ELSEVIER

Journal of Chromatography A, 788 (1997) 207–211

JOURNAL OF CHROMATOGRAPHY A

Short communication

Empirical equation for the accurate prediction of retention in planar chromatography

Megan C. Frost[a], Tom Lahr[a], Robert M. Kleyle[b], David Nurok[a,*]

[a]*Department of Chemistry, School of Science, Indiana University—Purdue University at Indianapolis, 402 Blackford Street, Indianapolis, IN 46202, USA*

[b]*Department of Mathematical Sciences, School of Science, Indiana University—Purdue University at Indianapolis, 402 Blackford Street, Indianapolis, IN 46202, USA*

Received 6 May 1997; received in revised form 1 September 1997; accepted 1 September 1997

Abstract

Either a second order or a third order polynomial equation accurately predicts R_F in a mobile phase which is a binary mixture of a strong and weak solvent. The mole fraction of the strong solvent is used as the independent variable. The *p*-nitrobenzyl esters of fifteen dansyl amino acids were used as model solutes for planar chromatography, using each of five different mobile phases on silica gel layers. For either of the polynomial equations, a minimum of 98% of the predicted R_F values, at mole fractions not used for establishing the equations, are within 0.05 or less of an R_F unit (when compared to the experimental R_F), and 90% of the predictions are within 0.03 or less of an R_F unit.

Keywords: Retention prediction; Amino acids

1. Introduction

Binary mobile phases consisting of a mixture of a weak and strong solvent are widely used in planar chromatography. Retention in these mobile phases can be predicted by suitable equations (vide infra) that predict R_F as a function of the concentration of the strong solvent. This allows retention to be predicted over a wide range of mobile phase composition, based on only a limited number of experimental R_F measurements. These equations are widely used in computer-assisted optimization studies, where their availability reduces the number of experimental observations required to determine the mobile phase composition that yields an optimum separation [1,2]. This is especially true in the optimization of two-dimensional separations where it would be impractical to perform an optimization without computer-assisted techniques [2]. Such equations are useful also in studies where regression models are constructed to predict the retention of a solute as a function of various solvent descriptors (dipole moment, polarizability, etc.) [3]. The identity of the descriptors that enter such a regression model depend on the concentration of the strong solvent.

The following equation, originally introduced in a different format by Soczewinski [4], has been widely used [5] in both column and planar chromatography for predicting retention:

$$\ln k' = a \cdot \ln X_s + b \quad (1)$$

*Corresponding author.

0021-9673/97/$17.00
PII S0021-9673(97)00901-1

where k' is the capacity factor, X_s is the mole fraction of a strong solvent in a binary mixture with a weak solvent, and a and b are experimentally determined constants.

Capacity factor is related to R_F by the relationship:

$$R_F = \frac{1}{1 + k'} \quad (2)$$

Eq. (1) has been used in the authors' laboratory for studies on a variety of topics. An excellent line fit ($R^2 > 0.99$) is obtained for data obtained at five mole fractions of the strong solvent in the binary system provided that an appropriate range of concentrations is used. While R_F can be predicted with a high level of confidence within the concentration range used for establishing the regression constants, errors can occur when extrapolating the equation to higher mole fractions than those used in establishing the equation. Moreover, for some binary mobile phases, a good line fit is obtained only over a relatively small concentration range of the strong solvent.

An alternative equation for predicting R_F in binary mobile phases has been used by Wang and Xie [6]:

$$R_F = a + b \cdot V + c \cdot V^2 \quad (3)$$

where V is the volume fraction of the strong solvent, and is based on data collected at between three and five concentrations. (The authors used X_s for volume fraction. The symbol has been replaced to avoid confusion with Eq. (1)).

The report below discusses both second and third order polynomial equations for the accurate prediction of R_F of a series of p-nitrobenzyl esters of dansyl amino acids over a large retention range.

2. Experimental

The dansyl amino acids were purchased from Sigma (St. Louis, MO, USA) and the solvents were purchased from Aldrich (Milwaukee, WI, USA). The reagent for preparing the p-nitrobenzyl esters of the dansyl amino acids was a gift from Pierce (Rockford, IL, USA), and the esters were prepared according to the procedure in the 1989 Pierce handbook which is based on Ref. [7].

Thin-layer chromatography (TLC) was performed on either K5 (Cat. No. 4850-820) or K6 (Cat. No. 4860-820) silica gel plates which were obtained as a gift from Whatman (Clifton, NJ, USA). The plates (cut into 10×10 cm sections) were heated for 30 min at 90°C and placed in a desiccator maintained at a relative humidity of 60% until immediately before use. A twin trough chamber (Camag Scientific, Wilmington, NC, USA) was used for TLC, with a 15 min preconditioning period, and a solvent development length of 7.3 cm.

3. Results and discussion

Table 1 lists the five binary mobile phases, the range of strong solvent mole fraction for each mobile phase and the TLC layers (K5 or K6) used in the study. The mole fraction range was divided into approximately equal increments. The proportions of solvent for each mole fraction were determined using an analytical balance. The p-nitrobenzyl esters of the fifteen dansyl amino acids listed in Table 2 were used as model solutes. A single spot was obtained for each of the solutes, with the exception of the derivatives of serine and threonine which each yield two spots. The spot of higher R_F was used for the

Table 1
The binary mobile phases, range of strong solvent mole fractions and TLC layers used in the study

Mobile phase	Range of mole fractions for strong solvent[a]	TLC layer
Acetonitrile–toluene	0.0528–0.5274	K5
Ethyl acetate–toluene	0.0220–0.2201	K5
Tetrahydrofuran–fluorobenzene	0.0082–0.0832	K5
Acetonitrile–toluene	0.0520–0.5279	K6
Ethyl propionate–toluene	0.0499–0.4999	K6

[a] The mole fractions used were approximately equally spaced in each respective range.

Table 2
R_F range of values for the fifteen derivatized amino acids

Amino acids[a]	R_F range				
	Acetonitrile–toluene	Ethyl acetate–toluene	Tetrahydrofuran–fluorobenzene	Acetonitrile–toluene	Ethyl propionate–toluene
Norvaline	0.22–0.86	0.14–0.69	0.15–0.84	0.12–0.82	0.09–0.76
Glutamic acid	0.06–0.83	0.04–0.41	0.05–0.65	0.03–0.79	0.01–0.46
Aspartic acid	0.09–0.83	0.05–0.50	0.05–0.66	0.04–0.80	0.03–0.54
Phenylalanine	0.19–0.89	0.14–0.68	0.16–0.83	0.09–0.80	0.10–0.72
α-Amino-*n*-butyric	0.20–0.87	0.12–0.64	0.14–0.79	0.08–0.81	0.08–0.66
Valine	0.22–0.89	0.16–0.70	0.17–0.83	0.09–0.83	0.09–0.72
Sarcosine	0.24–0.88	0.16–0.65	0.18–0.80	0.10–0.83	0.09–0.59
Methionine	0.13–0.88	0.07–0.61	0.10–0.70	0.05–0.82	0.05–0.62
γ-Amino-*n*-butyric	0.20–0.84	0.06–0.50	0.08–0.59	0.05–0.76	0.04–0.50
Threonine	0.02–0.71	0.00–0.18	0.02–0.25	0.03–0.61	0.03–0.46
Tryptophan	0.07–0.85	0.04–0.44	0.05–0.51	0.03–0.76	0.10–0.75
Serine	0.09–0.81	0.05–0.40	0.07–0.43	0.05–0.64	0.04–0.46
Norleucine	0.24–0.90	0.16–0.71	0.19–0.84	0.11–0.82	0.11–0.77
Leucine	0.23–0.89	0.14–0.70	0.18–0.82	0.10–0.82	0.11–0.76
Glycine	0.10–0.84	0.06–0.47	0.08–0.59	0.04–0.74	0.05–0.50

[a] Each analyte is the *p*-nitrobenzyl ester of a dansyl amino acid.

latter two solutes. Table 2 lists also the range of R_F values for each of the fifteen derivatized amino acids, obtained by increasing the concentration of strong solvent in each of the binary mobile phases. The range includes very low R_F values for solutes such as the *p*-nitrobenzyl ester of dansyl threonine, and high R_F values for solutes such as the *p*-nitrobenzyl ester of dansyl leucine.

The aim of this study was to establish an equation fitted to R_F values at six concentrations, that would accurately predict R_F over a wide range of mobile phase strength. The two equations that were investigated were:

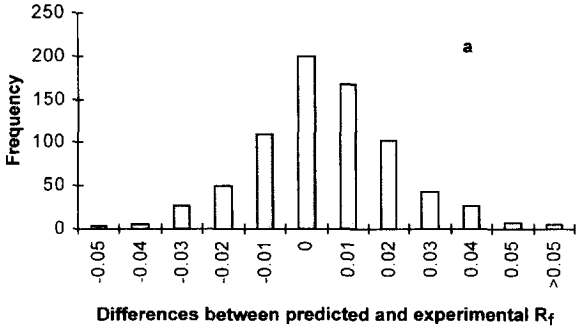
$$R_F = a + b \cdot X_s + c \cdot (X_s)^2 \quad (4)$$

$$R_F = a + b \cdot X_s + c \cdot (X_s)^2 + d \cdot (X_s)^3 \quad (5)$$

where X_s is the mole fraction of strong solvent in a binary mixture with a weak solvent. The first (the lowest), third, fifth, seventh, ninth and tenth (the highest) mole fractions of strong solvent were used for fitting the respective regression constants *a* through *c* (Eq. (4)), or *a* through *d* (Eq. (5)). The remaining four mole fractions were used for evaluating the prediction quality of the two equations.

Each of the above equations was used to predict the R_F values of the derivatized amino acids at each of the ten strong solvent concentrations in the five different mobile phases. For each equation, the differences between predicted and experimental R_F were computed and pooled to yield 750 (i.e., 15 solutes × 10 concentrations × 5 mobile phase systems) individual differences. The ten concentrations correspond to the six that were used for fitting the polynomials plus four intermediate concentrations. The histograms in Fig. 1a and Fig. 1b show the distribution of these differences, for the predictions based on Eqs. (4) and (5), respectively. Visual inspection of these histograms indicates that the third order polynomial provides a better prediction of R_F than the second order polynomial. There is a higher frequency of differences in R_F of 0.03, or less, in the third order polynomial. Nevertheless only six of the values predicted by either polynomial differ by more than 0.05 of an R_F unit from the experimental values This represents less than 1% of the data set.

Fig. 2a and Fig. 2b are the histograms, based on Eqs. (4) and (5), respectively, showing the distribution of differences between predicted and experimental R_F values for the four mole fractions that were not used for fitting the equation. This corresponds to 300 data points (15 solutes × 4 concentrations × 5 mobile phase systems). Both



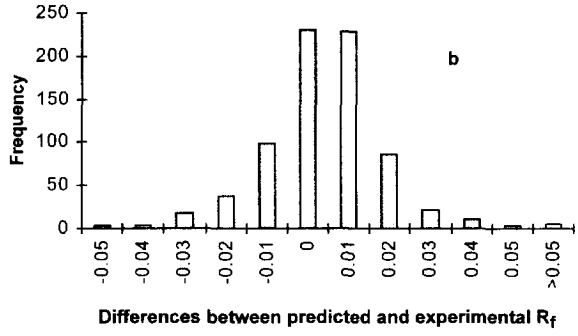
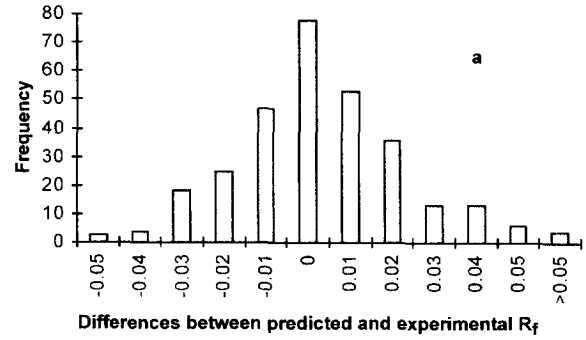


Fig. 1. Pooled histograms showing distribution of differences between experimental and predicted R_F for predictions based on both a second order (a) and third order (b) polynomial. The histograms are for all R_F values in the study. The numbers on the abscissa indicate the center point of each bar.



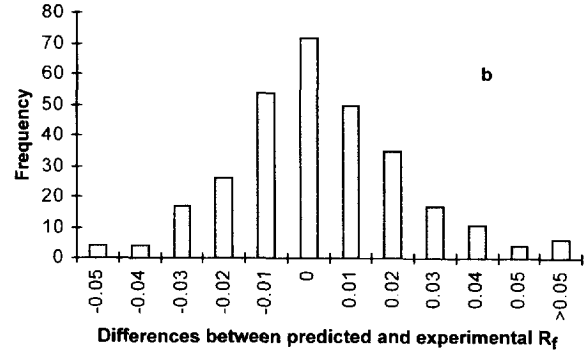


Fig. 2. Pooled histograms showing distribution of differences between experimental and predicted R_F for predictions based on both a second order (a) and third order (b) polynomial. The histograms are for the four mole fractions (in each mobile phase system) not used for establishing the relevant polynomials. The numbers on the abscissa indicate the center point of each bar.

histograms are very similar in shape and exhibit a normal distribution.

As noted above, there are 750 individual differences between predicted and experimental R_F in the pooled data. Only six of these differences are greater than 0.05 of a R_F unit when either the second order or third order equation is used for prediction. It is coincidental that there are six such differences, as the identity of the solute/mobile phase combinations in which these differences occur are not identical for the two equations. For the second order equation four of the differences occur with acetonitrile–toluene–K5 layer, and two for tetrahydrofuran–fluorobenzene–K5 layer. For the third order equation three of the differences occur with tetrahydrofuran–fluorobenzene–K5 layer, and one each occurs for acetonitrile–toluene–K5 layer, acetonitrile–toluene–K6 layer and ethyl propionate–toluene–K6 layer.

3.1. Statistical properties of the equations

A more rigorous discussion of the data requires the consideration of each individual solute. The following discussion is limited to the norvaline derivative run on K5 plates using a binary mixture of acetonitrile and toluene as the mobile phase. The comparison of second and third order models for this solute is typical of the models for the other solutes in this study.

A third order model will always have a higher R^2

Table 3
Second and third order models for norvaline derivative using acetonitrile–toluene on K5 plates

Coefficient for	Regression coefficient	Standard error	t Score	p Value
Second order model[a]				
$X_\searrow$	3.44	0.418	8.22	0.004
$(X_\searrow)^2$	−3.75	0.693	−5.42	0.012
Third order model[b]				
$X_\searrow$	5.27	0.673	7.83	0.016
$(X_\searrow)^2$	−11.54	2.723	−4.33	0.051
$(X_\searrow)^3$	9.01	3.121	2.89	0.102

[a] Model R^2 is 0.9858.
[b] Model R^2 is 0.9973.

for a given set of data than the corresponding second order model. The standard errors for the estimated regression coefficients may however be higher for the third order model than for the second order model. This is the case for the norvaline data in Table 3, where the values of the standard errors are uniformly higher in the third order than in the second order model. Thus the third order model has lower t scores and higher p values. This in turn results in a lower level of confidence for the regression coefficients in the third order model for the norvaline derivative than in the second order model for this solute.

The second order model is more robust and should be used for generalizations beyond a particular data set. However, for the purpose of fitting a model to a specific data set (as in this study), the third order model is preferable since it tends to yield lower residuals.

4. Conclusions

Both the second and the third order polynomials provide a very good prediction of R_F values over a large range of mobile phase strength. While the third order polynomial provides a higher frequency of very small differences between predicted and experimental values, both equations appear as good for the prediction of a value within 0.05 R_F unit of the experimental.

Acknowledgements

The authors thank Whatman for the gift of thin layer plates, and Pierce for the gift of reagent used for preparing the p-nitrobenzyl esters of the dansyl amino acids.

References

[1] Q.-S. Wang, in J. Sherma and B. Fried (Editors), Handbook of Thin Layer Chromatography, Marcel Dekker, New York, 1996, Ch. 3.
[2] D. Nurok, Chem. Rev. 89 (1989) 363.
[3] D. Nurok, R.M. Kleyle, P. Hajdu, B. Ellsworth, S.S. Myers, T.M. Brogan, K.B. Lipkowitz, R.C. Glen, Anal. Chem. 67 (1995) 4423.
[4] E. Soczewinski, Anal. Chem. 41 (1969) 179.
[5] E. Soczewinski, J. Chromatogr. 388 (1987) 91.
[6] Q.-S. Wang, W.-Q. Xie, J. Planar Chromatogr. 3 (1990) 153.
[7] D.R. Knapp, S. Kreuger, Anal. Lett. 8 (1975) 603.