UC San Diego
UC San Diego Previously Published Works

Title
Connection between chromatographic data and biological data

Permalink
https://escholarship.org/uc/item/4dh3053v

Journal
Journal of Chromatography B: Biomedical Sciences and Applications, 745(1)

ISSN
03784347

Authors
Abraham, Michael H
Gola, Joelle M.R.
Kumarsingh, Rachel
et al.

Publication Date
2000-08-01

DOI
10.1016/S0378-4347(00)00130-4

Data Availability
The data associated with this publication are within the manuscript.

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Journal of Chromatography B **745**(1):103-115, 2000

The connection between chromatographic data and biological data

Michael H. Abraham,[a] Joelle R. M. Gola,[a] Rachel Kumarsingh,[a] Enrique Cometto-Muñiz [b] and William S.Cain [b]

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

[b] *Chemosensory Perception Laboratory, Department of Surgery (Otolaryngology), University of California, San Diego, Mail Code 0957, La Jolla, CA 92093-097, USA*

Abstract

There are no previous references to the direct use of GLC data in the correlation of biological processes, but we show that GLC retention data can be used in the correlation of several such processes involving gaseous solutes. There are a number of reports of RP-HPLC and MEKC data being used in the correlation of biological processes, but they are mostly restricted as to the number and type of solute studied. We show that if chromatographic data are used to obtain solvation descriptors for solutes, and if these descriptors are then used in the correlation of biological processes, that this indirect connection is a much more powerful and generally applicable method than is the direct connection between chromatographic data and biological data.

Keywords: GLC, RP-HPLC, MEKC, solvation descriptors, nasal pungency thresholds, eye irritation thresholds, blood-brain distribution, brain perfusion, skin permeation, aqueous narcosis.

Contents

1. Introduction

2. The solvation parameter model

3. Direct connections between chromatographic data and biological data

3.1 GLC data

3.2 RP-HPLC and MEKC data

4. Indirect connections between chromatographic data and biological data

4.1 GLC data

4.2 RP-HPLC data

5. Conclusions

6. References

1. Introduction

There are many biological processes that are difficult to study. For example, to obtain a value for the distribution of a compound between blood and brain requires the sacrifice of several rats, and it is clearly both unethical and uneconomic to determine such distribution for large numbers of compounds. One general method for estimating data on processes that are difficult to study is to use model processes that can be studied rather easily. Thus water-octanol partition coefficients, as log P_{oct}, or the Δlog P parameter defined [1] by Eq.(1) have been used [2] to estimate blood-brain distribution; log P_{cyc} refers to the water-cyclohexane partition.

$$\Delta \log P = \log P_{oct} - \log P_{cyc} \qquad (1)$$

Chromatographic processes are also easy to study, and can be automated so that rather large numbers of compounds can be handled. Hence it might be expected that direct connections between chromatographic processes and biological processes would be a fruitful area to study, see Fig.1. The general procedure is to set up a correlation equation on the lines of Eq.(2), where *BR* is a biological response and *G* is a set of the corresponding chromatographic data for a series of compounds. Then further measurements of *G* can be used to estimate values of *BR*.

$$\log BR = g \log G + f \tag{2}$$

Gas liquid chromatographic (GLC) data on a given stationary phase might be suitable for the modelling of a biological process that involves the transfer of a solute from the gas phase to a biological phase, whereas HPLC, especially reversed phase HPLC, should be more suitable for modelling biological processes in which solutes are transferred from one condensed phase to another. In any event, we can find no published study in which GLC data have been used directly to correlate biological data, but we report on our own work in this area. On the other hand, there are several studies in which RP-HPLC capacity factors, as log *k'*, have been used in this way. It must be said, however, that results have been disappointing. Studies have nearly always been restricted to a rather small number of solutes, so that no generality can be established, and even then the obtained regressions have not been particularly good, as we shall detail later.

2. The solvation parameter model

It seems that it is quite difficult to choose a particular GLC or RP-HPLC system that closely matches a given biological process. There is, however, a way round this impasse, that involves the use of chromatographic data in an indirect manner. We have devised [3] two general solvation equations that have been used to correlate a large number of physicochemical and biochemical data,

$$\log SP = c + r\mathrm{R}_2 + s\pi_2^{\mathrm{H}} + a\Sigma\alpha_2^{\mathrm{H}} + b\Sigma\beta_2^{\mathrm{H}} + l\log \mathrm{L}^{16} \quad (3)$$

$$\log SP = c + r\mathrm{R}_2 + s\pi_2^{\mathrm{H}} + a\Sigma\alpha_2^{\mathrm{H}} + b\Sigma\beta_2^{\mathrm{H}} + v\mathrm{V}_x \quad (4)$$

In Eq.(3) and Eq.(4), the dependent variable log *SP* refers to some property of a series of solutes in a fixed phase, and can be a physicochemical property, or a biological property, *BR*. The first four descriptors are measures of the tendency of a solute to undergo various solute-solvent interactions, all of which are energetically favorable, i.e., exoergic [3]. R_2 is an excess molar refraction, π_2^{H} is the solute dipolarity-polarizability, $\Sigma\alpha_2^{\mathrm{H}}$ and $\Sigma\beta_2^{\mathrm{H}}$ are the solute effective hydrogen bond acidity and basicity. The other two descriptors are log L^{16} where L^{16} is [4] the solute gas-hexadecane partition coefficient at 298K, defined as in Eq.(5), and V_x is the McGowan characteristic volume [5] in units of ($\mathrm{ml.mol^{-1}}/100$). The log L^{16} term and the V_x term are composite quantities that include the unfavorable endoergic cavity effect defined as the disruption of solvent-solvent interactions when a solute is placed in a solvent, and the favorable general dispersion interaction between solute and solvent.

$$L = \text{[molar conc. in solution]/[molar conc. in gas phase]} \quad (5)$$

Of the five descriptors in Eq.(3) and Eq.(4), the R_2 descriptor can be obtained from the refractive index of the solute, if liquid, or from addition of fragment values, and the V_x descriptor can always be calculated from structure [5]. In order to obtain the remaining four descriptors for a given solute, we proceed as follows. Various equations for physicochemical processes can be set up on the lines of Eq.(3) and Eq.(4). Then if values of log *SP* for the given solute are known for several such equations, the missing descriptors can be assigned as those that give the best fit of observed and calculated log *SP* values.

This procedure has been outlined in some detail for the determination of descriptors for a number of terpenes [6], and we can illustrate the method using α-terpinene as an example. Regression coefficients for Eq.(3) applied to GLC data are given in Table 1, and the experimental GLC data for α-terpinene are in Table 2. Although two of the GLC equations have a b-coefficient, they are both very small and not useful for the calculation of the solute $\Sigma\beta_2^H$ descriptor. However the descriptors π_2^H , $\Sigma\alpha_2^H$ and log L^{16} can be obtained as the 'best-fit' values, that is the values that best reproduce the experimental data, as shown in Table 2. Note that the descriptor R_2 is easily obtained from the liquid refractive index. For the case of α-terpinene, it was reasonably easy to assign a value of 0.15 for $\Sigma\beta_2^H$ by analogy with other terpenes. The final set of calculated descriptors was: R_2 = 0.526, π_2^H = 0.25, $\Sigma\alpha_2^H$ = 0.00, $\Sigma\beta_2^H$ = 0.15, log L^{16} = 4.715 and V_x = 1.3230 (mol dm^{-3} /100). In general the estimation of $\Sigma\beta_2^H$ will not be possible, and at least one equation with a large b-coefficient will be required for its' determination. Most RP-HPLC equations for C18 stationary phases with aqueous acetonitrile or aqueous methanol mobile phases have large b-coefficients and so a

combination of GLC and RP-HPLC will provide sufficient experimental data for the determination of descriptors for a given solute.

With the availability of RP-HPLC stationary phases that have a variety of functional groups, it might be possible to select a set of RP-HPLC phases that give rise to considerably different coefficients in Eq.(4), especially in combination with different mobile phases. Such a set of RP-HPLC systems could then be used to determine the descriptors π_2^H , $\Sigma\alpha_2^H$, and $\Sigma\beta_2^H$ in Eq.(4). We have made some progress in this area, helped considerably by the use of gradient elution rather than isocratic elution [7,8], and it is possible that the RP-HPLC method could provide a rapid experimental method for the determination of descriptors. Of course, other combinations of experimental methods can also be used, such as GLC together with water-solvent partition coefficients.

Hence by the above procedure, using GLC and RP-HPLC data that are relatively easy to obtain, we can derive the required equations for the calculation of descriptors. Then if these solute descriptors are used in correlation equations for biological properties, we have started with GLC and RP-HPLC data and have, in effect, used them indirectly for the correlation of biological properties, as shown in Fig.1.

The aims of this paper are firstly to explore the direct connection between chromatographic data and biological data, and secondly to survey the use of chromatographic data in the determination of descriptors for correlation analysis. Such descriptors can be the solvation descriptors in Eq.(3) and Eq.(4), or can be any other type of descriptor calculated from chromatographic data.

3. Direct connections between chromatographic data and biological data

3.1 GLC data

As we have mentioned, we can find no published account of the direct connection between GLC data and biological data. However, we have ourselves explored this area, and have attempted to select GLC stationary phases that might be suitable models for two particular biological processes. These involve the transfer of solutes from the gas phase to a receptor or receptor area and so might be modelled by GLC. The first process is that of nasal pungency, defined [9] as irritation and burning sensations through chemical stimulation of the trigeminal nerve. The second process is that of eye irritation, again involving stimulation of the trigeminal nerve [10]. The two together are often called 'common chemical sense' or 'sensory irritation'. They are very important as regards air quality [11]; about half of the threshold limit values for chemicals set by the American Conference of Governmental Industrial Hygienists are based on sensory irritation [12]. Cometto-Muniz and Cain have systematically determined nasal pungency thresholds *(NPT)* and eye irritation thresholds *(EIT)* in human subjects for a variety of volatile compounds. The determinations are laborious and time consuming, and so some method of estimating *NPT* and *EIT* values would be of considerable use.

Following a preliminary analysis of *NPT* values [13], we reasoned that a stationary phase that was dipolar, basic, and also somewhat acidic, might be a good model. We chose tetraethyleneglycol *(TEG),* and also dibutylsuberate *(DBS),* as an alternative model and determined relative retention times at 50°C on each phase; decane was the standard for *TEG* and hexane was the standard for *DBS*. The *NPT* values are defined so that the larger the value in ppm, the less potent is the compound. We therefore regressed log *(1/NPT)* against log tr for compounds on each phase; tr is the relative

corrected retention time. The *NPT* values are all those that we have determined to date [14,15].

$$\log (1/NPT) = -4.36 + 1.02 \log tr(TEG) \qquad (6)$$

$n = 49$, $r^2 = 0.784$, $sd = 0.55$, $F = 171$

$$\log (1/NPT) = -4.87 + 1.13 \log tr(DBS) \qquad (7)$$

$n = 49$, $r^2 = 0.857$, $sd = 0.45$, $F = 281$

In the above equations, n is the number of data points (compounds), r is the correlation coefficient, sd is the standard deviation and F is the F-statistic. Both Eq.(6) and Eq.(7) are reasonable, in that they allow an estimation of log *(1/NPT)* to about 0.50 log units in a total range of 5 log units, from just a single determination of a relative retention time. However, the full Eq.(3) correlates log *(1/NPT)* to 0.30 log units, see below, and so is to be preferred where ever possible. Plots of log *(1/NPT)* against log *tr(TEG)* and log *tr(DBS)* are shown in Fig.2 and Fig.3; there seems to be just random scatter about the line of identity.

We also used relative retention times on the same two stationary phases to correlate eye irritation thresholds, using not only our own data on human subjects [16], but also results from the Draize rabbit eye test [17] all corrected to ppm,

$$\log (1/EIT) = -4.40 + 1.05 \log tr(TEG) \qquad (8)$$

$n = 54$, $r^2 = 0.864$, $sd = 0.48$, $F = 330$

$$\log (1/EIT) = -5.10 + 1.23 \log tr(DBS) \quad (9)$$

$n = 54$, $r^2 = 0.733$, $sd = 0.67$, $F = 143$

Once again, the correlations are reasonable in that just one GLC determination will yield a value of log *(1/EIT)* to around 0.6 log unit in a total range of 6 log units. There is random scatter about the line of identity in Eq.(8) and Eq.(9), see Figs. 4 and 5.

Although we have shown that GLC retention data can be used in the correlation of *NPT* and *EIT* values, the sd values in Eq.(6-9) are all rather high for the equations to be of much practical use. It might be possible to design stationary phases that are better models, but the indirect approach shown in Fig.1 seems the better possibility.

3.2 RP-HPLC and MEKC data

In contrast to the GLC work, there are a number of reports of the correlation of biological properties with RP-HPLC capacity factors, as log *k'*. Altomare et al. [18] were one of the first to relate capacity factors to biological activity, although they first transformed log *k'* values into log *P* values, where *P* is the equivalent to a water-octanol partition coefficient, P_{oct}. They also used log k_w values directly as a descriptor in a QSAR for antimicrobial activity [19].

The advent [20] of RP-HPLC stationary phases based on immobilized artificial membranes *(IAM)* has led to their use as models for a number of systems. Alvarez et al. [21] in a very early paper, correlated log *k'* values obtained using an IAM column with coefficients for permeation through human skin, *Kp*. For eight alkan-1-ols, they obtained a good correlation ($r^2 = 0.989$) between log *k'* using an aqueous mobile phase and log *Kp*. Alvarez et al. [21] also determined log *Kp* values for 15 steroids and

showed that for eight structurally related steroids there was a reasonable correlation (r^2 = 0.833) with log *k'* using 30% acetonitrile mobile phase. They also gave a correlation of log *Kp* for beta-blockers, but for four compounds only. In latter papers [22, 23], the same group correlated permeability through Caco-2 cells and rat intestinal absorption with IAM capacity factors; results were only shown graphically, however, with no numerical data. Log *k'* values on an *IAM* column with an aqueous buffer mobile phase were used [24] to correlate binding of 23 solutes to dimyristoylphosphatidylcholine *(DMPC)* liposomes as log *K(DMPC),*

$$\log K(DMPC) = 0.496 + 0.994 \log k'(IAM) \qquad (10)$$

$n = 23, \quad r^2 = 0.823$

Kaliszan et al. [25] used log k' values with acetonitrile (10%) / water (90%) mobile phase as a measure of the hydrophobicity of β-adrenolytes, and to correlate various pharmacokinetic parameters. Although the regressions were quite good, they are restricted to this particular solute set, and refer to only 6-14 solutes. The same group [26] used log k' values on a commercial *IAM* column with acetonitrile (5%) / water (95%) mobile phase to correlate the water-skin permeation coefficient, *Kp*, of a group of steroids, Eq.(11) and of a set of phenols, Eq.(12), with *Kp* in units of cm s^{-1}.

$$\log Kp = -10.19 + 1.77 \log k'(IAM) \qquad (11)$$

$n = 10, \quad r^2 = 0.887$

$$\log Kp = -6.09 + 1.05 \log k'(IAM) \qquad (12)$$

n = 14, r^2 = 0.584

Various parabolic relationships were explored, but, again, the data sets are too restricted for any general conclusions to be drawn.

More recently, La Rotonda et al. [27] measured log *k'* values for 12 structurally varied drugs on an IAM column at with different acetonitrile / water mobile phases and extrapolated to 100 % water to obtain log k_w values. La Rotunda et al.[27] reported that they could not find any correlation between *k'* and *Kp* values for the 12 drugs, although for 10 of the drugs they did find a correlation between log *Kp* and a parameter Δlog k_w defined as the difference between measured values of log k_w and those predicted from a plot of log k_w against log P_{oct}. The significance of such a correlation seems rather unclear.

Turowski and Kaliszan [28] have synthesised a new type of stationary phase that consists of keratin adsorbed onto silica. Since the stratum corneum layer of human skin contains a large proportion of keratin, log *k'* values on a keratin column, log *k'(keratin)* , might be a be a very useful predictor of skin permeation. An equation that incorporates both log *k'(keratin)* and log *k' (IAM)* was suggested for 17 mostly phenolic compounds.

$$\log Kp = -6.558 + 1.920 \log k'(IAM) - 1.039 \log k'(keratin) \qquad (13)$$

n = 17, r^2 = 0.869, sd = 0.40

A number of RP-HPLC stationary phases were also synthesised from collagen and various silica supports [29]. Values of log k' were obtained on collagen bound to an

amino-silica support with a mobile phase of 6% isopropanol / 94 % water, buffered at pH 4.2, for the set of mostly phenolic compounds but no comparison with log *Kp* was given. We have used the reported data of Turowski and Kaliszan [29] to generate simple regressions of log *Kp* for the set of phenolic compounds (we exclude baclofen and phenylalanine which will charged or partly charged at pH 4.2)

$$\log Kp = -6.44 + 1.48 \log k'\ (IAM) \qquad (14)$$

$n = 15,\ \ r^2 = 0.779,\ \ sd = 0.42,\ \ F = 45.7$

$$\log Kp = -5.70 + 1.06 \log k'\ (keratin) \qquad (15)$$

$n = 15,\ \ r^2 = 0.215,\ \ sd = 0.78,\ \ F = 3.6$

$$\log Kp = -6.01 - 0.69 \log k'\ (collagen) \qquad (16)$$

$n = 15,\ \ r^2 = 0.052,\ \ sd = 0.86,\ \ F = 0.7$

Rather extraordinarily, there is no correlation of skin permeation with the log *k'* values on the keratin or collagen stationary phases at all, so that the earlier IAM columns seem more useful in this respect.

Salminen et al. [30] used log *k*' values on an *IAM* column with an aqueous mobile phase at pH 7.4 as a descriptor in the correlation of blood-brain distribution for 26 diverse compounds; *BB* is defined as,

$$BB = [\text{conc in brain}]/[\text{conc in blood}] \qquad (17)$$

Eq.(18) was obtained after five outliers were omitted,

$$\log BB = 1.28 + 0.58 \log k'(IAM) + 0.89\, I_2 - 0.01\, V_m \qquad (18)$$

$n = 21$, $r^2 = 0.848$, $sd = 0.27$, $F = 32$

The additional descriptors in Eq.(18) are I_2 taken as zero except for compounds with an amino-nitrogen when I_2 is unity, and V_m a molecular volume.

Results to date thus show that the direct connection between RP-HPLC data and biological properties is feasible, although in only one or two cases have sufficient numbers of structurally diverse compounds been studied to indicate any generality of solute type.

Khaledi et al. [31] have briefly investigated the use of micellar electro-kinetic chromatographic (MEKC) systems for the generation of descriptors for biological properties. They showed that there were reasonable correlations of binding to serum albumin protein and intestinal absorption in the rat against log *k'(MEKC)* for nine and eight steroids, respectively. Khaledi et al. [31] pointed out that an advantage of MEKC is that the characteristics of the system can easily be varied by change in the amount and type of surfactant used. Poole and Poole [32] and Roses et al. [33] have characterised a number of (MEKC) systems, and have shown how they can be quite finely 'tuned' through the variation in the surfactant composition. In Table 4 are given details of the ratio of coefficients in Eq.(4) when applied [33] to log k values for MEKC using mixtures of sodium dodecyl sulfate (SDS) and Brij 35 as the surfactant. Roses et al. [33] attempted to match the ratios with those for various biological systems, and found that an equation for tadpole narcosis by aqueous solutes gave similar ratios [34],

see Table 4. Values of log *k(MEKC)* with 50 mM SDS and 10 mM Brij 35 were available for 13 solutes for which tadpole narcosis had been studied; the correlation with log $(1/C_{nar})$ is,

$$\log (1/C_{nar}) = 2.19 + 1.25 \log (1/k(MEKC)) \qquad (19)$$

$n = 13$, $r^2 = 0.949$, $sd = 0.17$, $F = 204$

In Eq.(19), C_{nar} is the minimum concentration of solute, in mol. dm^{-3} required to bring about narcosis. The equation is statistically very good, although limited to only thirteen solutes out of the 114 solutes studied [34].

4. Indirect connections between chromatographic data and biological data

4.1 GLC data

The first such connection was put forward by Laffort et al. [35,36]. Retention indices were obtained for 240 solutes on five stationary phases at 120°C, and were used [37] to calculate five solubility factors for each solute:

α an apolar factor proportional to molecular volume,

ω an orientation factor connected to the dipole moment,

ε an electron factor connected to molar refraction,

π an acidity factor,

β a basicity factor.

Laffort and Patte [37] then used these solubility factors to analyse human olfactory thresholds and odorous discrimination in man, showing that the ε-factor always seemed to be important. They also carried out a quantitative analysis of retention indices obtained from retention times of 12 solutes across the frog olfactory mucosa,

$$I(frog) = 876 - 1211\ \omega - 764\ \pi + 5844\ \beta \qquad (20)$$

We have used the original relative retention times to obtain the equivalent equation, Eq.(21) and to calculate the statistical fit. However, in both Eq.(20) and Eq.(21) there are too few data points to draw any real conclusions.

$$\log tr(frog) = 0.71 - 0.99\ \omega - 0.63\ \pi + 4.80\ \beta \qquad (21)$$

$n = 12$, $r^2 = 0.904$, $sd = 0.27$, $F = 25$

We have shown, above, that the solvation descriptors in Eq.(3) can be obtained from chromatographic data, so that the application of Eq.(3) to the correlation of biological data corresponds to the indirect connection shown in Fig.1. Several biological processes have been examined in this way.

One of the first such processes to be studied was that of upper respiratory tract irritation in mice, the most recent and extensive equation [38] being,

$$\log (1/RD_{50}) = -7.049 + 1.437\ \pi_2^H + 2.316\ \Sigma\alpha_2^H + 0.774 \log L^{16} \qquad (22)$$

$n = 58$, $r^2 = 0.840$, $sd = 0.35$

In Eq.(22), RD_{50} is the vapour concentration of the irritant in ppm that brings about a 50% decrease in the rate of respiration. Thus $1/RD_{50}$ is a measure of the solute potency. The 58 solutes used in Eq.(22) were all 'nonreactive' in that their irritancy effect is induced through essentially a physical mechanism [38-40].

The *NPT* and *EIT* data in Eq.(6-9) have also been correlated through the general Eq.(3). For exactly the same data set of 48 log *(1/NPT)* values we now find,

$$\log (1/NPT) = -7.992 + 1.702\,\pi_2^H + 3.361\,\Sigma\alpha_2^H + 0.923\,\Sigma\beta_2^H + 0.847\,\log L^{16}$$

$$n = 48,\ \ r^2 = 0.936,\ \ sd = 0.31,\ \ F = 157 \qquad (23)$$

Eq.(23) is not quite as good statistically as our original equation [13], but contains more solutes, specifically five terpenes. However it is very much better than Eq.(6) or Eq.(7) which have sd values of 0.55 and 0.45 log units respectively. Furthermore, Eq.(23) enables the exact factors that influence nasal pungency thresholds to be deduced. We can estimate the contribution of each term in Eq.(23) by multiplying the coefficient by the mean value of the descriptor in the 48 compound set. Then the percentage contribution is: rR_2 (0), $s\pi_2^H$ (19), $a\Sigma\alpha_2^H$ (10), $b\Sigma\beta_2^H$ (9), and $l\log L^{16}$ (62), so that the final term in Eq.(23) is by far the most important.

We have already [15] analysed the set of 54 eye irritation thresholds to give the correlation equation,

$$\log (1/EIT) = -\,7.918 - 0.482\,R_2 + 1.420\,\pi_2^H + 4.025\,\Sigma\alpha_2^H + 1.219\,\Sigma\beta_2^H$$
$$+\ 0.853\,\log L^{16} \qquad (24)$$

$$n = 54,\ \ r^2 = 0.930,\ \ sd = 0.36,\ \ F = 124$$

Eq.(24) is substantially better than Eq.(8) or Eq.(9) which have sd values of 0.48 and 0.67 log units respectfully. The percentage contribution of the terms in Eq.(24) is: rR_2 (3), $s\pi_2^H$ (15), $a\Sigma\alpha_2^H$ (9), $b\Sigma\beta_2^H$ (10), and $l\log L^{16}$ (63), so that again it is the final term that is the most important.

Comparison of the coefficients in Eq.(23) and (24) with the regression coefficients for *TEG* and *DBS* shown in Table1, shows why the two stationary phases are not good models for *NPT* and *EIT* values. In order for there to be a good linear regression between two sets of data, it is not necessary for the coefficients in the two equations to be the same; only the ratio of coefficients needs to be the same. In Table 3 are given ratios relative to the *l*-coefficient, and it is quite clear that neither *TEG* nor *DBS* will be good models for nasal pungency thresholds or eye irritation thresholds.

However, quite by chance, it appears that the ratios of coefficients for *DBS* are very close indeed to those for sensory irritation in mice. Hence log $(1/RD_{50})$ should be well correlated with *log tr(DBS)*. This is indeed the case. For 34 nonreactive solutes for which we have the required data [38-40] we find,

$$\log (1/RD_{50}) = -4.609 + 1.037 \log tr(DBS) \quad (25)$$

$n = 34$, $r^2 = 0.915$, $sd = 0.20$, $F = 345$

Eq.(25) is much better than Eq.(6-9) and is even better than the multiple regression Eq.(23). However, Eq.(25) deals with only 34 compounds, and a more complete analysis is required for a detailed comparison.

Examination of Table 3 shows how difficult it is to design model stationary phases. From Eq.(23) and Eq.(24) we knew beforehand that phases to model *NPT* or *EIT* must have moderate *s*- and *b*-coefficients, and have a large *a*-coefficient. Our choice of model phase, *TEG*, gives rise to such coefficients, exactly as we predicted, but the numerical values of the coefficients and their ratios render *TEG* not suitable as a model for *NPT* or *EIT* values. On the other hand, the excellent agreement between coefficient ratios for *DBS* and RD_{50} values is a tribute to the value of serendipity!

4.2 RP-HPLC data

We have suggested that RP-HPLC data from several different systems could be used to obtain the descriptors in Eq.(4), especially for rather involatile compounds. However, to date, we have only used RP-HPLC data in conjunction with other data, especially water-solvent partition coefficients, to calculate descriptors. Hence the following brief account of the application of Eq.(4) to biological systems [41] is not meant to be a rigorous illustration of the indirect method shown in Fig.1, but only an indication of the type of process that RP-HPLC data could help (indirectly) to model.

The passive transport of molecules across the blood-brain barrier can be measured in several ways, one of which is through the distribution shown in Eq.(13). For a very wide range of compounds log *BB* was well correlated [42] with our solvation descriptors,

$$\log BB = -0.038 + 0.198\ R_2 - 0.687\ \pi_2^H - 0.715\ \Sigma\alpha_2^H - 0.698\ \Sigma\beta_2^H + 0.995\ V_x \quad (26)$$

$n = 57$, $r^2 = 0.916$, $sd = 0.20$, $F = 99$

Eq.(26) covers a much wider range of compounds than the 'direct' Eq.(14), and gives a better fit as well. Furthermore, the indirect method using Eq.(4) has the great advantage that the individual factors that influence the passsive distribution can be identified and quantified. The main effects are of solute dipolarity, hydrogen-bond acidity and hydrogen-bond basicity that reduce passage from blood to brain, and solute volume that aids passage from blood to brain. Other equations for log *BB* have subsequently been published (see ref. 41), but none have any considerable advantage over Eq.(26).

Another measure of the ability of compounds to cross the blood-brain barrier is the rate of perfusion from the internal carotid artery into the brain. Unlike distribution studies, the time scale of perfusion experiments is very small, so reducing the possibility of biological degradation of the compound during the course of the experiment. Perfusion experiments can be made starting with the compounds in a saline solution, or in plasma or in blood (injected into the carotid artery). These may give rise to different perfusion rates and so it is essential to consider only rates from a fixed medium. Gratton et al.[43] used protein-free saline as the starting medium and obtained perfusion rates for neutral species, i.e. corrected for ionisation in the saline solution, as permeability surface area products, *PS* in $cm^3\ s^{-1}\ g^{-1}$. Values of log *PS* were reasonably well correlated [44] as shown in Eq.(27), although there are too few data points to come to any very definite conclusions.

$$\log PS = -1.21 + 0.77\ R_2 - 1.87\ \pi_2^H - 2.80\ \Sigma\beta_2^H + 3.31\ V_x \qquad (27)$$

$n = 18$, $r^2 = 0.953$, $sd = 0.48$, $F = 65$

It can be said, however, that the factors that influence rates of permeation from saline are quantitatively not the same as those that influence blood-brain distribution.

Another rate process that is of considerable importance is the permeation of human skin by aqueous compounds. We were able to construct [45] a reasonably general equation that covered a range of solute type (alcohols, phenols, steroids, etc.); *Kp* is the permeability coefficient in units of cm s^{-1},

$$\log Kp = -5.132 + 0.439\,R_2 - 0.498\,\pi_2^H - 1.478\,\Sigma\alpha_2^H - 3.442\,\Sigma\beta_2^H + 1.941V_x \quad (28)$$

$n = 53$, $r^2 = 0.958$, $sd = 0.21$, $F = 213$

Subsequently, Raevsky and Schaper [46] re-investigated the hydrogen bond contribution to skin permeation, but could only obtain equations for subgroups of compounds, (i) phenols and (ii) steroids and alcohols. Such equations are obviously of limited predictive use. Cronin et al. [47] analysed a much larger data set without the necessity of subdividing data. After omitting seven outliers they obtained a good equation for no less than 107 compounds, using as descriptors log P_{oct} and molecular weight.

$$\log Kp = -5.89 + 0.772 \log P_{oct} - 0.0103\,MW \quad (29)$$

$n = 107$, $r^2 = 0.859$, $sd = 0.39$, $F = 317$

A wide range of compounds has also been used in experiments on the narcosis of tadpoles by aqueous solutes [48-54], a useful bioassay for aqueous anesthesia. For no less than 114 solutes a reasonable correlation equation was obtained [34],

$$\log (1/C_{nar}) = 0.582 + 0.770\, R_2 - 0.696\, \pi_2^H + 0.243\, \Sigma\alpha_2^H - 2.592\, \Sigma\beta_2^O + 3.343\, V_x \quad (30)$$

$n = 114$, $r^2 = 0.909$, $sd = 0.34$, $F = 217$

The term in $\Sigma\alpha_2^H$ is not statistically significant, and the $\Sigma\beta_2^O$ parameter is the alternative hydrogen-bond basicity descriptor that takes into account the variable basicity of a few solutes. Bearing in mind that C_{nar} is found by visual inspection, Eq.(30) is as good as can be expected.

A large number of compounds was also used by Gunatilleka and Poole [55] in their study of non-specific toxicity of aqueous compounds to the fathead minnow, Eq. (31), and to the guppy, Eq.(32).

$$\log (1/LC_{50}) = 0.71 + 0.60\, R_2 + 0.36\, \Sigma\alpha_2^H - 3.15\, \Sigma\beta_2^O + 3.33\, V_x \quad (31)$$

$n = 110$, $r^2 = 0.939$, $sd = 0.31$, $F = 404$

$$\log (1/EC_{50}) = 0.72 + 0.84\, \pi_2^H + 0.62\, \Sigma\alpha_2^H - 4.14\, \Sigma\beta_2^O + 3.67\, V_x \quad (32)$$

$n = 61$, $r^2 = 0.918$, $sd = 0.43$, $F = 158$

Here, LC_{50} and EC_{50} are the lethal concentration and effective concentration respectively, in mol dm^{-3}. Similar equations for smaller numbers of solutes were constructed [55] for aqueous toxicity towards the Golden Orfe and *Vibrio fischeri*.

5. Conclusions

We have seen that is possible to use GLC data directly to correlate biological phenomena that involve gaseous solutes. There will usually be little technical difficulty in obtaining the required GLC data, especially if temperature programming is used. However, it is not at all easy to choose a suitable model stationary phase. If the biological process has been characterised through Eq.(3), then comparison of system coefficients, or of the ratios of system coefficients, with characterised GLC stationary phases is of considerable help (see Table 3).

For processes within condensed phases, RP-HPLC or other chromatographic methods such as micellar electrokinetic chromatography could provide suitable model systems. Once again, choice of the model chromatographic system would be very considerably aided through the characterisation of the biological process and comparison of system coefficients with those for candidate chromatographic systems, as shown by Roses et al.[33]. However, there may well be technical difficulties if the biological system covers a wide range of compounds. For example, Roses et al.[33] showed that an MEKC system was a good model for aqueous narcosis, see Eq.(19). The MEKC system was used to obtain data for quite a wide range of compounds, the simplest being benzene, and the most complicated being 2,3-benzofuran or pyrimidine. But data for aqueous narcosis were obtained for a much larger range of compounds, from pentane, urea, acetonitrile and methanol up to caffeine, morphine and antipyrine.

It seems very unlikely that an MEKC system could be devised that covers such a wide range of compound.

In a similar vein, Eq.(18) was set up by Salminen et al. [30] who used RP-HPLC with an IAM column to obtain log *k'* values as a descriptor in the correlation of blood-brain distribution. Eq.(18) covers 21 quite complicated solutes, ranging from acetylsalicylic acid and 4-hydroxyacetanilide up to trifluoroperazine and verapamil. Log *BB* values, however, are available not only for complicated molecules but also for molecules such as neon, methane, ethanol, and important compounds used as gaseous anesthetics (halothane, enflurane, etc.). There is little prospect of an RP-HPLC system being able to cope with the wide range of compound for which log *BB* values are required, and for which estimations need to be made. In addition, there may also be difficulties over the type of acidic or basic compound that can be studied by RP-HPLC. Considerable progress has been made in construction of commercial columns that are stable over a wide range of pH, but stationary phases that contain IAM, for example, are notoriously unstable at high pH, and cannot be used to study very basic compounds, such as numerous classes of drugs.

The use of data obtained from any form of liquid chromatography as a descriptor for biological processes thus seems to be limited to those biological processes for which data have been obtained for a restricted set of solutes. Since predictions from any correlation equation can (or should) be made only for compounds that fall within the descriptor range used to set up the equation, this implies that predictions will be severely limited.

The indirect method, see Fig.1, is not so restricted as regards predictive capability, consider Eq.(11) and Eq.(12) that are very simple equations, but which can only be

used [26] for the estimation of skin permeability coefficients for steroids and phenols respectively. The indirect equations for skin permeability, Eq. (28) and Eq. (29), were constructed from data not only for steroids and phenols but for alcohols, ethers, carboxylic acids, and aromatic hydrocarbons as well. Because of the wide range of compound type used to set up the correlation, predictions can reasonably be made for a very large number of compounds that will all fall within the descriptor space used. Table 5 lists the minimum and maximum values of the descriptors used in Eq.(28), together with the number of compounds for which we have determined descriptors within these limits. There are 1850 compounds for which all the descriptors used in Eq.(28) are available, and for which predictions of skin permeation can be made. Eq. (29) is even more general, and can be applied to any compound for which log P_{oct} is available. Thus although direct connections between chromatographic processes and biological processes have the merit of simplicity, such connections are unlikely to have the generality of indirect connections such as those obtained via Eq.(3) and Eq.(4).

Acknowledgments

We are very grateful to the Center for Indoor Air Research (CIAR) for their support.

6. References

[1] P.Seiler, Eur. J. Med. Chem., 9 (1974) 473-479.

[2] R.C.Young, R.C. Mitchell, T.H.Brown, C.R.Ganellin, R.Griffiths, M.Jones, K.K. Rana, D.Saunders, I.R.Smith, N.E. Sore, T.J.Wilks, J. Med. Chem. 31 (1988) 656-671.

[3] M.H.Abraham, Chem.Soc.Revs., 22 (1993) 73-83.

[4] M.H.Abraham, P.L.Grellier and R.A.McGill, J.Chem.Soc., Perkin Trans.2., (1987) 797-803.

[5] M.H.Abraham and J.C.McGowan, Chromatographia, 23 (1987) 243-246.

[6] M.H.Abraham, R.Kumarsingh, J.E.Cometto-Muñiz, W.S.Cain, M.Roses, E. Bosch and M.L.Diaz, J.Chem.Soc., Perkin Trans.2, (1998) 2405-2411.

[7] K.Valko, M.Plass, C.Bevan, D.Reynolds and M.H.Abraham, J.Chromatogr.A, 797 (1998) 41-55.

[8] C.M.Du, K.Valko, C.Bevan, D.Reynolds and M.H.Abraham, Anal.Chem., 70 (1998) 4228-4234.

[9] J.E.Cometto-Muñiz and W.S.Cain, Indoor Air, 4 (1994) 140-145.

[10] J.E.Cometto-Muñiz and W.S.Cain, Chemical Senses, 20 (1995) 191-198.

[11] J.E.Cometto-Muñiz and W.S.Cain, Ann.N.Y. Acad.Sci., 641 (1992) 137-151.

[12] Y.Alarie, Environ. Health Perspect., 42 (1981) 9-13.

[13] M.H.Abraham, J.Andonian-Haftvan, J.E.Cometto-Muñiz and W.S.Cain, Fundam. Appl. Toxicol., 31 (1996) 71-76.

[14] M.H.Abraham, R.Kumarsingh, J.E.Cometto-Muñiz and W.S.Cain., Arch. Toxicol., 72 (1998) 227-232.

[15] J.E.Cometto-Muñiz, W.S.Cain, M.H.Abraham and R.Kumarsingh, Pharmacol. Biochem.Behav., 60 (1998) 765-770.

[16] M.H.Abraham, R.Kumarsingh, J.E.Cometto-Muñiz and W.S.Cain, Toxicol. in Vitro, 12 (1998) 403-408.

[17] M.T.D.Cronin, D.A.Basketter and M.York, Toxicol. in Vitro, 8 (1994) 21-28.

[18] C.Altomare, R.-Y.Tsai, N. El Tayar, BG.Testa, A.Carotti, S.Cellamare and P.G. De Benedetti, J.Pharm.Pharmacol., 43 (1991) 191-197.

[19] C.Altomare, S.Cellamare, A.Carotti and M.Ferappi, Il Farmaco, 49 (1994) 393-401 ; C.Altomare, A.Carotti, S.Cellamare,A.Contento, R.Ciabatti, A.Malabarba, M.Berti and B.Goldstein, Med.Chem.Res., 1 (1992) 393.

[20] H.Thurnhofer, J.Schnabel, M.Betz, G.Lipka, C.Pidgeon and H.Hauser, Biochim. Biophys.Acta, 1064 (1991) 275-286.

[21] F.M.Alvarez, C.B.Bottom, P.Chikhale and C.Pidgeon, in Molecular Interactions in Bioseparations, T.T.Ngo (Ed.), Plenum Press, New York, 1993.

[22] S.Ong, H.Liu and C.Pidgeon, J.Chromatogr.A, 728 (1996) 113-128.

[23] C.Pidgeon, S.Ong, H.Liu, X.Qui, M.Pidgeon, A.H.Dantzig, J.Munroe, W.J.Hornback, J.S.Kasher, L.Glunz and T.Szczerba, J.Med.Chem., 38 (1995) 590-594.

[24] S.Ong, H.Lui, X.Qui, G.Bhat and C.Pidgeon, Anal.Chem., 34 (1995) 755-762.

[25] R.Kaliszan, A.Nasal and A.Bucin′ski, Eur.Med.Chem., 29 (1994) 163-170.

[26] A.Nasal. M.Sznitowska, A.Bucinski and R.Kaliszan, J.Chromatogr.A, 692 (1995) 83-89.

[27] F.Barbato, B.Cappello, A.Miro, M.I. La Rotonda and F.Quaglia, Il Farmaco, 53 (1998) 655-661.

[28] M.Turowski and R.Kaliszan, J.Pharm.Biomed.Anal., 15 (1997) 1325.

[29] M.Turowski and R.Kaliszan, Biomed. Chromatogr., 12 (1998) 187-192.

[30] T.Salminen, A.Pulli and J.Taskinen, J.Pharm.Biomed.Anal., 15 (1997) 469-477.

[31] S.Yang, J.G.Bumgarner, L.F.R.Kruk and M.G.Khaledi, J.Chromatogr.A, 721 (1996) 323-335.

[32] S.K.Poole and C.F.Poole, J.High Resol. Chromatogr., 20 (1997) 174-178.

[33] M.Roses, C.Rafols, E.Bosch, A.M.Martinez and M.H.Abraham, J.Chromatogr. A, 845 (1999) 217-226.

[34] M.H.Abraham and C.Rafols, J.Chem.Soc., Perkin Trans.2, (1995) 1843-1851

[35] P.Laffort and F.Patte, J.Chromatogr., 126 (1976) 625-639..

[36] F.Patte, M.Etcheto and P.Laffort, Anal.Chem., 54 (1982) 2239-2247.

[37] P.Laffort and F.Patte, J.Chromatogr., 406 (1987) 51-74.

[38] Y.Alarie, M.Schaper, G.D.Nielsen and M.H.Abraham, SAR and QSAR in Environ.Res., 5 (1996) 151-165.

[39] Y.Alarie, M.Schaper, G.D.Nielsen and M.H.Abraham, Arch.Toxicol., 72 (1998) 125-140.

[40] G.D.Nielsen, M.H.Abraham, L.F.Hansen, M.Hammer, C.J.Cooksey, .Andonian-Haftvan and Y.Alarie, Arch.Toxicol., 70 (1996) 319-328.

[41] M.H.Abraham, H.S.Chadha, F.Martins, R.C.Mitchell, M.W.Bradbury and J.A.Gratton, Pesticide Science, 55 (1999) 78-88.

[42] M.H.Abraham, H.S.Chadha and R.C.Mitchell, J.Pharm.Sci., 83 (1994) 1257-1268.

[43] J.A.Gratton, S.I Lightman and M.W.Bradbury, J.Physiol., 470 (1993) 651-663.

[44] J.A.Gratton, M.H.Abraham, M.W.Bradbury and H.S.Chadha, J.Pharm. Pharmacol., 49 (1997) 1211-1216.

[45] M.H.Abraham, F.Martins and R.C.Mitchell, J.Pharm.Pharmacol., 47 (1995)

8-16.

[46] O.A.Raevsky and K.-J. Schaper, Eur.J.Med.Chem., 33 (1998) 799-807.

[47] M.T.D.Cronin, J.C.Dearden, G.P.Moss and G.Murray-Dickson, Eur.J. Pharm. Sci., 7 (1999) 325-330.

[48] C.E.Overton, Studien uber die Narkose, Fischer, Jena, 1901.

[49] C.E.Overton, Studies on Narcosis, R.L.Lipnick (Ed.), Chapman and Hall, London, 1991.

[50] J.K.Alifimoff, L.L.Firestone and K.W.Miller, Anaesthesiology, 66 (1987) 55-59.

[51] J.K.Alifimoff, L.L.Firestone and K.W.Miller, Br.J.Pharmacol., 96 (1989) 9-16.

[52] S.Curry, G.W.J.Moss, R.Dickenson, W.R.Lieb and N.P.Franks, Br.J.Pharmacol., 102 (1991) 167-173.

[53] G.W.J.Moss, S.Curry, N.P.Franks and W.R.Lieb, Biochem., 30 (1991) 10551-10557.

[54] R.Dickenson, E.H.Smith, N.P.Franks and W.R.Lieb, J.Med.Chem., 36 (1993) 111-118

[55] A.D.Gunatilleka and C.F.Poole, Anal.Commun., 36 (1999) 235-242.

Table 1

Coefficients in eq.(3) for some GLC stationary phases [a]

Phase	c	r	s	a	b	l
OV1 [b]	0.705	0.000	0.746	0.683	0.000	2.001
SE54 [b]	0.578	-0.419	1.430	0.983	0.000	2.010
PEG [c]	-5.496	0.960	5.629	8.841	0.000	2.037
TEG [d]	-2.731	0.208	1.882	4.636	0.313	0.584
DBS [e]	-2.034	0.000	0.968	1.911	0.045	0.765

[a] Ref.6. [b] I/100. [c] I/100 with methyl esters used as the standards. [d] Log tr relative to decane. [e] Log tr relative to hexane.

Table 2

Observed and calculated retention data for α-terpinene [a]

Phase	Obs	Calc
OV1	10.100	10.325
SE54	6.010	6.023
PEG	10.210	10.194
TEG	0.670	0.665
DBS	1.830	1.824

[a] With descriptors indicated in the text.

Table 3

The ratio of coefficients in Eq.(3) for various processes

Process	r/l	s/l	a/l	b/l
log tr(TEG)	0.356	3.223	7.938	0.536
log tr(DBS)	0.000	1.265	2.498	0.059
log (1/RD_{50})	0.000	1.856	2.992	0.000
log (1/NPT)	0.000	2.009	3.968	1.089
log (1/EIT)	-0.565	1.664	4.719	1.429

Table 4

Coefficient ratios in Eq.(4) for MEKC

Surfactant [a]	r/v	s/v	a/v	b/v
SDS(50): Brij(0)	0.12	-0.14	-0.01	0.68
SDS(50): Brij(5)	0.18	-0.15	0.11	0.97
SDS(50): Brij(10)	0.22	-0.18	0.11	0.95
SDS(50): Brij(15)	0.24	-0.19	0.12	0.98
log (1/C_{nar}) [b]	0.23	-0.21	0.07	0.78

[a] Concentrations in mM, ref.33 [b] Tadpole narcosis, ref. 34.

Table 5

The range of descriptors used to construct Eq.(28)

Descriptor	Min	Max	Number [a]
R_2	0.04	2.04	3110
π_2^H	0.25	3.51	2660
$\Sigma\alpha_2^H$	0.00	1.40	3820
$\Sigma\beta_2^H$	0.14	1.92	2130
V_x	0.324	2.913	4110

[a] The number of compounds for which the descriptor is known.

Legends to the figures

Fig. 1. The direct and indirect connection between chromatographic data and biological data.

Fig. 2. Plot of log (1/NPT) against log tr on TEG

Fig. 3. Plot of log (1/NPT) against log tr on DBS

Fig. 4. Plot of log (1/EIT) against log tr on TEG

Fig. 5. Plot of log (1/EIT) against log tr on DBS

Fig. 6. Plot of log ($1/RD_{50}$) against log tr on DBS

FIGURE 1

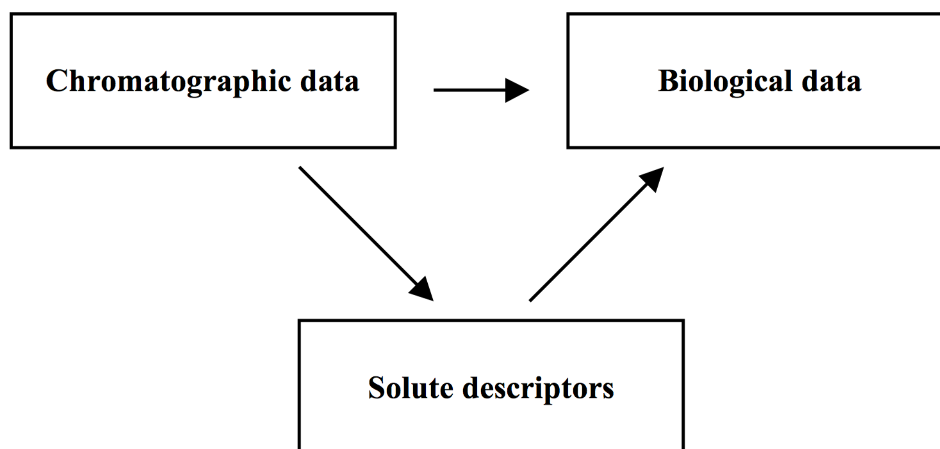


FIGURE 2

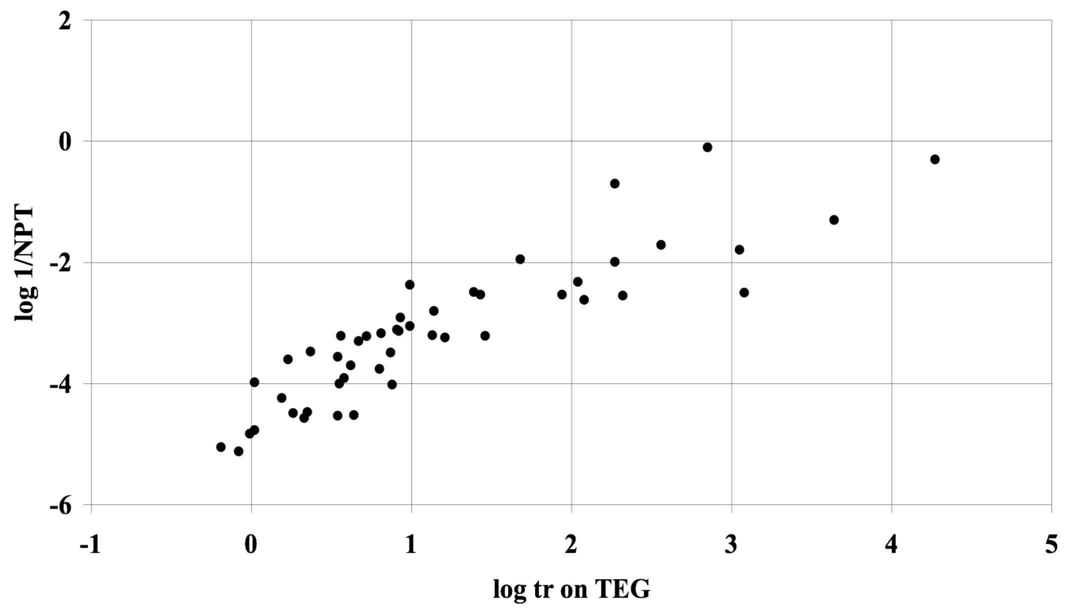


FIGURE 3

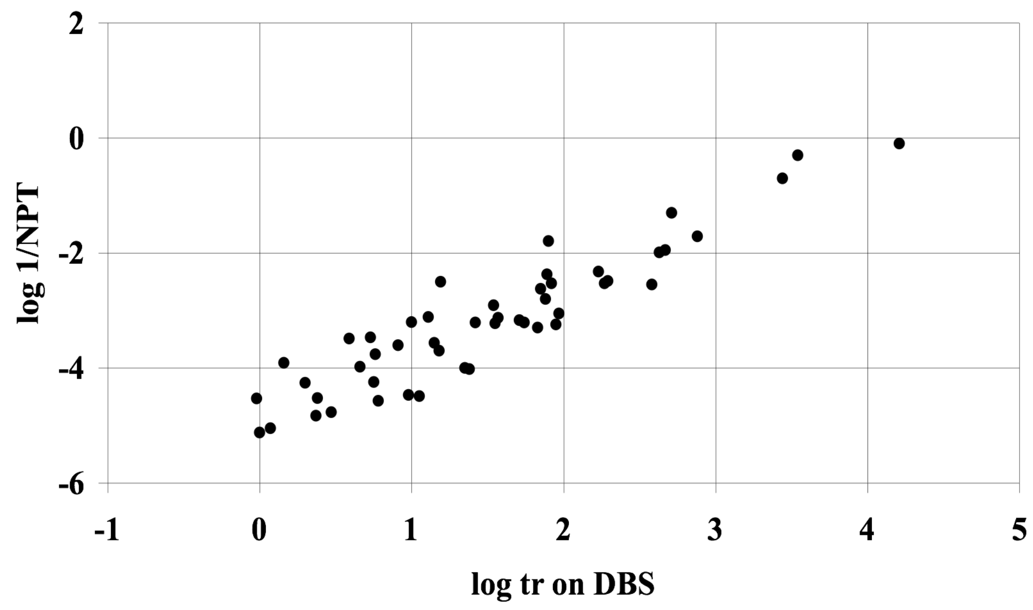


FIGURE 4

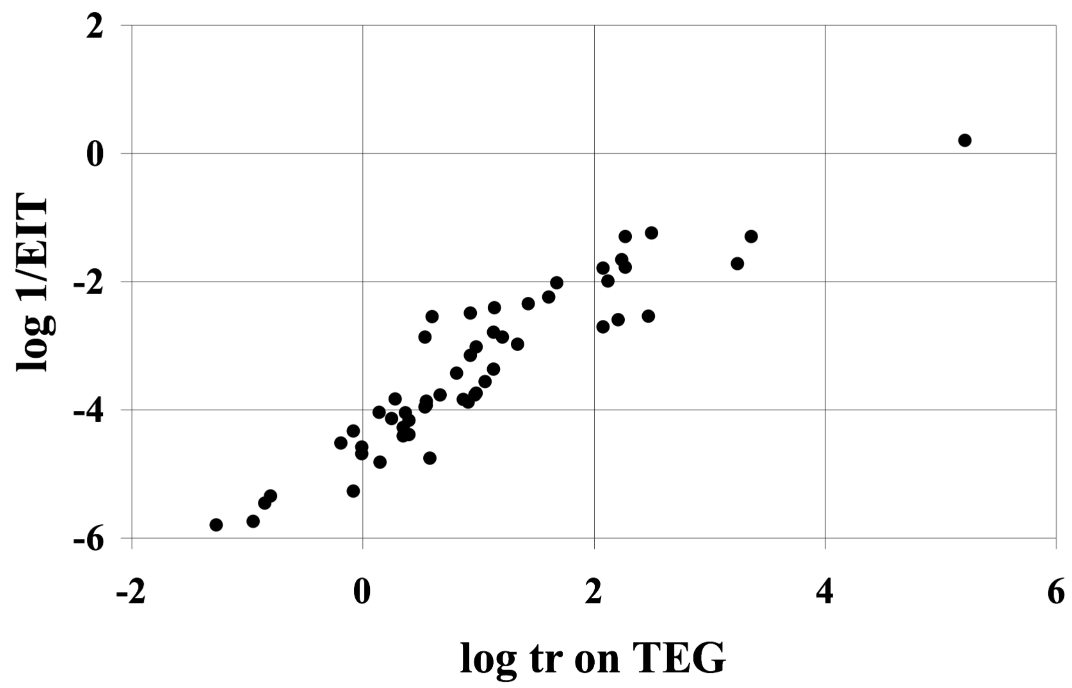


FIGURE 5

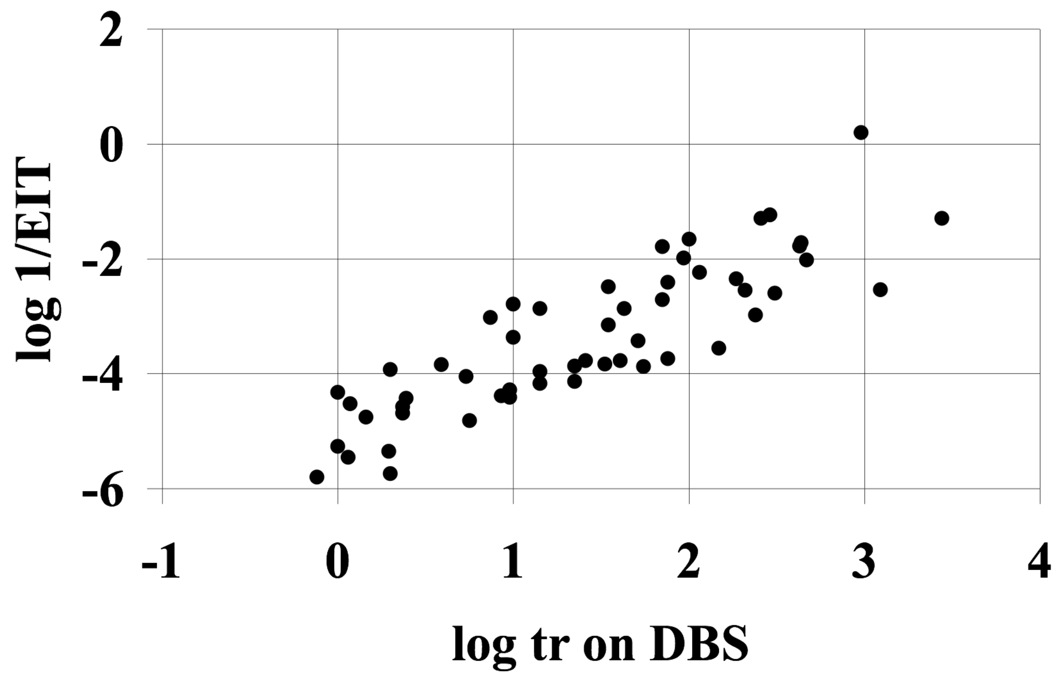
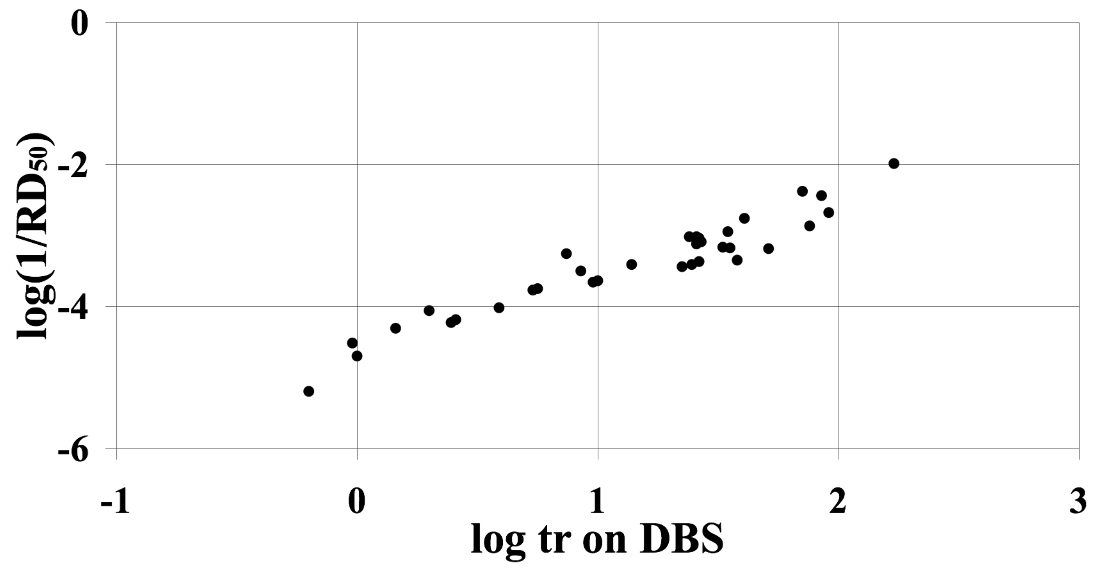


FIGURE 6



This is a pre-copyedited, author-produced version of an article accepted for publication in Journal of Chromatography B following peer review. The version of record Journal of Chromatography B **745**(1):103-115, 2000 is available online at: http://www.sciencedirect.com/science/article/pii/S0378434700001304 - DOI: 10.1016/S0378-4347(00)00130-4