Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 478–485
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Air to lung partition coefficients for volatile organic compounds and blood to lung partition coefficients for volatile organic compounds and drugs

Michael H. Abraham [a,*], Adam Ibrahim [a], William E. Acree Jr. [b]

[a] Department of Chemistry, University College London, 20 Gordon Street, London, Middlesex, WC1H OAJ, UK
[b] Department of Chemistry, P.O. Box 305070, University of North Texas, Denton, TX 76203-5070, USA

Received 14 January 2007; received in revised form 2 April 2007; accepted 3 April 2007
Available online 24 April 2007

Abstract

Values of *in vitro* gas to lung partition coefficients, K_{lung}, of VOCs have been collected from the literature. For 44 VOCs, application of the Abraham solvation equation to log K_{lung} yielded a correlation with $R^2 = 0.968$ and S.D. = 0.25 log units. Combination of the log K_{lung} values with log K_{blood} values leads to *in vitro* blood to lung partition coefficients, log P_{lung} for 43 VOCs; an Abraham solvation equation correlated these values with a very poor $R^2 = 0.262$ but with a very good S.D. = 0.190 log units.

Values of *in vivo* log P_{lung} for 80 drugs were collected, and were correlated with $R^2 = 0.647$ and S.D. = 0.51 log units. When the log P_{lung} values for VOCs and drugs were combined, an Abraham solvation equation could correlate the 123 compounds with $R^2 = 0.676$ and S.D. = 0.43 log units. Division of the 123 compounds into a training set and a test set, showed that the training equation could predict log P_{lung} values with an average error of 0.001 and a standard deviation of 0.44 log units; for drugs in the combined test set the average error was 0.02 and the standard deviation 0.43 log units.

Keywords: Volatile organic compounds; Drugs; Lung; Blood–lung; Plasma–lung; Partition coefficients; Linear free energy relationship; Abraham equation; Hydrogen-bond

1. Introduction

Partition coefficients from air to tissue and from blood to tissue are of importance in environmental, toxicological and pharmacokinetic modelling. There are a number of reports on partition coefficients of volatile organic compounds, VOCs, from air to lung [1–9], but seldom on the correlation of the coefficients, K_{lung}. The air to lung partition coefficient is defined as

$$K_{lung} = \frac{[\text{Conc. of compound in lung, mol kg}^{-1}]}{[\text{Conc. of compound in air, mol dm}^{-1}]} \quad (1)$$

Values of K_{lung} are invariably determined by an *in vitro* method in which the gaseous solubility is determined in isolated samples of lung; nearly all the data refers to human lung.

Abraham and Weathersby [1] used a dataset of 36 VOCs to correlate log K_{lung} values against the five Abraham descriptors, as shown in Eq. (2); we shall deal with these descriptors in detail later.

$$\log K_{lung} = -1.300 + 0.667E + 0.680S + 3.539A + 3.350B + 0.458L \quad (2)$$
$$N = 36,\ R^2 = 0.976,\ \text{S.D.} = 0.233,\ F = n/a$$

In Eq. (2) and elsewhere, N is the number of data points, usually the number of compounds, R is the correlation coefficient, S.D. is the standard deviation and F is the F-statistic.

Combination of the *in vitro* K_{lung} values with *in vitro* gas to blood partition coefficients yields *in vitro* blood to lung

* Corresponding author. Tel.: +44 020 7679 4639; fax: +44 020 7679 7463.
E-mail address: m.h.abraham@ucl.ac.uk (M.H. Abraham).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2007.04.002

partition coefficients, log P_{lung}. Fiserova-Bergerova and Diaz [2] showed that there was a correlation between blood to lung and blood to fat partition coefficients for 30 VOCs, with $R^2 = 0.541$ and S.D. = 0.352. Abraham and Weathersby [1] correlated 36 values of log P_{lung} for VOCs with the Abraham descriptors and found that $R^2 = 0.975$ and S.D. = 0.233 log units. A rather complicated expression was used by Balaz and Lukacova [10] to correlate 28 log P_{lung} values. The water to octanol partition coefficient, log P_{oct}, was used as a descriptor, together with adjustable parameters A_0, A_1, A_2, A_b and β, and indicator variables I_1 for alcohols and I_2 for alkanes. A_0 was not significant, but six adjustable variables were used to correlate values for 28 simple compounds, so that it is not surprising that the statistics for the fitting equation were good, $R^2 = 0.709$, S.D. = 0.073, $F = 30.5$. Zhang [11] also used a complicated expression to fit 29 values of log P_{lung} with $R^2 = 0.653$ and S.D. = 0.098 log units. The small values of S.D. obtained by Balaz and Lukacova [10] and by Zhang [11] must indicate over-fitting of the data, bearing in mind that the error in P_{lung} will be the resultant of the errors in both K_{lung} and K_{blood}. The only predictive statistics given were by Zhang. For a very small test set of six compounds, S.D. is only 0.070 log units.

There is a considerable amount of information on *in vitro* values of log P_{lung} nearly all of which are for human lung as well as *in vivo* values obtained by direct determination on rats [12–43], but few analyses of the data. Zhang and Zhang [33] examined *in vivo* data of drugs for partition between blood and several tissues, and obtained a general training equation that covered 248 data points over seven tissues. For blood to lung there were 37 drugs in the 248 data points, and six drugs in an independent test set. For the given predicted and experimental values of log P_{lung} [33], we calculated S.D. values of 0.62 and 0.60 log units for the test set on the two given equations. Rodgers et al. [34] have determined plasma to tissue distribution for various drugs, and have set out equations for the prediction of plasma water to tissue distribution [35, 36]. We have converted the latter into plasma to tissue distribution, using data given by Rodgers et al. [34–36]; note that the statistics given by Rodgers et al. refer to plasma water to tissue distribution.

There is, therefore, little in the way of predictive statistics for either air to lung or blood to lung distribution. It is our aim of this work to set out a comprehensive list of log K_{lung} and log P_{lung} values, to obtain equations that correlate these values and to assess the predictive capability of the equations.

2. Methods

We use the same methods that we have previously employed [44–49] for partition from air to blood [44], air to brain [45], blood to brain [46], air and blood to muscle [47], air and blood to fat [48], and air and blood to liver [49]. In brief, the linear free energy relationships (LFERs) that we use are given as Eqs. (3) and (4),

$$\log K_{lung} = c + eE + sS + aA + bB + lL \quad (3)$$

$$\log P_{lung} = c + eE + sS + aA + bB + vV \quad (4)$$

Eq. (3) is used to correlate partition from air to lung, and Eq. (4) is used to correlate partition from blood or plasma to lung [50,51]. The independent variables in Eqs. (3) and (4) are properties of VOCs and drugs, as we have discussed before [50,51]. E is the solute excess molar refractivity in units of $(cm^3\,mol^{-1})/10$, S is the solute dipolarity/polarizability, A and B are the overall or summation hydrogen-bond acidity and basicity, L is the logarithm of the gas–hexadecane partition coefficient at 25 °C, and V is the McGowan volume in units of $(cm^3\,mol^{-1})/100$.

The experimental data on log K_{lung} values [1–9] that we use are collected in Table 1 for 45 VOCs; most of the values are from old compilations [1,2,7]; surprisingly there seems to be no recent compilation at all. We have shown previously that for *in vitro* partition from the gas phase into brain, muscle, and liver, values for human and rat tissues are so close that they can be averaged, and in Table 1 we give such averages.

The *in vitro* air to lung log K_{lung}(human/rat) values, Table 1, can be combined with *in vitro* air to blood partition coefficients, log K_{blood}(human/rat), that we have previously reported [44] to give *in vitro* blood to lung distributions for VOCs, log P_{lung}(human/rat). Values of log P_{lung}(human/rat) are given in Table 1 and are indirect *in vitro* values. There are indirect *in vitro* log P_{lung} values for VOCs as well as a large number of direct *in vivo* values, for distribution from blood or plasma or serum to rat lung [12–43]. These are listed in Table 2. As before, we take distribution from blood, plasma and serum together [44–49], although they are separately listed in Table 2. There are a few compounds in Table 2 that are difficult to identify, and their SMILES notation is given in Table 3.

3. Results and discussion

3.1. Air to lung distribution

The log K_{lung}(human/rat) *in vitro* values in Table 1 were correlated against the Abraham descriptors [50,51] to yield Eq. (5).

$$\begin{aligned}&\log K_{lung}(\text{human/rat})\ \textit{in vitro}\\ &\quad = -1.250(0.077) + 0.639(0.178)E + 1.038(0.208)S\\ &\qquad +3.661(0.411)A + 3.041(0.300)B + 0.420(0.036)L\\ &N = 44,\ R^2 = 0.968,\ R^2_{cv} = 0.957,\ \text{S.D.} = 0.250,\ F = 231.8\end{aligned} \quad (5)$$

In Eq. (5), R^2_{cv} (sometimes known as Q^2) is the leave-one-out cross validated correlation coefficient. There is little cross-correlation between the descriptors in Eq. (5); the maximum correlation is $R^2 = 0.398$ between A and B. There are just enough data points to provide a training set and a test set to ascertain

Table 1
Air to lung and air to blood (Ref. [44]) partition coefficients, as log K, and blood to lung partition coefficients, as log P, for VOCs

Solute	log K_{lung}	Ref.	log K_{blood}	log P_{lung}
Helium	−1.98	[1]	−2.00	0.02
Oxygen	−1.62	[1]	−1.58	−0.04
Nitrous oxide	−0.37	[1,2,7]	−0.34	0.00
Carbon monoxide	−1.69	[1]	−1.67	−0.02
Sulphur hexafluoride	−2.14	[1]	−2.17	0.03
Ethane	−1.28	[1]	−1.02	−0.26
Pentane	−0.30	[1,2]	−0.29	−0.01
Hexane	0.00	[1,2]	0.21	−0.21
2-Methylpentane	−0.10	[1,2]	−0.39	0.29
3-Methylpentane	−0.05	[1,2]	−0.37	0.32
2,2-Dimethylbutane	−0.22	[1,2]	−0.59	0.37
Heptane	0.40	[1,2]	0.50	−0.10
Cyclopropane	−0.51	[1]	−0.21	−0.30
Methylcyclopentane	0.23	[1,2]	−0.07	0.30
Cyclohexane	0.43	[1,2]	0.17	0.26
Ethene	−0.25	[6]	−0.53	0.28
Propene	−0.24	[4]	−0.21	−0.03
Buta-1,3-diene	−0.32	[3]	0.09	−0.41
Ethyne	−0.06	[1]	−0.06	0.00
Dichloromethane	0.76	[1,2]	1.12	−0.36
Trichloromethane	1.12	[1,2,8]	1.15	−0.03
1,1,1-Trichloroethane	0.67	[1,2]	0.63	0.04
Trichloroethene	1.15	[1,2]	1.14	0.01
2-Chloro-1,3-butadiene	1.12	[9]		
1-Chloro-2,2,2-trifluroethane	0.34	[1,2]	0.14	0.20
1,1-Difluro-2-chloroethene	0.00	[1,2]	0.06	−0.06
Diethyl ether	1.14	[1]	1.11	0.03
Ethylene oxide	1.79	[6]	1.80	−0.01
Halothane	0.46	[1,2]	0.57	−0.11
Methoxyflurane	1.20	[1,2]	1.28	−0.08
Isoflurane	0.20	[1,2]	0.20	0.00
Enflurane	0.56	[1,2]	0.35	0.21
Sevoflurane	0.11	[1,2]	−0.20	0.31
Propanone	2.20	[1,2]	2.36	−0.16
Butanone	2.01	[1,2]	2.24	−0.23
Methanol	3.24	[1,2]	3.41	−0.17
Ethanol	3.07	[1,2]	3.27	−0.20
Propan-1-ol	2.91	[1,2]	3.06	−0.15
Propan-2-ol	2.77	[1,2]	3.02	−0.25
2-Methylpropan-1-ol	2.60	[1,2]	2.92	−0.32
Benzene	1.08	[1,2]	1.05	0.03
o-Xylene	1.94	[8]	1.42	0.52
Toluene	1.32	[1,2]	1.14	0.18
4-Chlorobenzotrifluoride	1.63	[5]	1.43	0.20
Tetrachloroethene[a]	1.86	[8]	1.19	0.67

[a] Not used in the analysis.

the predictive capability of Eq. (5). We selected the sets using the Kennard and Stone method [52], to ensure that the training set, the test set and the full set covered the same chemical space. The training set yielded the equation,

$$\log K_{lung}(\text{human/rat})\ \textit{in vitro} = -1.265(0.095) + 0.688(0.209)E + 1.159(0.253)S + 3.354(0.519)A + 3.096(0.348)B + 0.411(0.045)L$$
$$N = 30,\ R^2 = 0.969,\ R^2_{cv} = 0.945,\ \text{S.D.} = 0.259,\ F = 149.4 \quad (6)$$

When Eq. (6) was used to predict the independent test set of 14 compounds, comparison of the predicted and experimental values gave the average error, AE = 0.001, the absolute average error AAE = 0.195, the root mean square error RMSE = 0.244, and S.D. = 0.253 log units. There is therefore no bias in the predictions, with AE = 0.001, and hence Eq. (6), and by implication Eq. (5) can be used to predict further values of log K_{lung}(human/rat) *in vitro* to be around 0.26 log units. There was one outlier to Eq. (5) that we omitted: tetrachloroethene had a calculated log K value of 1.12 from Eq. (5) as compared to the observed value [8] of 1.86 log units.

A referee has suggested that the Kennard–Stone method does not always lead to a representative test set. We therefore constructed further test sets as follows: (i) The data were ordered in terms of the descriptor A and every third VOC was selected as the test set; this ensures that the test set is representative of the entire set as regards A. (ii) The data were ordered in terms of the descriptor B and every third VOC was selected as the test set; this ensures that the test set is representative of the entire set as regards B. (iii) The data were ordered in terms of the descriptor L, and every third VOC was selected as the test set; this ensures that the test set is representative of the entire set as regards L. Since the terms aA, bB, and lL make the largest contributions in Eq. (5), these three test sets should encompass the chemical space of the entire set. Results of predictions of the test sets from the equations for the corresponding training sets are given in Table 4. It can be seen that the assessment of predictive capability using the Kennard–Stone method is almost the same as the average of the three other methods of selection; the predictive S.D. is 0.25 on the Kennard–Stone method, and the average predicted S.D. from the three other methods is 0.26 log units.

3.2. Blood to lung distribution

Application of Eq. (4) to the *in vitro* values in Table 1 yielded Eq. (7); there are only 43 data points in Eq. (7) because we had no value for the gas to blood distribution of 2-chloro-1,3-butadiene. The terms eE, sS and aA were statistically not significant and were omitted.

$$\log P_{lung}(\text{human/rat})\ \textit{in vitro} = -0.143(0.084) - 0.383(0.161)B + 0.308(0.115)V$$
$$N = 43,\ R^2 = 0.264,\ R^2_{cv} = 0.164,\ \text{S.D.} = 0.190,\ F = 7.2 \quad (7)$$

The R^2 value is very poor in Eq. (7), but the S.D. value of only 0.190 log units suggests that Eq. (7) could be used to predict further *in vitro* values of log P_{lung}(human/rat). In order to assess the predictive power of Eq. (7), we selected a training set by the Kennard and Stone method [52]. The training equation is shown as Eq. (8),

$$\log P_{lung}(\text{human/rat})\ \textit{in vitro} = -0.156(0.089) - 0.271(0.186)B + 0.280(0.130)V$$
$$N = 29,\ R^2 = 0.207,\ R^2_{cv} = 0.050,\ \text{S.D.} = 0.178,\ F = 3.4 \quad (8)$$

The training equation, Eq. (8) was then used to predict log P_{lung}(human/rat) for the 14 compounds in the test set. Comparison of the predicted and experimental values gave

Table 2
Values of log *P* for the *in vivo* distribution of drugs from plasma, blood or serum to lung

Solute	*E*	*S*	*A*	*B*	*V*	Ia	Medium	log *P*	Ref.
5-Methyl-5-ethyl barbituric acid	1.03	1.17	0.46	1.18	1.2330	0	Plasma	−0.21	[25]
5-Ethyl-5-ethyl barbituric acid	1.03	1.14	0.47	1.18	1.3739	0	Plasma	0.00	[25]
5-Propyl-5-ethyl barbituric acid	1.03	1.14	0.47	1.18	1.5148	0	Plasma	0.16	[25]
5-Butyl-5-ethyl barbituric acid	1.03	1.14	0.47	1.18	1.6557	0	Plasma	0.18	[25]
5-Pentyl-5-ethyl barbituric acid	1.03	1.14	0.47	1.18	1.7966	0	Plasma	0.22	[25]
5-Hexyl-5-ethyl barbituric acid	1.03	1.14	0.47	1.18	1.9375	0	Plasma	0.08	[25]
5-Heptyl-5-ethyl barbituric acid	1.03	1.14	0.47	1.18	2.0784	0	Plasma	0.12	[25]
5-Octyl-5-ethyl barbituric acid	1.03	1.14	0.47	1.18	2.2193	0	Plasma	0.49	[25]
5-Nonyl-5-ethyl barbituric acid	1.03	1.14	0.47	1.18	2.3602	0	Plasma	0.35	[25]
Acebutolol	1.60	2.42	0.90	2.10	2.7556	0	B, P	0.82	[18,34]
AI-2 (SCH442416)	3.99	3.20	0.25	1.69	2.7696	0	Blood	0.62	[13]
AI-6	2.25	1.96	0.95	1.97	1.6648	0	Plasma	0.60	[16]
AI-9 (YH1885)	2.50	2.20	0.28	1.21	2.8039	0	Plasma	1.34	[29]
Acrylic acid	0.36	0.58	0.60	0.43	2.2820	1	Blood	0.14	[37]
Alfentanil	2.18	2.62	0.00	2.56	3.2586	0	plasma	−0.11	[24,36]
Alprazolam	2.90	2.50	0.00	1.55	2.2041	0	Plasma	0.32	[36,38]
Atenolol	1.45	1.88	0.69	2.00	2.1763	0	Blood	0.28	[18]
Azithromycin	2.24	3.20	1.19	4.90	5.9980	0	Serum	2.08	[15]
Betaxolol	1.18	1.51	0.24	1.79	2.5745	0	Plasma	2.29	[34]
Biperiden	1.85	1.25	0.31	1.57	2.6196	0	Plasma	1.79	[14,24,35]
Bisoprolol	0.83	1.50	0.30	2.19	2.7418	0	Plasma	1.62	[34]
Budesonide	2.33	3.03	0.40	2.17	3.2683	0	Plasma	0.94	[39]
Chlorpromazine	2.16	1.57	0.00	1.01	2.4056	0	Blood	1.81	[33]
Clomipramine	1.79	1.39	0.00	1.10	2.5239	0	Blood	2.16	[33]
Clotiazepam	2.07	1.53	0.00	1.40	2.2804	0	Blood	1.04	[33]
Cocaine	1.36	1.92	0.00	1.50	2.2977	0	Plasma	1.06	[17,31]
Cefazolin	3.62	4.00	0.75	2.68	2.8265	1	Plasma	−0.78	[36]
Cotinine	1.05	1.49	0.00	1.38	1.3867	0	Plasma	−0.18	[14]
CDRI-81/470	3.05	3.89	0.95	2.25	2.7392	0	Serum	−0.29	[40]
Diazepam	2.08	1.55	0.00	1.28	2.0739	0	Plasma	0.52	[14,21,36]
Dicloxacillin	3.51	2.52	0.54	2.28	3.0085	1	Plasma	−0.92	[14,24]
Digoxin	3.67	4.68	1.58	5.07	5.7525	0	Plasma	0.32	[36]
Doxorubicin	3.63	3.50	1.30	3.64	3.7284	0	Plasma	0.53	[20]
Erythromycin	2.90	3.73	1.25	4.96	5.7730	0	Serum	1.62	[15]
Fentanyl	1.83	1.75	0.00	1.81	2.8399	0	Plasma	1.14	[14,24,35]
Fluoxetine	1.00	1.30	0.10	0.93	2.2403	0	Plasma	1.24	[12]
Ftorafur	1.02	1.70	0.30	1.18	1.2830	0	Plasma	−0.59	[36]
Glycyrrhetinic acid	1.56	2.17	0.93	1.60	3.8984	1	Plasma	−0.66	[14]
Haloperidol	1.90	1.39	0.40	1.76	2.7980	0	Blood	1.73	[33]
Hexobarbital	1.50	1.37	0.17	1.37	1.7859	0	Plasma	0.45	[14,24]
Hydroquinone	1.06	1.27	1.06	0.57	0.8338	0	Blood	−0.16	[26]
Hydroxyzine	2.00	2.21	0.10	1.89	2.9231	0	Blood	1.06	[41]
Imipramine	1.15	1.60	0.00	1.15	2.4015	0	Plasma	2.11	[35]
Inaperisone	1.07	1.35	0.00	0.94	2.1323	0	Plasma	1.52	[35]
Lidocaine	1.10	1.47	0.06	1.24	2.0589	0	Plasma	0.58	[35]
Lorazepam	2.51	1.28	0.45	1.63	2.1141	0	Plasma	0.44	[27]
Methylphenidate	1.17	1.85	0.13	1.20	1.9092	0	plasma	1.54	[43]
Metoprolol	1.17	1.33	0.17	1.76	2.2604	0	Blood	1.06	[18]
Midazolam	2.57	2.01	0.00	1.38	2.2629	0	P, B	0.60	[22–24,36]
Miloxacin	2.00	1.81	0.36	1.38	1.6675	1	Plasma	−0.29	[14]
Nalidixic acid	1.56	1.80	0.59	1.25	1.6999	1	Plasma	−0.48	[24,36]
Nitrazepam	2.30	1.53	0.33	1.43	1.9848	0	Blood	0.26	[33]
Nicotine	0.87	0.88	0.00	1.09	1.3710	0	Plasma	0.32	[35]
o-Ethoxybenzamide (AI-5)	0.91	1.51	0.49	0.80	1.3133	0	Plasma	−0.03	[14,36]
Olanzapine	2.45	1.40	0.23	1.75	2.3742	0	Plasma	1.43	[28]
Oxprenolol	1.31	1.49	0.17	1.62	2.2174	0	P, B	1.27	[18,34]
PenicillinV	2.20	1.90	0.80	1.89	2.4358	1	Plasma	−0.80	[14,24,36]
Pentazocine	1.40	1.15	0.60	1.25	2.4464	0	Plasma	1.43	[14,24,35]
Phenobarbital	1.63	1.80	0.73	1.15	1.6999	0	Plasma	−0.08	[14,24,36]
Phencyclidine	1.23	0.96	0.00	0.82	2.1489	0	Plasma	1.61	[35]
Phenytoin	1.71	2.19	0.85	1.00	1.8693	0	Plasma	−0.10	[14,24,36]
Pindolol	1.70	1.65	0.30	1.48	2.0090	0	Plasma	1.01	[32]
p-Phenylbenzoic acid	1.48	1.30	0.59	0.50	1.5395	1	Plasma	−0.55	[14,24]
Practolol	1.45	1.90	0.60	1.84	2.1763	0	Blood	0.48	[18]

(*continued on next page*)

Table 2 (*continued*)

Solute	*E*	*S*	*A*	*B*	*V*	Ia	Medium	log *P*	Ref.
Propranolol	1.84	1.43	0.44	1.31	2.1480	0	Plasma	1.69	[18,21,34]
Pyrene	2.81	1.71	0.00	0.28	1.5846	0	Plasma	0.35	[30]
Quinidine	2.47	1.23	0.37	1.97	2.5512	0	Plasma	1.62	[35]
Salicylic acid	0.90	0.85	0.73	0.37	0.9904	1	Plasma	−0.72	[24,36]
Tenoxicam	2.86	2.67	0.55	1.85	2.1747	0	Plasma	−0.49	[36]
Terbinafine	1.89	1.38	0.00	1.03	2.6061	0	Plasma	0.41	[19]
Theophylline	1.50	1.60	0.54	1.34	1.2223	0	Plasma	−0.15	[14]
Thiopental (Thiopentone)	1.48	1.36	0.55	1.04	1.9014	0	Plasma	0.08	[14,24,36]
Thioridazine	2.70	2.10	0.00	1.30	2.9017	0	Plasma	1.13	[12]
Timolol	1.47	1.85	0.17	1.79	2.3759	0	Plasma	1.43	[34]
Tolbutamide	1.33	2.20	0.59	1.17	2.0580	0	Plasma	−0.60	[36]
Toliprolol	1.05	1.16	0.15	1.40	1.9199	0	Plasma	0.89	[32]
Trihexyphenidyl (Antitrem)	1.50	1.15	0.29	1.30	2.6300	0	Blood	1.87	[33]
Trifluoperazine	2.24	1.83	0.00	1.53	2.8911	0	Blood	1.33	[42]
Valproic acid	0.14	0.57	0.60	0.50	1.3102	1	Plasma	−0.38	[24,36]
Verapamil	1.81	2.91	0.00	2.51	3.7861	0	Plasma	1.70	[35]

AE = 0.052, AAE = 0.181, RMSE = 0.216 and S.D. = 0.224 log units. There is therefore little bias in the predictions, and Eq. (8) and hence the full Eq. (7) can therefore be used to predict values of log P_{lung}(human/rat) *in vitro* to about 0.22 log units. The very poor R^2 value for Eq. (7) is due to the very small spread of data in log P_{lung}(human/rat) *in vitro*, 0.78 log units, whereas for log K_{lung}(human/rat) the spread is 5.38 log units. This also shows that whilst R^2 provides information about the correlation between two (or more) variables, it is quite uninformative about the actual goodness-of-fit. As regards the predictive capability of an equation, by far the most useful statistic is the S.D. value for an independent test set. The statistics of Eqs. (7) and (8) are much inferior to those obtained previously by Balaz and Lukacova [10], and by Zhang [11]. However, as we pointed out above, the S.D. values obtained by these workers are unrealistically small. We have previously shown [44] that the inter-laboratory error, as S.D., on air to blood *in vitro* distribution, as log K_{blood}, was around 0.16 log units, so that the experimental error in the *in vitro* log P_{lung} values must be more than 0.16 log units. If a fitting error or a prediction error is much smaller than the experimental error, then the former value must be unrealistic. Quite recently, Liu et al. [53] have calculated partition coefficients from blood to seven tissues, using a global equation. For 35 VOCs the S.D. value between observed and fitted log P_{lung} *in vitro* values was 0.175 log units, very comparable to the statistics in Eqs. (7) and (8).

The log P_{lung} values for 80 drugs are assembled in Table 2, and all are *in vivo* values for rat. We note whether partition is from blood or from plasma, but take all values together. Application of Eq. (4) yielded Eq. (9); the *eE* and the *bB* terms were statistically not significant and were left out.

$$\log P_{lung}(\text{rat})\ in\ vivo = 0.269(0.164) - 0.523(0.107)S - 0.723(0.187)A + 0.720(0.088)V - 0.988(0.181)\ \text{Ia}$$
$$N = 80,\ R^2 = 0.647,\ R^2_{cv} = 0.600,\ \text{S.D.} = 0.509,\ F = 34.4 \quad (9)$$

The additional descriptor, Ia, is one we have used before [46–49] and takes the value Ia = 1 for carboxylic acids and Ia = 0 for all other compounds. We divided the 80 compounds into a training set and a test set. For the training set we obtained Eq. (10),

$$\log P_{lung}(\text{rat})\ in\ vivo = 0.231(0.216) - 0.389(0.152)S - 0.877(0.303)A + 0.657(0.116)V - 1.088(0.261)\ \text{Ia}$$
$$N = 40,\ R^2 = 0.642,\ R^2_{cv} = 0.521,\ \text{S.D.} = 0.531,\ F = 15.7 \quad (10)$$

Then Eq. (10) was used to predict log P_{lung} for the 40 compounds in the test set. For the observed and predicted values we find AE = 0.042, AAE = 0.407, RMSE = 0.506 and S.D. = 0.512 log units. We also chose training and test sets as set out for Eq. (5), with the data ordered in terms of the descriptors *S*, *A*, and *V*. Results of predictions of the 40 compound data sets are in Table 5. As before, the results using the Kennard–Stone method are almost exactly the average of the three other methods. There is almost no bias in the predictions, as shown by

Table 3
SMILES nomenclature for some compounds

Solute	SMILES
AI-2 (SCH442416)	COc5ccc(CCCn1ncc2c1nc(N)n3nc(nc23)c4ccco4)cc5
AI-6	Nc1cc(Cl)c(cc1c2nnn[nH]2)S(N)(=O)=O
AI-9 (YH1885)	CC1N(CCc2ccccc12)c4nc(Nc3ccc(F)cc3)nc(C)c4C
CDRI-81/470	c1cc(ncc1)N2CCN(CC2)C(=O)c3cc4c(cc3)nc(n4)NC(OC)=O

Table 4
Predictions of log K_{lung}(human/rat) *in vitro* through Eq. (5) using various training and test sets; the training set always has 30 VOCs and the test set always has 14 VOCs

Selection of sets	AE	AAE	RMSE	S.D.
Kennard–Stone	0.001	0.195	0.244	0.253
(i) Ordered in *A*	0.090	0.224	0.281	0.291
(ii) Ordered in *B*	0.035	0.195	0.255	0.264
(iii) Ordered in *L*	0.034	0.182	0.225	0.233

the small AE values and Eq. (10) and hence Eq. (9) is capable of predicting further values of log P_{lung}(rat) *in vivo* to about 0.51 log units. The statistics of Eqs. (9) and (10) are not very good. This is not due to a small range of the dependent variable, which is 3.2 log units, neither can it be due to the large range of compound type used, see Table 2, because for the comparable blood/plasma to muscle equation, a S.D. value of 0.25 log units is found for 60 drugs (47). However, partition of 85 drugs from blood/plasma to liver, led to a regression equation with a S.D. of 0.42 log units (49). The equations that we, and others, have constructed for partition of VOCs and drugs to lung refer to passive transfer. Any active transport or efflux mechanisms are not taken into account. It is therefore possible that for transfer to muscle, these active mechanisms play little part, but for transfer to liver and to lung they have the effect of increasing the correlation/predictive error. We also repeated Eq. (9) by inclusion of an indicator variable so that blood to lung and plasma to lung values could be taken separately. The indicator variable was not statistically significant, and no improvement was found by taking the two sets of data separately. We also repeated Eq. (9) using data only on partition from plasma to lung, but the resulting equation was almost identical to Eq. (9).

We note that the very small test set of Zhang and Zhang [33] leads to predictive S.D. values of 0.62 and 0.60 log units, on two different equations. Poulin and Theil [14] have put forward a method for the calculation of plasma to tissue partition coefficients that requires a number of parameters including the unbound fraction of the drug in plasma and the tissue in question. For the plasma to lung partition of 43 drugs, some to rat lung and some to rabbit lung, we calculated a S.D. value of 0.49 log units from the given data [14]. Our predictive assessment of 0.51 log units is better than those of Zhang and Zhang [33], and about the same as that of Poulin and Theil [14] with a less general equation. We suggest that Eq. (9) is the most reasonable general equation put forward to date for the prediction of log P_{lung}(rat) *in vivo*.

Table 5
Predictions of log P_{lung}(rat) *in vitro* through Eq. (9) using various training and test sets; the training set always has 40 VOCs and the test set always has 40 VOCs

Selection of sets	AE	AAE	RMSE	S.D.
Kennard–Stone	0.042	0.407	0.506	0.512
(i) Ordered in *S*	−0.129	0.470	0.585	0.593
(ii) Ordered in *A*	−0.046	0.359	0.451	0.457
(iii) Ordered in *V*	−0.108	0.420	0.482	0.488

We have investigated the possibility of combining the *in vitro* and *in vivo* values in Tables 1 and 2. We used the descriptors in Eq. (9) plus an indicator variable for the *in vitro* and *in vivo* values; Iv takes the value 1.0 for the *in vitro* values and 0.0 for the *in vivo* values. The resulting equation is,

$$\log P_{lung}(\text{human/rat})\ \textit{in vivo/in vitro} = 0.282(0.137) - 0.462(0.085)S - 0.694(0.153)A + 0.664(0.070)V - 1.005(0.153)\,\text{Ia} - 0.514(0.129)\,\text{Iv}$$
$$N = 123,\ R^2 = 0.676,\ R^2_{cv} = 0.639,\ \text{S.D.} = 0.432,\ F = 48.8 \quad (11)$$

As usual, we can assess the predictive power of the equation by constructing a training and a test set. For the former we find,

$$\log P_{lung}(\text{human/rat})\ \textit{in vivo/in vitro} = 0.338(0.235) - 0.415(0.124)S - 0.971(0.255)A + 0.649(0.118)V - 1.080(0.225)\,\text{Ia} - 0.526(0.207)\,\text{Iv}$$
$$N = 62,\ R^2 = 0.691,\ R^2_{cv} = 0.590,\ \text{S.D.} = 0.440,\ F = 25.0 \quad (12)$$

When Eq. (12) was used to predict the 61 data points in the test set, we found that AE = 0.001, AAE = 0.338, RMSE = 0.441 and S.D. = 0.445 log units. Hence we can say that the predictive ability of Eq. (11) is around 0.45 log units for the VOCs and drugs combined. In order to ascertain whether Eq. (9) or Eq. (11) is the better equation for predictions for drugs, we calculated the predictive statistics for the 45 drugs in the 61 compound test set. We found that AE = 0.018, AAE = 0.293, RMSE = 0.424, and S.D. = 0.428 log units, which is better than the test statistics of Eq. (9) where S.D. = 0.512 log units. We therefore suggest Eq. (11) as the preferred equation for the prediction of further values of log P_{lung}(rat) *in vivo* for drugs, with an estimated predictive error of about 0.43 log units. A plot of the calculated values of log P_{lung} from Eq. (11) against the observed values is shown in Fig. 1, and illustrates how the indicator variable, Iv, enables the VOCs to be correlated in the same equation as drugs.

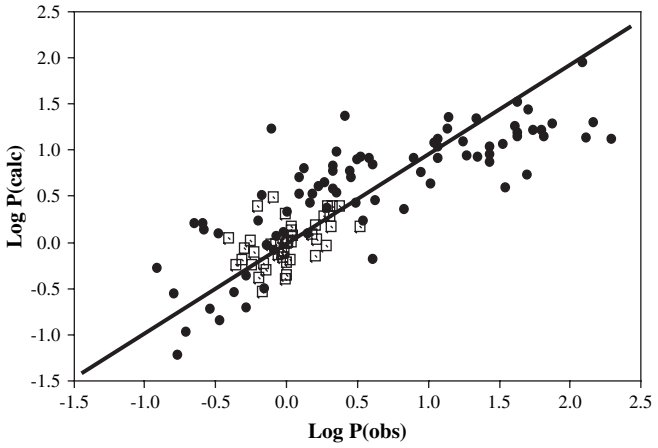

Fig. 1. A plot of log *P*(calc) against log *P*(obs) on Eq. (11); ● drugs, □ VOCs.

Table 6
Coefficients in LFERs for air to biological phases and air to water at 37 °C, and the distance parameter D'

Solvent	c	e	s	a	b	l	D'
Water	−1.347	0.928	2.795	3.717	4.297	−0.254	0.00
Blood	−1.069	0.456	1.083	3.738	2.580	0.376	2.55
Lung	−1.254	0.617	1.046	3.698	3.023	0.420	2.29
Muscle	−1.039	0.207	0.723	3.242	2.469	0.463	2.98
Liver	−0.943	0.000	0.836	2.836	2.081	0.564	3.32
Brain	−0.987	0.263	0.411	3.358	2.025	0.591	3.48
Fat	−0.052	0.051	0.728	1.783	0.332	0.743	5.05

The points for the VOCs are randomly scattered about the line of best fit.

3.3. General discussion

The equation we have constructed for air to lung partition of VOCs, Eq. (5), is reasonably good, with $R^2 = 0.968$, and S.D. = 0.25 log units, the latter being near to the estimate of the predictive capability of Eq. (5), at 0.26 log units. Hence Eq. (5) can be used for the prediction of further values of log K_{lung}(human/rat) *in vitro* for VOCs. Since descriptors are available for a very large number of VOCs, and in any case can be calculated easily from structure [54], further values can be obtained by simple arithmetic. One considerable advantage of the present method is that the descriptors in the LFER, Eq. (3), refer to particular chemical properties of the solute. Hence the coefficients in the LFER will then refer to the corresponding chemical properties of the condensed phase, and comparison of the chemical properties of condensed phases can be carried out simply by a comparison of the coefficients e to l in equations corresponding to Eq. (3). We give in Table 6 the LFER coefficients for a number of air to condensed phase partitions, including air to water at 310 K [55]. The phases can be compared coefficient by coefficient, if required. For example, the b-coefficient corresponds to the phase hydrogen-bond acidity, because a hydrogen-bond basic VOC will interact with a hydrogen-bond acidic phase. As can be seen from Table 6, water is a stronger hydrogen-bond acid than any of the biological phases, with fat having almost no hydrogen-bond acidity at all. It is not so easy to obtain overall comparisons just by simple examination of the coefficients. However, this can be done by regarding the five coefficients for phases as points in five-dimensional space [56,57]. The nearer the points are, as calculated by Euclidean geometry, the closer will be the phases in a chemical sense. We denote the Euclidean distance as D', and give in Table 6 values of D' from water as a standard phase. The biological phases blood, lung, muscle, liver and brain all fall between water and fat, as one might expect. Both the goodness-of-fit of Eq. (7), as shown by S.D. = 0.19, and the predictive capability of Eq. (7), as shown by S.D. = 0.22 log units for the test set, are quite good, and so Eq. (7) is a reasonable equation for the prediction of further values of plasma/blood to lung distribution for VOCs.

The distribution of drugs between plasma/blood/serum and lung leads to small values of the coefficients, as in Eq. (9), and to a low value of R^2. The predictive S.D. value of 0.51 for drugs from Eq. (9) is quite high, but the combined equation, Eq. (11), has an estimated predictive capability of 0.43 log units for drugs, and this is our preferred predictive equation. The previous assessments of predictive ability of equations for plasma/blood/serum to lung distribution also yield rather large S.D. values of around 0.6 log units [33] or 0.49 log units [14], and so it is possible that distribution to lung might be affected by active transport and metabolism.

All the statistical analyses were carried out using Minitab, version 14 [58].

Acknowledgements

We thank Dr. Chris Luscombe for helping with the Stone–Kennard method. We are grateful to Anne-Marie Arthur, Training and Technical Support Specialist, Minitab Ltd., for devising a rapid method for the calculation of some of the statistics.

References

[1] M.H. Abraham, P.K. Weathersby, J. Pharm. Sci. 83 (1994) 1450–1455.
[2] V. Fiserova-Bergerova, M.L. Diaz, Int. Arch. Occup. Environ. Health 58 (1986) 75–87.
[3] L.M. Sweeney, M.W. Himmelstein, M.L. Gargas, Chem. Biol. Interact. 135–136 (2001) 303–322.
[4] J.G. Filser, R. Schmidbauer, F. Rampf, C.M. Baur, C. Putz, G.A. Csanady, Toxicol. Appl. Pharmacol. 169 (2000) 40–51.
[5] J.B. Knaak, L.W. Smith, R.D. Fitzpatrick, J.R. Olson, P.E. Newton, Inhal. Toxicol. 10 (1998) 65–85.
[6] Gy. A. Csanady, B. Denk, C. Putz, P.E. Kreuzer, W. Kessler, C. Baur, M.L. Gargas, J.G. Filser, Toxicol. Appl. Pharmacol. 165 (2000) 1–26.
[7] A. Steward, P.R. Allott, A.L. Cowles, W.W. Mapleson, Br. J. Anaesth. 45 (1973) 282–293.
[8] K.D. Thrall, R.A. Gies, J. Muniz, A.G. Woodstock, G. Higgins, J. Toxicol. Environ. Health A 65 (2002) 2075–2086.
[9] M.W. Himmelstein, S.C. Carpenter, M.V. Evans, P.M. Hinderliter, E.M. Kenyon, Toxicol. Sci. 79 (2004) 28–37.
[10] S. Balaz, V. Luckacova, Quant. Struct.–Act. Relat. 18 (1999) 361–368.
[11] H. Zhang, J. Pharm. Sci. 93 (2004) 1595–1604.
[12] J. Wojcikowski, W.A. Daniel, Pol. J. Pharmacol. 54 (2002) 647–654.
[13] F. Fazio, S. Todde, R.M. Moresco, P. Simonelli, P.G. Baraldi, B. Cacciari, G. Spalluto, K. Varani, A. Monopoli, M. Matarrese, A. Carpinelli, F. Magni, G.K. Kienle, J. Med. Chem. 43 (2000) 4359–4362.
[14] P. Poulin, F.P. Theil, J. Pharm. Sci. 89 (2000) 16–35.
[15] D. Davila, L. Kolacny-Babic, Biopharm. Drug Dispos. 12 (1991) 505–514.
[16] H.S. Lee, M.G. Lee, Biopharm. Drug Dispos. 16 (1995) 547–561.
[17] J.I. Javaid, J.M. Davis, Biopharm. Drug Dispos. 14 (1993) 357–364.
[18] J.M. Cruickshank, Am. Heart J. 100 (1980) 160–178.

[19] M. Hosseini-Yeganeh, A.J. McLachlan, J. Pharm. Sci. 90 (2001) 1817–1828.
[20] M.R. Gasco, A. Fundaro, R. Cavalli, A. Bargoni, D. Vighetto, Pharmacol. Res. 4 (2000) 337–343.
[21] P. Poulin, F.P. Theil, J. Pharm. Sci. 91 (2002) 1358–1367.
[22] S. Bjorkman, D.R. Wada, B.M. Berling, G.J. Benoni, J. Pharm. Sci. 90 (2001) 1226–1241.
[23] S. Bjorkman, A. Fyge, Z. Qi, J. Pharm. Sci. 85 (1996) 887–889.
[24] S. Bjorkman, J. Pharm. Pharmacol. 54 (2002) 1237–1245.
[25] G.E. Blakey, I.A. Nestorov, P.A. Arundel, L.J. Aarons, M.J. Rowland, J. Pharmacokinet. Biopharm. 25 (1997) 277–312.
[26] R.A. Corley, J.C. English, T.S. Hill, L.A. Fiorcia, D.A. Morgott, Toxicol. Appl. Pharmacol. 165 (2000) 163–174.
[27] R.T. Schillings, S.F. Sisenwine, H.W. Ruelius, Drug Metab. Dispos. 5 (1977) 425–435.
[28] M. Aravagiri, Y. Teper, S.R. Marder, Biopharm. Drug Dispos. 20 (1999) 369–377.
[29] K.S. Han, Y.G. Kim, J.K. Yoo, J.W. Lee, M.G. Lee, Biopharm. Drug Dispos. 19 (1998) 493–500.
[30] S. Haddad, J. Withey, S. Lapare, F. Law, K. Krishnan, Environ. Toxicol. Pharmacol. 5 (1998) 245–255.
[31] P.L. Bonate, A. Swann, P.B. Silverman, J. Pharm. Sci. 85 (1996) 878–883.
[32] P. Doze, P.H. Elsinga, B. Maas, A. Van Waarde, T. Wegman, W. Vaalburg, Neurochem. Int. 40 (2002) 145–155.
[33] H. Zhang, Y. Zhang, J. Med. Chem. 49 (2006) 5815–5829.
[34] T. Rodgers, D. Leahy, M. Rowland, J. Pharm. Sci. 94 (2005) 1237–1248.
[35] T. Rodgers, D. Leahy, M. Rowland, J. Pharm. Sci. 94 (2005) 1259–1275.
[36] T. Rodgers, M. Rowland, J. Pharm. Sci. 95 (2006) 1238–1257.
[37] K.A. Black, L. Finch, Toxicol. Lett. 78 (1995) 73–78.
[38] W.R. Banks, H. Yamakitra, G.A. Digenis, J. Pharm. Sci. 81 (1992) 797–801.
[39] J.M.M. van den Bosch, C.J.J. Westerman, J. Aumann, S. Edsbacker, M. Tonnesson, O. Selroos, Biopharm. Drug Dispos. 14 (1993) 455–459.
[40] N.V. Nagaraja, S.K. Singh, J.K. Paliwal, R.C. Gupta, J. Pharm. Pharmacol. 52 (2000) 1257–1264.
[41] S.F. Pong, C.L. Huang, J. Pharm. Sci. 63 (1974) 1527–1532.
[42] G. Schmalzing, Drug. Metab. Dispos. 5 (1977) 104–115.
[43] H. Kotaki, F. Nakazato, T. Aoyama, Y. Saitoh, F. Nakagawa, Chem. Pharm. Bull. 36 (1988) 3190–3195.
[44] M.H. Abraham, A. Ibrahim, W.E. Acree Jr., Chem. Res. Toxicol. 18 (2005) 904–911.
[45] M.H. Abraham, A. Ibrahim, W.E. Acree Jr., Eur. J. Med. Chem. 41 (2006) 494–502.
[46] M.H. Abraham, A. Ibrahim, Y.H. Zhao, W.E. Acree Jr., J. Pharm. Sci. 95 (2006) 2091–2100.
[47] M.H. Abraham, A. Ibrahim, W.E. Acree Jr., Chem. Res. Toxicol. 19 (2006) 801–808.
[48] M.H. Abraham, A. Ibrahim, Eur. J. Med. Chem. 41 (2006) 1430–1438.
[49] M.H. Abraham, A. Ibrahim, W.E. Acree Jr., Eur. J. Med. Chem., in press.
[50] M.H. Abraham, Chem. Soc. Revs. 22 (1993) 73–83.
[51] M.H. Abraham, A. Ibrahim, A.M. Zissimos, J. Chromatogr., A 1037 (2004) 29–47.
[52] R.W. Kennard, L.A. Stone, Technometrics 11 (1959) 137–148.
[53] H.X. Liu, X.J. Yao, R.S. Zhang, M.C. Liu, Z.D. Hu, B.T. Fan, J. Comput. Aided Mol. Des. 19 (2005) 499–508.
[54] Pharma Algorithms, ADME Boxes, Version 3.0, PharmaAlgorithms Inc, 591 Indian Road, Toronto, ON M6P 2C4, Canada, 2006.
[55] M.H. Abraham, A. Ibrahim, W.E. Acree Jr., Fluid Phase Eq. 251 (2007) 93–109.
[56] M.H. Abraham, Can we identify models for intestinal absorption, blood-brain barrier distribution and intestinal absorption? in: M. Ford, D. Livingstone, J. Dearden, H. van de Waterbeemd (Eds.), EuroQSAR 2002. Designing Drugs and Crop Protectants: Processes, Problems and Solutions Blackwell, Oxford, 2003, pp. 5–7.
[57] M.H. Abraham, F. Martins, J. Pharm. Sci. 93 (2004) 1508–1523.
[58] Minitab, Version 14, Minitab Inc, Quality Plaza, 1829 Pine Hall road, State College, Pa 16801–3008, USA, 2003.