Water–Solvent Partition Coefficients and Δ Log P Values as Predictors for Blood–Brain Distribution; Application of the Akaike Information Criterion

MICHAEL H. ABRAHAM,[1] WILLIAM E. ACREE JR.,[2] ALBERT J. LEO,[3] DAVID HOEKMAN,[4] JOSEPH E. CAVANAUGH[5]

[1]Department of Chemistry, University College London, 20 Gordon Street, London WC1H OAJ, UK

[2]Department of Chemistry, University of North Texas, 1155 Union Circle Drive #305070, Denton, Texas 76203-5017

[3]BioByte Corp., 201 W. Fourth St., Claremont, California 91711

[4]David Hoekman Consulting, Inc., 107 NW 82nd, Seattle, Washington 97117

[5]Department of Biostatistics C22-GH, College of Public Health, University of Iowa, 200 Hawkins Drive, Iowa City, Iowa 52242-1009

Received 26 June 2009; revised 8 October 2009; accepted 8 October 2009

ABSTRACT: It is shown that log P values for water–alkane or water–cyclohexane partitions, and the corresponding Δ log P values when used as descriptors for blood–brain distribution, as log BB, yield equations with very poor correlation coefficients but very good standard deviations, S from 0.25 to 0.33 log units. Using quite large data sets, we have verified that similar S-values apply to predictions of log BB. A suggested model, based on log P for water–dodecane and water–hexadecane partition coefficients, has 109 data points and a fitted $S = 0.254$ log units. It is essential to include in the model an indicator variable for volatile compounds, and an indicator variable for drugs that contain the carboxylic group. A similar equation based on water–chloroform partition coefficients has 83 data points and a fitted $S = 0.287$ log units. We can find no causal connection between these log P values and log BB in terms of correlation or in terms of chemical similarity, but conclude that the log P descriptor will yield excellent predictions of log BB provided that predictions are within the chemical space of the compounds used to set up the model. We also show that model based on log P(octanol) and an Abraham descriptor provides a simple and easy method of predicting log BB with an error of no more than 0.31 log units. We have used the Akaike information criterion to investigate the most economic models for log BB. © 2009 Wiley-Liss, Inc. and the American Pharmacists Association J Pharm Sci 99:2492–2501, 2010

Keywords: blood–brain distribution; water–solvent partition coefficients; Δ log P; linear free energy relationship; Akaike information criterion

INTRODUCTION

In a series of papers, Collander[1–5] showed that when water to solvent partition coefficients of a solute in a given system, $P(s1)$, were plotted against water to solvent partition coefficients in another system, $P(s2)$, as log P values, a straight line for all solutes is obtained only if the two systems are chemically very close. More usually, a series of straight lines is obtained, one line for each homologous series of solutes, as in Eq. (1). Whereas the value of the slope, a, is roughly constant, the intercept, b, depends on the functional group of the homologous series and reflects the hydrogen bond properties of the functional group.

$$\text{Log}\, P(s1) = a \log P(s2) + b \qquad (1)$$

Hansch,[6] Leo and Hansch,[7] and Leo et al.[8] extended the analysis of Collander and used Eq. (1) to group

---

Additional Supporting Information may be found in the online version of this article.

Abbreviations: P, water to solvent partition coefficient; $\Delta \log P$, $\log P$ for water to octanol minus $\log P$ for water to alkane; $\sum I_{\mathrm{H}}$, the contribution to $\Delta \log P$ from hydrogen bonding; BB, the blood to brain distribution coefficient; n, the number of data points in a regression; r, the correlation coefficient; S, the standard deviation; AE, the average error; AAE, the absolute average error; RMSE, the root mean square error; PSD, the predictive standard deviation; F, the Fischer F-statistic; SP, the dependent variable in a regression equation; AIC, the Akaike information criterion; $\mathbf{E}$, $\mathbf{S}$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{V}$, the Abraham descriptors; LFER, linear free energy relationship; VCs, volatile compounds; K_{s} and K_{w}, gas to solvent and gas to water partition coefficients; I_{v} and I_{c}, indicator variables for VCs and carboxylic acids; θ or $\cos\theta$, the Ishihama and Asakawa nearness parameter; D', the Abraham and Martins nearness parameter; PCA, principal components analysis.

Correspondence to: Michael H. Abraham (Telephone: 44-20-7679-4639; Fax: 44-20-7679-7463; E-mail: m.h.abraham@ucl.ac.uk)

Journal of Pharmaceutical Sciences, Vol. 99, 2492–2501 (2010)

solutes into hydrogen bond acids and hydrogen bond bases, arguing, as did Collander, that it is the solute hydrogen bond properties that lead to different values of b for homologous series. They[6–8] also advocated the water to octanol system as a "standard" partitioning system, and this has been adopted, displacing systems such as water to olive oil[9,10] or water to oleyl alcohol[11] as models for biological processes.

Leo and Hansch[7] had shown that the a- and b-values in Eq. (1) differed as between cyclohexane and heptane, although the number of solutes examined was rather small, 56 and 21, respectively. Seiler,[12] however, suggested that data on cyclohexane and the n-alkanes could all be combined, although again the number of solutes studied was small, from 10 to 35 only. When the alkane system was compared to the water to octanol system, Seiler obtained Eq. (2), where $\sum I_{\mathrm{H}}$ represents the contribution to $\Delta \log P$(alkane) due to solute hydrogen bonding.

$$\Delta\,\mathrm{Log}\,P(\text{alkane}) = \log\,P(\text{octanol}) - \log\,P(\text{alkane}) = \sum I_{\mathrm{H}} - b \quad (2)$$

The first use of the $\Delta \log P$ parameter in the correlation of blood to brain distribution, as log BB, was by Young et al.[13] who put forward Eq. (3)

$$\mathrm{Log\,BB} = -0.485\,\Delta\,\log\,P(\text{cyclohexane}) + 0.889, \quad n = 20,\ r^2 = 0.692,\ S = 0.439,\ F = 40.2 \quad (3)$$

Here, n is the number of data points, r is correlation coefficient, S is standard deviation, and F is F-statistic. Although the number of data points is not very large, Eq. (3) has provoked considerable interest.

Calder and Ganellin[14] used data[15] on five more compounds which they added to the 20-compound data set of Young et al.[13] and obtained Eq. (4) (statistics calculated in this work). Although Eq. (4) is a two-descriptor equation that requires only log P(octanol) and log P(cyclohexane), it cannot be regarded as particularly good.

$$\mathrm{Log\,BB} = -0.409\,\Delta\,\log\,P(\text{cyclohexane}) + 0.547, \quad n = 25,\ r^2 = 0.533,\ S = 0.497,\ F = 26.3 \quad (4)$$

Better results were found by Abraham et al.[16] who used a data set of 32 compounds and obtained Eq. (5)

$$\mathrm{Log\,BB} = -0.318\,\Delta\,\log\,P(\text{cyclohexane}) + 0.225, \quad n = 32,\ r^2 = 0.784,\ S = 0.272,\ F = 108.6 \quad (5)$$

Kenny et al.[17] determined partition coefficients in the water to octanol and water to hexadecane systems, and obtained $\Delta \log P(16)$ for 44 compounds, where $P(16)$ refers to the water to hexadecane partition coefficient in Eq. (2). These values were then used to derive predictions of $\Delta \log P(16)$ for further compounds, using a combination of molecular electrostatic potentials and a fragmentation scheme. Finally, a set of eighteen log BB values from the literature was regressed against predicted $\Delta \log P(16)$ values to give Eq. (6)

$$\mathrm{Log\,BB} = -0.475\,\Delta\,\log P(16)\text{calc} + 1.490, \quad n = 18,\ r^2 = 0.877,\ S = 0.321,\ F = 114.3 \quad (6)$$

It is known that water to alkane and water to cyclohexane partition coefficients, and hence the corresponding $\Delta \log P$ values, are not quite the same.[18] Nevertheless, the variation of the slopes and intercepts in Eqs. (3–6) is so large, that it seems very likely that the number of compounds in the equations is too small to yield reliable general equations for log BB. As pointed out by Chadha et al.[19] small data sets cannot be expected to yield reliable predictions of log BB for much larger data sets.

Kenny et al.[17] also explored the use of log P(octanol)calc and log P(16)calc to fit log BB values, but for 18 drug compounds obtained poorer fits than with the $\Delta \log P(16)$ parameter; values were $r^2 = 0.66$ and $S = 0.54$ and $r^2 = 0.82$ and $S = 0.39$, respectively.

More recently, Zerara et al.[20] have explored the correlation between log BB and calculated values of log P(octanol), log P(cyclohexane), and log P(alkane), and showed that for 37 compounds the latter two equations were much better than the equation in log P(octanol). They reported that for the correlation against log P(cyclohexane)calc, $r^2 = 0.78$ and $S = 0.46$ and for the correlation against log P(alkane)calc, $r^2 = 0.78$ and $S = 0.49$ log units. Both VCs and drugs were included in the correlations.

In a development that has not been pursued, Abraham et al.[21] used a method that they describe as the log P plus system. They used log P(octanol), as the most widely available log P descriptor, in combination with the Abraham descriptors and obtained Eq. (7); **A** and **B** are the compound hydrogen bond acidity and basicity.

$$\mathrm{Log\,BB} = 0.055 + 0.203\,\log\,P(\text{octanol}) - 0.507A - 0.500B, \quad n = 49,\ r^2 = 0.900,\ S = 0.201,\ F = 136.1 \quad (7)$$

An extensive set of log BB values for 207 compounds was compiled by Abraham et al.[22] who obtained a good correlation against the Abraham descriptors, if two particular factors were taken into account. First, log BB values for volatile compounds, VCs, could only be included in a general equation if an indicator variable, I_{v} was used. This took the value -0.438; in other words, log BB values for VCs are systematically