.
36 PROLOGP, version 4.2 from CompuDrug Chemistry Ltd., Budapest, Hungary, 1992.
37 J. P. Kasanen, A. L. Pasanen, P. Pasanen, J. Liesivuory, V. M. Kosma and Y. Alarie, *Arch. Toxicol.*, 1998, **72**, 514.
38 M. H. Abraham and H. S. Chadha, in *Lipophilicity in Drug Action and Toxicology*, ed. V. Pliska, B. Testa and H. van de Waterbeemd, VCH, Weinheim, 1996.
39 M. H. Abraham, J. Andonian-Haftvan, J. P. Osei-Owusu, P. Sakellariou, J. S. Urieta, M. C. Lopez and R. Fuchs, *J. Chem. Soc., Perkin Trans. 2*, 1993, 299.
40 J. E. Cometto-Muniz, W. S. Cain, M. H. Abraham and R. Kumasingh, *Pharmacol. Biochem. Behav.*, 1998, **60**, 765.

Paper 8/05665J