ELSEVIER

Journal of Chromatography A, 749 (1996) 211–217

JOURNAL OF CHROMATOGRAPHY A

Novel computational methods for the determination of partition coefficients by planar chromatography

Robert M. Kleyle[a], David Nurok[b], Aaron D. Kossoy[c,*], Stuart C. Burris[c,1]

[a]*Department of Mathematical Sciences, Indiana University-Purdue University at Indianapolis, Indianapolis, IN 46202, USA*

[b]*Department of Chemistry, Indiana University-Purdue University at Indianapolis, Indianapolis, IN 46202, USA*

[c]*Lilly Research Laboratories, A Division of Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, USA*

Received 24 March 1996; revised 24 April 1996; accepted 24 April 1996

Abstract

An evaluation of an empirical method for the determination of log P values for 24 phenols by thin-layer chromatography is described. Separation is on bonded C_8 plates and aqueous methanol is used as the mobile phase. The method is based on the Soczewinski equation and uses standards to define a solvent composition at which R_M is equal to log P. In addition, a new approach, based on a paired t-test is described. The t-statistic is used to compare the differences between the literature log P values and the computed R_M values over a range of solvent compositions. The p-value corresponding to the t-score can then be used as the measure of agreement between the computed and chromatographic values of log P. The mean difference between the estimated and the literature log P values varies between 0.1 and 0.2 of a log P unit, depending on the method used.

Keywords: Partition coefficients; Octanol–water partition coefficients; Phenols

1. Introduction

Log P, the logarithm of the octanol–water partition coefficient, is an important descriptor for lipophilicity. There are a number of problems with the classical shake-flask method of determining log P [1]; the recent trend has been to calculate the value of log P using various computational approaches [2,3] or to determine the value by either HPLC [4] or TLC [5].

The latter techniques use reversed-phase systems with a stationary phase such as bonded C_8 and a mobile phase such as aqueous methanol to determine either log k' (HPLC) or R_M (TLC), either at a single arbitrary mobile phase composition or over a range of compositions. In the latter case, the data are usually extrapolated to zero concentration of the organic modifier. Log k' or R_M for a training set of solutes is then related to log P by a suitable regression equation.

The extrapolation approach can be satisfactory because there is a smooth change in the retention of a solute with changes in the composition of a mobile phase consisting of an aqueous solution of an organic modifier. Such a relationship was originally demonstrated by Soczewinski and Wachtmeister [6].

Schoenmakers and co-workers [7] later demonstrated that the capacity factor of a solute can be

*Corresponding author.

[1]Present address: Department of Chemistry, North Carolina State University, Raleigh, NC 27695, USA.

0021-9673/96/$15.00
PII S0021-9673(96)00420-7

related to the composition of the mobile phase in a reversed-phase separation by the following second order equation:

$$\ln k' = a_i\varphi^2 + b_i\varphi + c_i \quad (1)$$

where φ is the volume percent of an aqueous organic modifier and a_i, b_i and c_i are empirical constants for a solute i. This relationship has been widely used in liquid chromatography.

The retention of a solute can also be predicted using the following equation, originally introduced by Soczewinski [8]:

$$R_M = a_i \log Xs + b_i \quad (2)$$

where Xs is the mole fraction of a strong solvent in a binary mixture with a weak solvent, and a_i and b_i are empirical constants for a solute i. This equation was originally suggested for normal-phase chromatography but is valid also for reversed-phase chromatography [9,10].

A new approach to determining log P by reversed-phase TLC has been suggested and is based on Eq. (2) [11]. It is not possible to use this equation to extrapolate to zero concentration of organic modifier, because of the logarithmic nature of the relationship. It is however possible to select a concentration of organic modifier (i.e. the strong solvent) such that R_M equals log P for a given solute. This concentration is referred to as the equivalence mole fraction. The range of equivalence mole fractions is small for a given class of solutes (the mole fraction range for the phenols under study is 0.0427 to 0.0838) and the midrange log mole fraction can be used to estimate log P. In a preliminary study [11] it was found that these estimated values of log P were within 0.5 units of the corresponding shake-flask values for the majority of the solutes tested. While the previous results are acceptable, they were obtained for small sets of compounds, each set containing between five and seven compounds of known shake-flask values. In order to test the method more rigorously, a study utilizing 24 phenols was performed, and is discussed below.

An implicit assumption in this approach is that a plot of the values of log P obtained by chromatography versus the shake-flask values has a slope of unity and an intercept of zero. This assumption is generally not valid for virtually all reports of log P values estimated by chromatographic methods; where R_M is correlated with log P either at zero concentration of organic modifier (an extrapolated value) or at an arbitrary concentration of the modifier. The only exception is for systems with a stationary phase of inert material impregnated with either n-octanol or silicone oil. Plots obtained from these systems have slopes with values close to unity, but the intercepts are not close to zero [12]. A statistical analysis of the preliminary results for the new computational approach was inconclusive even though two classes of solutes (phenols by TLC and quinolines by OPLC) show a high probability of a slope of unity and an intercept of zero. It is demonstrated below, with respect to a set of 21 phenols, that it is possible to obtain a slope and intercept that are not significantly different from unity and zero, respectively.

2. Experimental

Reversed-phase KC_8F TLC plates (20×20 cm; catalogue number 4808-820) containing a fluorescent indicator were obtained as a gift from Whatman Inc. (Clifton, NJ, USA). Chromatography was performed in a flat-bottomed chamber (Camag, Wilmington, NC, USA). Aqueous methanol, with the aqueous component 0.5 M in NaCl, was used as the mobile phase. The 0.5 M NaCl was added in order to protect the TLC layer at high water concentration [13]. Trifluoroacetic acid (0.004 M) was added to the aqueous component, as suggested by Garst and Wilson for HPLC [14].

Each solvent system was run at five mole fractions and the results for each solute were fitted to Eq. (2). All results reported are derived from data with a value of R^2 of 0.99 or better. The phenols and the trifluoroacetic acid were purchased from Aldrich (Milwaukee, WI, USA). Methanol was purchased from Baxter Burdick and Jackson (Muskegon, MI, USA); sodium chloride was purchased from EM Science (Cherry Hill, NJ, USA).

3. Results and discussion

The structures of the phenols used are shown in Fig. 1. Table 1 lists the literature log P values [15],

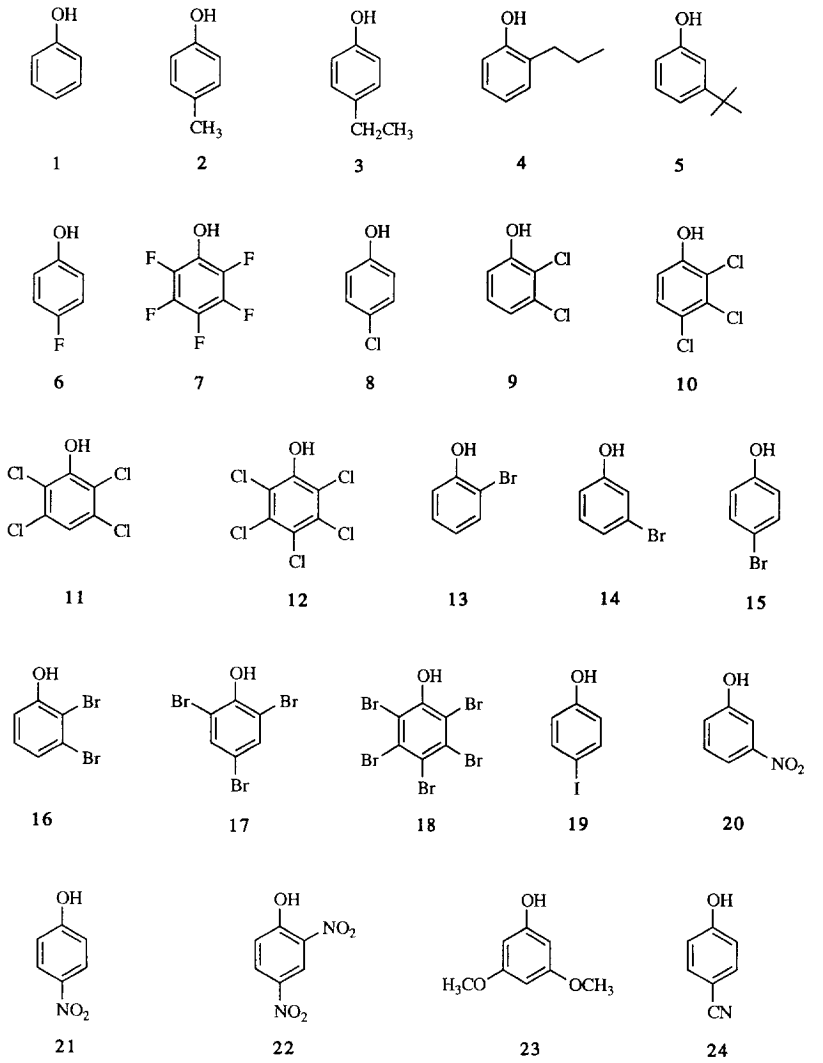

Fig. 1. Structures of the phenols used in the study.

the log equivalence mole fractions and the computed R_M values using midrange mole fraction for sets of 24 and 21 phenols, respectively.

A literature log P value is not available for pentabromophenol, but there is a literature value for the log partition coefficient in methylene chloride–water [15]. It was assumed that the latter is the same as the value for chloroform–water, which can be transformed into the required value after rearranging the following relationship established by Leo and Hansch [16]:

$$\log P_{CHCl_3/H_2O} = 1.126 \log P_{octanol/H_2O} - 1.343 \quad (3)$$

where $\log P_{CHCl_3/H_2O}$ is the chloroform–water partition coefficient and $\log P_{octanol/H_2O}$ is the octanol–water partition coefficient. The error in assuming that the partition coefficient in methylene chloride–water is the same as in chloroform–water is assumed to be small. The estimated log P value for pentabromophenol yields a log equivalence mole fraction (see below) similar to those of the other two pentahalophenols in the study. The values of log equival-

Table 1

Compound[a]	Literature log P	Log equivalence mole fraction[b]	24 Compounds		21 Compounds	
			R_M[c]	$R_M - \log P$	R_M[d]	$R_M - \log P$
1	1.46	−1.1397	1.64	0.18	1.51	0.05
2	1.94	−1.1344	2.17	0.23	2.01	0.07
3	2.58	−1.1720	2.74	0.16	2.55	−0.33
4	2.93	−1.0815	3.44	0.51	3.22	0.29
5	3.23	−1.0767	3.80	0.57	3.56	0.33
6	1.77	−1.1368	1.99	0.22	1.83	0.06
7	3.23	−1.3027	2.96	−0.27		
8	2.39	−1.2092	2.43	0.04	2.25	−0.14
9	3.00	−1.2414	2.94	−0.06	2.74	−0.26
10	3.62	−1.2170	3.64	0.02	3.41	−0.21
11	3.88	−1.1982	3.98	0.10	3.73	−0.15
12	5.12	−1.3698	4.46	−0.66		
13	2.35	−1.2115	2.38	0.03	2.21	−0.14
14	2.63	−1.2419	2.58	−0.05	2.39	−0.22
15	2.59	−1.2455	2.53	−0.06	2.35	−0.24
16	3.22	−1.1511	3.48	0.26	3.26	0.04
17	4.11	−1.2133	4.15	0.04	3.90	−0.21
18	5.30[e]	−1.3178	4.84	−0.46		
19	2.91	−1.2265	2.90	−0.01	2.70	−0.21
20	2.06	−1.1726	2.19	0.13	2.03	−0.03
21	1.91	−1.1473	2.11	0.20	1.95	0.04
22	1.54	−1.1308	1.78	0.24	1.62	0.08
23	1.64	−1.1099	1.93	0.29	1.77	0.13
24	1.60	−1.1427	1.79	0.19	1.64	0.04

[a] See Fig. 1.
[b] To obtain R_M values accurate to two decimal places it was necessary to compute log equivalence concentrations to four decimal places.
[c] Calculated using Eq. (1) and a log Xs value of −1.2233.
[d] Calculated using Eq. (1) and a log Xs value of −1.1611.
[e] Estimated log P value. See text.

ence mole fraction for the three pentahalophenols are substantially more negative than those of the other phenols studied.

For the complete set of 24 phenols, fourteen of the R_M values are within 0.20 unit of the corresponding shake-flask values and an additional six values are within 0.29 unit. The remaining four phenols have values that differ from the corresponding shake-flask values by between 0.46 and 0.66 unit. Thus it is seen that this method provides decent estimates of log P for a diverse set of phenols. It must, however, be noted that the four phenols with the poorest agreement between computed and experimental log P values have values of log Xs at, or, near the limits of the range of equivalence mole fractions, whereas the best agreement is found for those phenols with log Xs values close to the midrange value of log Xs.

The shake-flask values of log P were regressed against the corresponding chromatographic estimates; the estimated slope of the fitted line is 1.126 and is significantly different from a slope of unity ($p=0.0360$). The estimated intercept of the fitted line is −0.4385 which is significantly different from zero ($p=0.0169$).

There are insufficient compounds in the set of phenols to search for correlations between log equivalence mole fraction and structure. It is however noted that the compounds with the most positive log equivalence mole fractions (−1.0815 and −1.0767) are the phenols with either an *n*-propyl or a *tert.*-butyl group attached. These are the phenols with the largest alkyl substituents in the study. The three pentahalophenols have the most negative values (−1.3698 to −1.3027) of the log equivalence mole fraction. Elimination of the latter three compounds results in a smaller range of log equivalence

mole fractions (−1.2455 to −1.0767). Computation of log P using the new midrange value results in substantially smaller differences between the chromatographic and the shake-flask values. The differences are less than 0.10 unit for nine of the phenols, and between 0.10 and 0.26 unit for ten phenols. Only two of the phenols exhibited larger differences of 0.28 and 0.33 unit, respectively.

The shake-flask values of log P were regressed against the corresponding chromatographic estimates obtained with a midrange equivalence mole fraction after the pentahalophenols were excluded; the estimated slope of the fitted line is 1.036, and this is not significantly different from a slope of unity ($p=0.497$). The estimated intercept of the fitted line is −0.055 which is not significantly different from zero ($p=0.691$). Thus, exclusion of the three pentahalophenols (the compounds with the most negative log equivalence mole fraction values) results in significantly better results both with respect to agreement with shake-flask values and with respect to the slope and intercept of the regression. These compounds are the only pentasubstituted phenols considered.

The present results indicate that a good to excellent agreement between log P and shake-flask values can be obtained by this method, depending on which set of compounds are included in the study. This agreement compares favorably to that obtained using HPLC, a technique which Kaliszan [4] suggests yields better results than TLC. In an inter laboratory comparison of shake-flask to HPLC values involving eighteen different laboratories, it was concluded that the agreement between HPLC and shake-flask results was usually within 0.5 of a log P unit [17]. The results presented in the current paper show a similar agreement. Thus, for this set of phenols, the current method is competitive with the HPLC method. Moreover, as discussed above and in reference [11], when the chromatographic log P is regressed against the shake-flask values, the current method results in a slope and intercept that are not significantly different from unity and zero, respectively. There are no reports of such a relationship when the established chromatographic methods are used to estimate log P using bonded stationary phases. The values of the above slope and intercept are based on relatively small sample sizes. With a larger number of solutes,

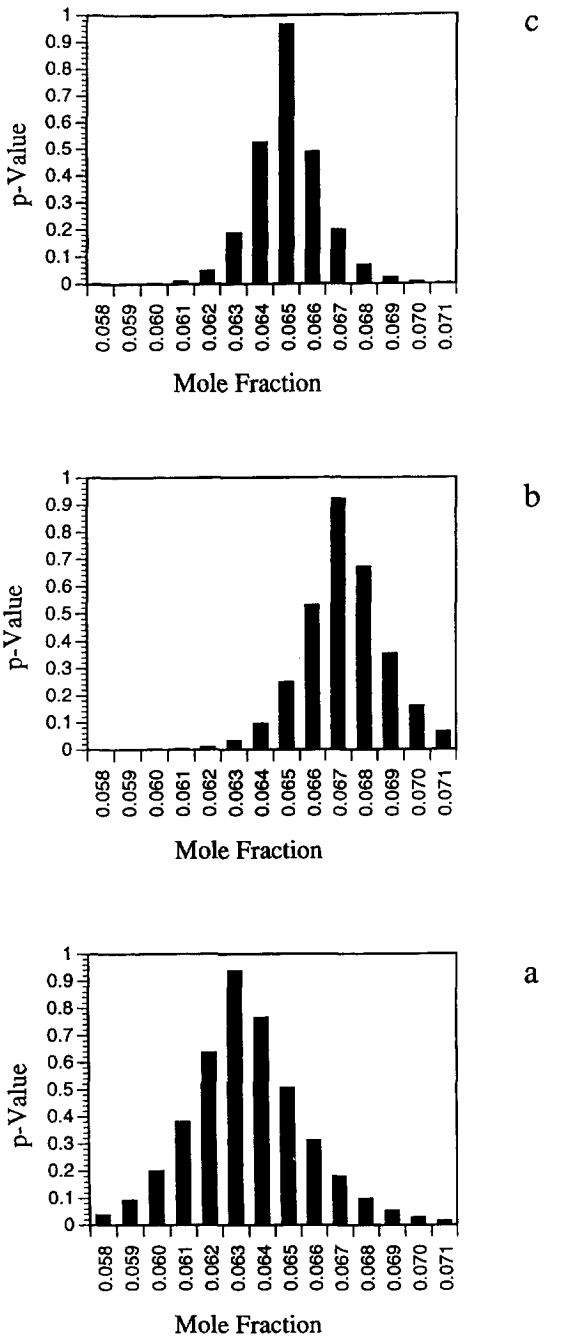

Fig. 2. Plots of p-value versus mole fraction used in Eq. (1). Plots a, b and c are for sets of 24, 21 and 18 phenols, respectively, as discussed in the text.

the slope and intercept may indeed be significantly different from unity and zero respectively, but these differences are expected to be small. Thus, these initial results indicate that the method based on an equivalent mole fraction appears competitive with the established chromatographic approaches to estimating log *P*.

An alternative approach to that discussed above is to use Eq. (2) to compute R_M for each compound over a range of mole fractions, and then to perform a paired *t*-test where the *t*-statistic is used to compare the differences between the literature log *P* value and the computed R_M for each value of the mole fraction. The *p*-value corresponding to the *t*-score can then be used as the measure of agreement. The higher the *p*-value, the better the overall agreement between log *P* and R_M, since any differences are less significant statistically. The methanol concentration corresponding to the highest *p*-value can then be used to compute R_M. Plots of the *p*-value versus mole fraction for the sets of 18, 21 and 24 phenols are shown in Fig. 2. The set of eighteen phenols was obtained by eliminating the three phenols with the smallest equivalence mole fractions (the three pentahalophenols) as well as the three phenols with the largest equivalence mole fractions (*ortho-n*-propylphenol, *meta-tert.*-butyphenol and 3,5-dimethoxyphenol). The distribution of sizable *p*-values occurs over a smaller concentration range for the smaller set of phenols. These plots were constructed by incrementing the methanol concentration in units of 0.001 mole fraction. This procedure can be reiterated with smaller increments until an optimal mole fraction is reached; this yields the highest *p*-value and the best overall agreement. This approach avoids the use of the arbitrary midrange concentration for estimating log *P*. However, as with the midrange method, it is necessary to find a training set of compounds that have a small range of equivalence mole fractions and then to limit the technique to other compounds of similar structure. It is expected that the distribution of *p*-values and the value of the maximum *p*-value should provide an estimate of the appropriateness of the training set of compounds.

While the paired-*t* method yields the best overall agreement, it does not always yield the best estimate of log *P* for individual compounds. For one member (pentachlorophenol) of the set of 24 phenols, the paired-*t* method results in a larger absolute difference (0.76 unit) between the shake-flask and R_M value than the largest difference (0.66 unit) — also for pentachlorophenol — that occurs for the midrange method. For 3-*tert.*-butylphenol in the set of 21 phenols, the paired-*t* method results in a larger absolute difference (0.37 of a unit) between the shake-flask and the R_M value than the largest absolute difference (0.33 unit) — also for 3-*tert.*-butylphenol — that occurs for the midrange method. A brief comparison of the results by the two methods is shown in Table 2. There is only a very small difference between the two methods when comparing the mean of the absolute differences between the shake-flask and the estimated log *P* values. The difference between the literature and the TLC result decreases by about 30%, to a value of 0.14 of a unit, when the set of 24 phenols is reduced to 21.

Table 2 shows also that the mean absolute difference between the literature and the TLC result becomes even smaller when all the compounds are of a more narrowly defined class. In the case of the nine phenols considered that are mono-, di-, tri- and tetrasubstituted with chlorine or bromine, the mean

Table 2

Number of phenols	Smallest difference[a]		Largest difference[b]		Mean difference[c]	
	Midrange	Paired-*t*	Midrange	Paired-*t*	Midrange	Paired-*t*
24	0.010	0.100	0.661	0.763	0.208	0.197
21	0.031	0.004	0.325	0.375	0.142	0.138
9[d]	0.0002	0.037	0.048	0.412	0.111	0.114

[a] Refers to the compound for which there is the smallest absolute difference between literature and estimated log *P* value.
[b] Refers to compound for which there is the largest absolute difference between literature and estimated log *P* value.
[c] Mean difference between literature and estimated log *P* value.
[d] Phenols that are mono-, di-, tri- or tetrasubstituted with chlorine or bromine.

absolute difference decreases to about 0.11 of a unit using either the midrange or the p-value method of calculation.

Acknowledgments

David Nurok thanks Whatman Inc. for a gift of KC_8F reversed-phase thin-layer plates.

References

[1] J.C. Dearden and G.M. Bresnen, Quant. Struc.–Act. Relat., 7 (1988) 133.
[2] T. Fujita, J. Iwasa and C.J. Hansch, Am. Chem. Soc., 86 (1964) 5175.
[3] R.F. Rekker, The Hydrophobic Fragment Constant, Elsevier, New York, 1977.
[4] R. Kaliszan, Quantitative Structure–Chromatographic Retention Relationships; Chemical Analysis, Vol. 93, Wiley, New York, 1987, Ch 11.
[5] J.L. Biagi, M.C. Guerra, A.M. Barbaro, S. Barbieri, M. Recanatini, P.A. Borea and M.C. Pietrogrande, J. Chromatogr., 498 (1990) 179.
[6] E. Soczewinski and C.A. Wachtmeister, J. Chromatogr., 7 (1962) 311.
[7] P.J. Schoenmakers, H.A.H. Billiet, R. Tijssen and L. De Galan, J. Chromatogr., 149 (1978) 519.
[8] E. Soczewinski, Anal. Chem., 41 (1969) 179.
[9] E. Soczewinski, J. Chromatogr., 388 (1987) 91.
[10] S. Habibi-Goudarzi, K.J. Ruterbories, J.E. Steinbrunner and D. Nurok, J. Planar Chromatogr., 1 (1988) 161.
[11] A.D. Kossoy, D.S. Risley, R.M. Kleyle and D. Nurok, Anal. Chem., 64 (1992) 1345.
[12] M.S. Mirlees, S.J. Moulton, C.T. Murphy and P.J. Taylor, J. Med. Chem., 19 (1976) 615.
[13] J. Sherma, Practice and Applications of Thin Layer Chromatography on Whatman KC_{18} Reversed Phase Plates, TLC Technical Series, Whatman Inc., Clifton, NJ, 1 (1981) 5.
[14] J.E. Garst and W.C. Wilson, J. Pharm. Sci., 73 (1984) 1616.
[15] C. Hansch and A. Leo, Pomona College Medicinal Data Base, MEDCHEM, ver. 3.53, Claremont, CA.
[16] A. Leo and C. Hansch, J. Org. Chem., 36 (1971) 1539.
[17] W. Klein, W. Kordel, M. Weiss and H.J. Poremski, Chemosphere, 17 (1988) 361.