

Accurate prediction of the retention behaviour of polydisperse macromolecules based on a minimum number of experiments

Fiona Fitzpatrick^a, Hans Boelens^b, Peter Schoenmakers^{a,*}

^a Polymer-Analysis Group, Department of Chemical Engineering, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

^b Process Analysis and Chemometrics, Department of Chemical Engineering, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

Received 19 January 2004; received in revised form 19 April 2004; accepted 26 April 2004

Available online 24 May 2004

Abstract

This study illustrates how retention models can be used to accurately predict the retention behaviour of polydisperse macromolecules in LC separations. It highlights that the number of experiments required can be drastically reduced when the relationship between the model parameters and molecular structure parameters (e.g. molar mass) can be incorporated into one global model. A practical implication of this work is that an appropriate model can then be used for the determination of molar-mass distributions for polydisperse samples. The globalised model can predict retention time as a function of molar mass and gradient slope. Both the original and globalised versions of the model were rigorously validated in terms of the difference between the predicted and experimental retention times. The original model had very low residuals and there was no apparent dependence of the errors on the applied gradient, the molar mass or the retention times. Confidence intervals for the model parameters (S and $\ln k_0$) were determined using a bootstrapping analysis of the residual errors in the predicted retention times. Confidence intervals were seen to broaden significantly as the mass of the polymer increased. The parameters were also seen to be highly correlated. For the global model, retention-time residuals remained quite low, even when the number of experiments used to determine the model parameters was reduced from approximately 100 to 10.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Retention behaviour; Retention prediction; Retention models; Molecular mass; Gradient elution; Polymers; Polystyrene

1. Introduction

1.1. Retention models for HPLC

The theory of liquid chromatography has been developed during several decades and some well-established descriptions are now in place for the elution behaviour of solutes in different chromatographic systems [1–3]. The retention behaviour of a solute can be described in terms of the relationship between its retention factor (k) and the (isocratic) mobile phase composition (i.e. the volume fraction of strong solvent, φ). For reversed-phase chromatography, the logarithm of the retention factor is found to vary in an approx-

imately linear manner with the mobile phase composition according to:

$$\ln k = \ln k_0 - S\varphi \quad (1)$$

where k_0 is the retention factor of the analyte in 100% of the starting (weak) solvent and S is the solvent-strength parameter (a measure of the rate of change of the retention factor with increasing mobile phase strength) [4–6]. Depending on the type of chromatography, other retention mechanisms may be more appropriate. For example, for normal-phase chromatography, a linear relationship between the logarithm of the retention factor and the logarithm of the mobile phase composition is predicted:

$$\log k = \log k_1 - m \log \varphi \quad (2)$$

where k_1 is the retention factor of the solute at $\varphi = 1$ (i.e. in 100% strong solvent) and m is the observed slope [7]. In

* Corresponding author. Tel.: +31-20-525-6642; fax: +31-20-525-6638.

E-mail address: pjschoen@science.uva.nl (P. Schoenmakers).

both cases, the model parameters can be estimated by measuring retention times at various (isocratic) mobile phase compositions. If the model holds, then a linear (or approximately linear) relationship should be found, according to the relevant equation (Eqs. (1) or (2) above).

For the chromatographic characterisation of high-molar-mass samples, such as polymers, isocratic chromatography is not always practical, because there is not always a wide enough range of mobile phase compositions where meaningful (i.e. finite but non-zero) retention can be measured. For example, for high-molar-mass polymers, the transition between infinite retention and zero retention occurs within a very narrow range of mobile phase compositions [8,9]. In this case, gradient LC is more appropriate (although, since elution behaviour remains sensitive to the mobile phase composition, small errors in the gradient profile can lead to significant errors in the observed retention time). The retention model can be solved using gradient LC retention data, taking into account the shape of the gradient [3]. For a linear gradient (i.e. $\varphi = A + Bt$, where A is the initial mobile phase composition and B is the composition change in mobile phase (φ) per unit time) and assuming a reversed-phase retention mechanism (Eq. (1)), retention times for solutes eluting within the gradient can be calculated as [3]:

$$t_R = \frac{1}{SB} \ln \left\{ 1 + SBk(A) \left[t_m - \frac{t_D}{k(A)} \right] \right\} + t_m + t_D \quad (3)$$

where $k(A)$ is the retention factor of the solute in the initial mobile phase composition and t_D and t_m are the dwell time of the system and the dead-time of the column, respectively. When $k(A)$ and S of a solute are known, the retention time of that solute can be predicted under any gradient or isocratic conditions. Conversely, S and $k(A)$ (and thus also k_0) can be estimated when the retention time of a solute is measured as a function of B .

1.2. Chromatographic retention of polymers

The mechanisms controlling the chromatographic retention of large molecules, such as polymers, are still not fully understood. It has been suggested that retention is controlled by a precipitation–redissolution mechanism, where analyte molecules are either fully retained or fully eluted, depending on the strength of the mobile phase [8,10,11]. Other research suggests that the retention of macromolecules is similar to the retention of smaller analyte molecules, i.e. Eqs. (1) and (2) hold equally well for large as well as small molecules [12]. We take the latter approach and test this by testing the validity of the models (Eqs. (1) and (2)) for the chromatographic retention of a series of polystyrene standards (different masses) in a non-aqueous reversed-phase system. In previous research, we have compared experimental retention values to retention times that were predicted using the reversed-phase model (Eqs. (1) and (3)) [13,14]. Excellent agreement between experimental and predicted retention times was obtained, indicating that the model was

appropriate for this LC separation. In general, the mechanisms involved in retention will depend on the sample, the concentration of the analytes, the choice of mobile phase and on the strength of the interaction between the sample and the stationary phase. The validity of a particular retention mechanism (and the models associated with that mechanism) should be explored on a case-by-case basis [15,16].

1.3. Applying retention models to polydisperse analytes

No single molecular structure can be assigned to a polydisperse sample. Instead, polydisperse analytes are characterised by distributions of, for example, molar mass, chemical composition, functional or end-groups, etc. [17]. The members of the distribution will have different retention behaviour across the distribution and no single set of parameters (e.g. S and $\ln k_0$) can adequately describe retention across the distribution. For the RP model, there is a strong relationship between S and $\ln k_0$ for homologous series such as homopolymers [6,13,14,18,19]. This means that model parameters can only change *concomitantly* within a distribution, so that the retention lines ($\ln k$ versus φ) are anchored to a common point (where all members of the series have the same retention factor). This is known as the critical mobile phase composition.

1.4. Determining correlations between model parameters and molar mass

When the retention of a single (very narrow) polymeric standard is measured as a function of the gradient slope B , the model parameters S and $\ln k_0$ for that standard can be estimated. If the model is appropriate, then the retention behaviour of that standard under any gradient or isocratic conditions can be predicted. By estimating S and $\ln k_0$ values for a range of standards (same polymer, different masses), the relationship between the model parameters and mass can be established. When there is a definable correlation between the model parameters and molar mass, it is possible to predict the parameters for a given polymer within the mass range studied. The well-known Martin rule predicts a linear relationship between $\ln k$ and the number of repeat units on the macromolecule [20], although it has been reported that it can fail for both low and high masses [21]. Skvortsov and Trathnigg [22] have suggested that the Martin rule only holds under special conditions, related to the radius of gyration of the molecule. In this study, we have examined the relationship between the RP model and molar mass for a series of polymeric standards. Using confidence intervals and weighted linear and non-linear regression analysis, we have found the best relationship between the model and the molar mass of the polymer. Using this relationship, we have developed a globalised version of the model, i.e. made it applicable to a *range* of masses of the polymer rather than just the specific masses that were measured. With this global approach, we have been able to accurately

describe the retention behaviour of polydisperse analytes. We have also been able to dramatically reduce the number of experiments that are required to understand and predict the chromatographic retention behaviour of a polymer.

2. Experimental

Experiments were carried out on a Waters 2690 liquid chromatograph. Gradient control, data acquisition and data analysis were handled by Waters Millennium 3.2 software. The stationary-phase was Supelco Discovery C₁₈, particle size 5 μm , pore diameter 180 \AA . Column dimensions were 150 mm \times 2.1 mm i.d. and column temperature was maintained at 25 $^{\circ}\text{C}$. The solvents were tetrahydrofuran (THF; Biosolve, Valkenswaard, The Netherlands) and acetonitrile (ACN; Rathburn Walkerburn, UK). Both were HPLC grade. The flow-rate was 0.2 mL/min. Samples consisted of low-dispersity polystyrene (PS) standards (Polymer Labs., Church Stretton, UK). The sample-injection volume was 5 μL and sample concentrations were 1.5 mg/mL. For the calculation of the model parameters (S and $\ln k_0$), gradient programmes from 5 to 95% THF in acetonitrile were run over 20, 45, 60 and 90 min. Peak detection was performed with a Waters PDA 996 diode-array detector. All samples were run in duplicate. MATLAB (Mathworks, Natick, MA, USA) and Microsoft Excel (Microsoft, Seattle, WA, USA) were used for data analysis.

3. Results and discussion

3.1. Estimation of the model parameters S and $\ln k_0$

If the RP model is an appropriate model for the chromatographic separation of a particular analyte, then values of S and $\ln k_0$ can be found that can predict the retention behaviour of that analyte under any gradient or isocratic mobile phase conditions (for the given LC system). By measuring the retention time of the analyte under a number of different gradient conditions, values for S and $\ln k_0$ that give the least overall error in prediction can be estimated. A simple (though somewhat tedious) approach to the estimation of the model parameters is to plot the variation in the error of prediction as a function of S and $\ln k_0$. As S and $\ln k_0$ are varied, the quality of the prediction changes (with the most correct values of the two parameters giving the smallest error between predicted and experimental retention times). When the inverse of the error is plotted, the maximum in the grid corresponds to the best combination of S and $\ln k_0$ for that sample (Fig. 1).

In our previous papers, we calculated optimal values of S and $\ln k_0$ using the solver tool in Microsoft Excel. Solver uses a generalised reduced-gradient non-linear optimisation procedure [23]. This approach is useful, because it can find optimum values for any number of parameters reasonably

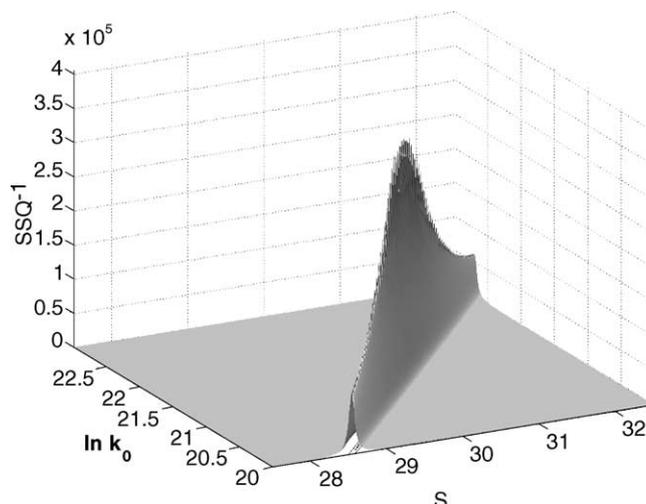


Fig. 1. Surface plot showing the variation in the SSQ^{-1} (i.e. the inverse of the squared difference between the predicted and experimental retention times) as a function of S and $\ln k_0$.

quickly. However, the results are obtained from an empirical stepwise process akin to computerised trial-and-error. Solver can also be misleading, because local rather than absolute optima are sometimes reported. This problem can be reduced, but not totally avoided by starting the process several times, using different initial estimates. The best results are obtained if approximately correct values for S and $\ln k_0$ can be used as the starting point. In this work, an alternative approach to estimating S and $\ln k_0$, using non-linear regression, was used [24]. This approach is in some ways similar to solver. It is also iterative and it can also get lost in local optima. However, since the optimisation procedure is based on the actual relationship between the model parameters, it generally reaches a more-accurate optimum than solver and it can do so more quickly. Both techniques gave quite similar values for S and $\ln k_0$ for the polystyrene standards (see Table 1). For higher masses, the differences were larger, but

Table 1
Calculated parameters for the global model using different data sets

Molar mass of standard	S		$\ln k_0$	
	Solver	Non-linear regression	Solver	Non-linear regression
1 700	13.93	13.96	3.63	3.63
2 100	14.05	14.35	4.38	4.44
3 250	17.82	17.35	6.22	6.09
4 000	23.09	22.30	8.27	8.01
7 000	25.96	25.55	10.07	9.92
10 900	29.58	28.25	12.14	11.62
17 600	36.17	35.92	15.67	15.57
30 000	49.15	48.43	22.27	21.95
39 200	57.37	56.95	26.46	26.27
76 600	80.50	84.24	38.14	39.93
117 000	113.75	120.83	54.56	58.00
160 000	156.93	169.64	75.75	81.99
325 000	397.37	319.96	194.06	157.94

they remained well within the calculated 95% confidence interval for each standard (see discussion below).

3.2. Validation of the model

Using the model (Eq. (1)) and the estimates of S and $\ln k_0$ from non-linear regression, a curve showing the relationship between retention time and the gradient slope B [for a given initial mobile phase composition (A)] can be predicted.

Fig. 2 shows the predicted curves for all of the masses measured, along with the experimental retention times measured at different gradient slopes. It is clear that the model can describe the retention behaviour of all the standards, in all the applied gradients. Errors in the prediction were assessed more rigorously by calculating retention-time residuals, i.e. the differences between the predicted and experimental retention times. Residual errors for all of the polymer samples as a function of B , are plotted in Fig. 3, along with an indication of the measurement error (i.e. experimental error) calculated from the 95% confidence interval of duplicate measurements of each standard at each applied gradient.

The residual errors (errors in prediction) are quite dependent on the measurement errors and increase with increasing measurement error. In most cases, residual errors are well within the experimental error. This indicates that the model describes the retention behaviour of these samples very well, since the error appears to be mainly due to experimental error rather than any error in the model. The two measurements with the largest residuals correspond to the duplicate injections of the PS 1700 standard eluting from the shallowest gradient (5–95% THF in ACN in 90 min). The chromatographic peaks corresponding to these measurements are broad and flat because selectivity in this mass range and under these chromatographic conditions is very high [14]. The true peak top was therefore difficult to determine for these peaks and so the measurement error was quite large. When

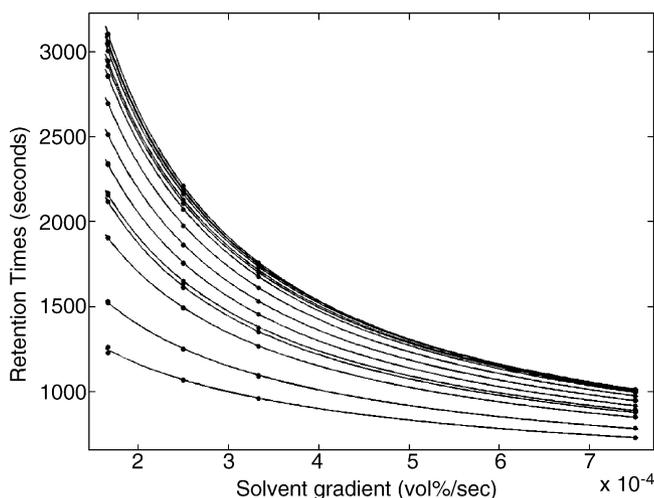


Fig. 2. Predicted curve for the relationship between retention time and gradient slope B for a series of polystyrene standards. The points correspond to the equivalent experimental values.

measuring the retention time of a polydisperse sample, it is assumed that the peak maximum corresponds to the average mass of that standard, i.e. its M_p value. While this can be a reasonable assumption for Gaussian-shaped chromatographic peaks, it is not always the case for asymmetrical peaks (coupling the separation to a mass spectrometer overcomes this problem). Residual errors were also plotted as a function of molar mass and retention time. In all cases, errors were low, non-systematic and comparable to the measurement error. The low, non-biased nature of the residuals in Fig. 3 indicate that: (a) the RP model is an appropriate model for an accurate description of the chromatographic separation in question and (b) the model works equally well for all gradient slopes and molar masses within the ranges that were tested.

3.3. Determination of confidence intervals for estimated S and $\ln k_0$ values using bootstrapping analysis

Confidence intervals for each of the S and $\ln k_0$ estimates were determined using external bootstrapping analysis [25]. Bootstrapping analysis uses retention time residuals to generate 1000 new data sets where:

$$t_{R,\text{boot}} = t_{R,\text{predicted}} + \tilde{\Delta}t_R$$

where $\tilde{\Delta}t_R$ is a noise data point generated by:

$$\tilde{\Delta}t_{R,i} = \frac{|\Delta t_{R,i}|}{\sqrt{1-h_{ii}}} \varepsilon_i$$

where $|\Delta t_{R,i}|$ is the retention time residual calculated from non-linear regression, ε_i a random number taken from a $N(0, 1)$ Gaussian distribution, and h_{ii} a value between 0 and 1 and is a measure of the model leverage for the residual i (the model leverage is an indication of the influence a given point has on the estimation of the model parameters the higher h_{ii} is, the more that point influences the model).

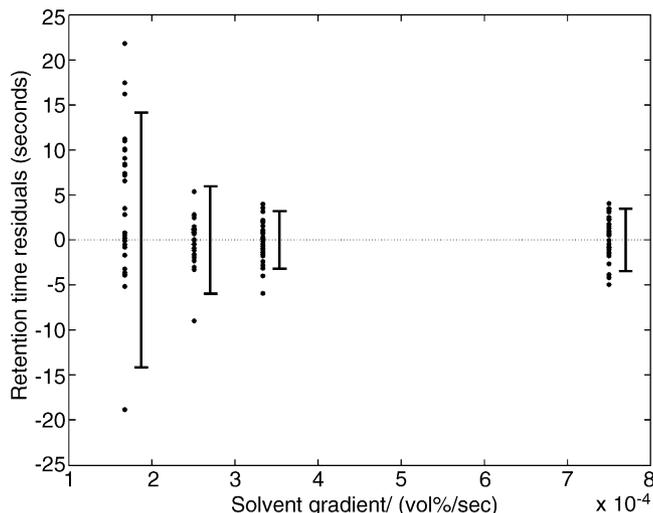


Fig. 3. Residual errors in prediction using the RP model and the calculated values of S and $\ln k_0$ (dots) and the corresponding measurement errors (solid bars).

The new data set (using $t_{R,boot}$ as the retention time) is again fitted using non-linear regression to give new estimations of S and $\ln k_0$ (S_{boot} and $\ln k_{0,boot}$) that are different from the original estimation (S_{best} and $\ln k_{0,best}$). This procedure was repeated 1000 times, to give 1000 separate estimations of S_{boot} and $\ln k_{0,boot}$. Examples of the estimations are given in Fig. 4. The lines represent the range of (highly correlated) values of S and $\ln k_0$ that can predict the retention time of a particular sample to within 5% of the experimentally measured value.

Confidence intervals were seen to get significantly larger as molar mass increased; in other words, the range of S and $\ln k_0$ values that can reasonably estimate retention times gets broader as mass increases. The confidence intervals prove that both the non-linear regression and Solver approaches to the estimation of S and $\ln k_0$ (Table 1) give statistically similar results. Fig. 5 shows the estimated best values of S and $\ln k_0$ and their confidence intervals for all of the standards. The strong correlation between the parameters for the homologous series is clear. This correlation is useful because it simplifies the model (since, if one of the parameters is known, then the other can be predicted). The correlation also proves the existence of a critical mobile phase compo-

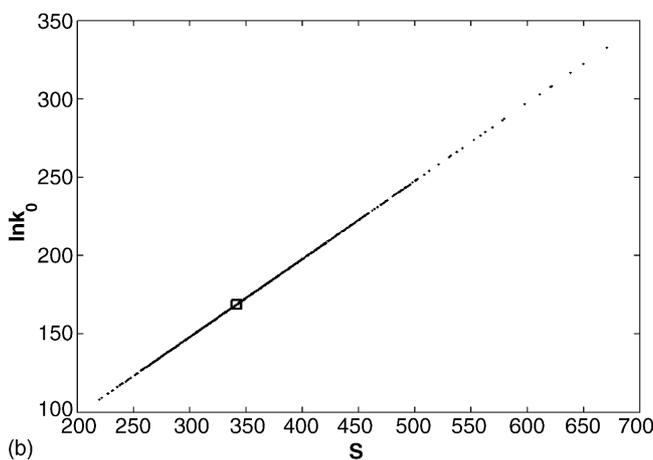
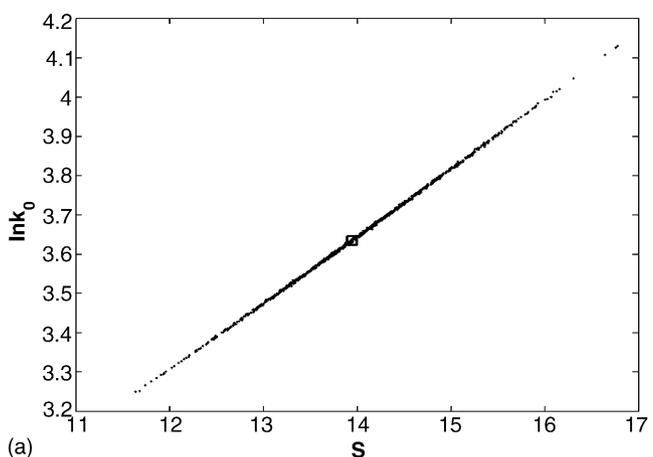


Fig. 4. Values of S and $\ln k_0$ calculated using external bootstrapping analysis: (a) PS 1700; (b) PS 325 000; (—□—) corresponds to the optimum calculated value.

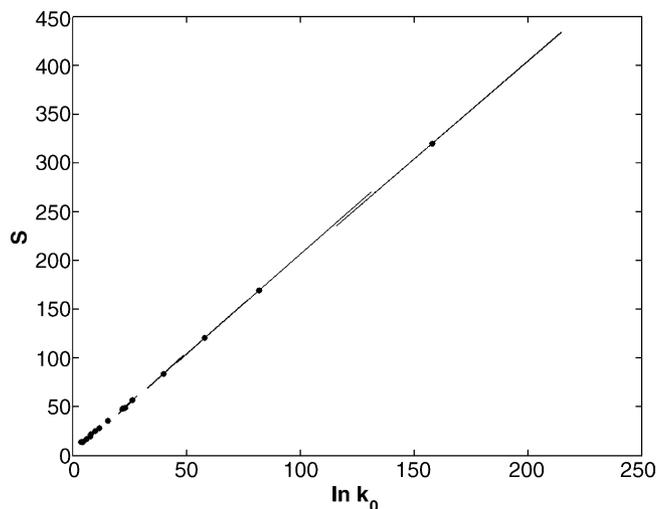


Fig. 5. Correlation between S and $\ln k_0$ for all the standards studied. Points correspond to S_{best} and $\ln k_{0,best}$ calculated using non-linear regression (see Table 1 also). Lines correspond to the confidence intervals calculated using external bootstrapping analysis.

sition where all members of a homologous series co-elute, regardless of their $\ln k_0$ value. This is a particular form of isocratic chromatography called critical chromatography. Its relevance and its link to the model have been discussed in more detail in earlier papers [13,19].

3.4. Creating a global model to predict retention times as a function of molar mass

From the original model, if S and $\ln k_0$ are known for a particular standard, retention time can be calculated as a function of the mobile phase conditions (i.e. A and B), i.e. $t_{R,i} = g(S_i, \ln k_{0,i}, A, B)$. S and $\ln k_0$ are estimated for each standard i , by measuring its retention time $t_{R,i}$ at different gradient slopes B . This approach can only be used to predict the retention behaviour of specific samples where S and $\ln k_0$ are known or can be estimated. The model, in this case, is rather limited in its application, because it cannot predict the retention behaviour of polydisperse samples, i.e. samples from the same homologous series with different molar masses. If the model could be expanded, so that the relationship between the model parameters and molar mass could be established, then the model would become much more versatile. In that case, we would have a *global* version of the model, represented by $t_{R,i} = f(B, A, M, \gamma_1, \dots, \gamma_n)$, where M is molar mass and γ_1 – γ_n are the parameters describing the relationship between the model parameters and molar mass. One function could then describe the retention behaviour of *all* members of the polymeric series (within the mass range where the global model is valid). The effects of both molar mass and molar-mass distribution (polydispersity) on the chromatographic separation could then be predicted.

Since we already know that there is a strong correlation between S and $\ln k_0$ (Fig. 5), the slope and intercept of

this line are used as two of the parameters (γ_1 and γ_2) in the global model. To link the parameters with molar mass, a relationship between one of the parameters and molar mass is also required. In this study, we have tested various (weighted) relationships between $\ln k_0$ and molar mass using linear and non-linear regression analysis and modified simplex searches. Weighting was assigned according to the confidence intervals determined by bootstrapping analysis. Thus, S and $\ln k_0$ values that had small confidence intervals were more heavily weighted than values that had large confidence intervals. The relationship between molar mass and $\ln k_0$ was difficult to describe using one straight line, even using weighted points. $\ln k_0$ values increased steeply with increasing mass in the lower mass region. The curve then levelled off for molar masses above approximately 4000. When all the masses were tested at once, weighted linear regression did not give a good prediction of the relationship between molar mass and $\ln k_0$. Because the confidence intervals of low-molar-mass standards were much smaller than confidence intervals for high-molar-mass standards, the line was forced to go through low-molar-mass points, at the expense of higher molar masses. This resulted in a bad fit of the $\ln k_0$ versus molar-mass relationship (see Fig. 6) and ultimately in a global model that could not predict accurate retention times over the entire mass range.

In a second approach, the $\ln k_0$ versus molar-mass relationship was taken to be bi-linear, i.e. split into two separate straight-line relationships, with a ‘transition mass’ (M_c) at which there is a switch from one line to the other. Non-linear regression was not suitable for the determination of parameters for the bi-linear relationship, because the discontinuity of the transition point resulted in it being over-sensitive to the starting values. A modified simplex-search method [26] gave robust results that did not depend on the starting values of the parameters. This approach again used weighted data and could estimate the parameters to describe the two separate linear relationships between mass and $\ln k_0$, as well as the critical mass (M_c) where the transition between the two lines took place. A much better fit was found using this approach and weighted residuals for the line were small. Fig. 6 compares the fit of the linear and bi-linear models. There is no physical meaning of the critical molar mass, i.e. the fit is mathematical rather than based on any real transition in the polymeric series. However, for the lower molar-mass region in particular, the bi-linear model worked better than the simple linear relationship (or other relationships based on continuous curves). The bi-linear model intersects the confidence intervals of all of the standards spanning the entire molar-mass range thus providing a much-more-accurate estimate of the relationship between $\ln k_0$ and molar mass (Table 2).

3.5. Predicting retention times using the global model

Using the global model (i.e. the bi-linear relationship between molar mass and $\ln k_0$ and the S versus $\ln k_0$

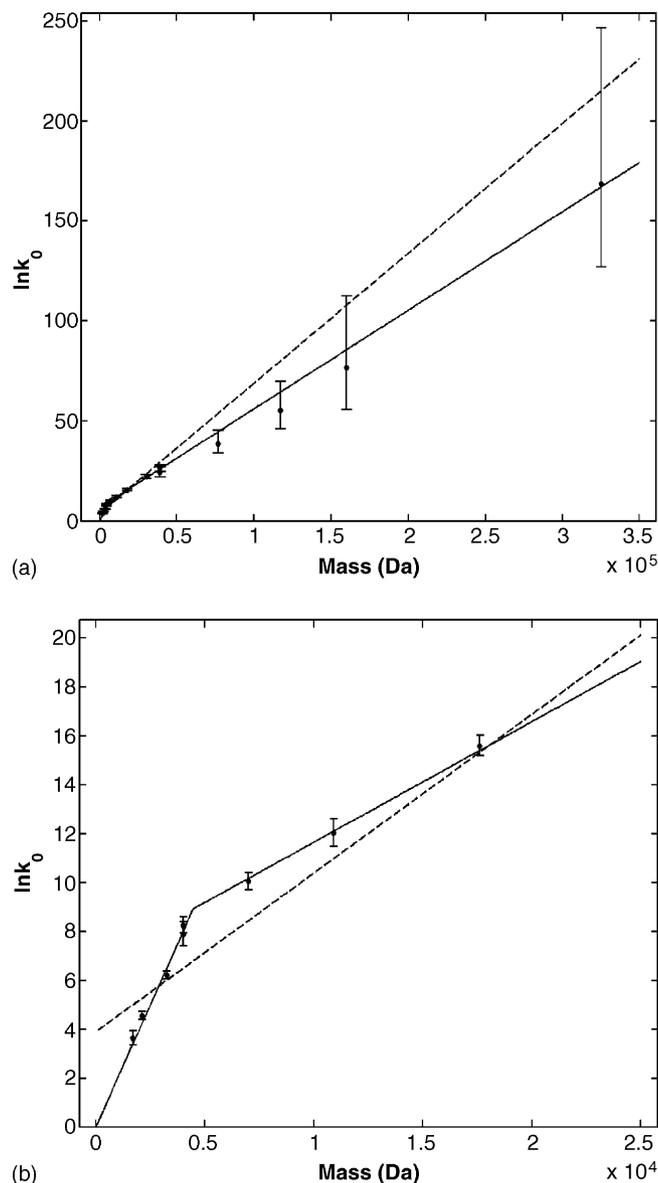


Fig. 6. Linear (dashed line) and bi-linear (solid line) curves for the $\ln k_0$ vs. molar mass relationship: (a) entire molar-mass range and (b) low molar-mass region.

Table 2
Definitions of the bi-linear model parameters

Parameter	Definition
γ_1	Intercept of the $\ln k_0$ vs. mass correlation in the low mass range, i.e. when $M \leq \gamma_4$
γ_2	Slope of the $\ln k_0$ vs. mass correlation in the low mass range, i.e. when $M \leq \gamma_4$
γ_3	Slope of the $\ln k_0$ vs. mass correlation in the high mass range, i.e. when $M > \gamma_4$
γ_4	Transition mass (M_c)
γ_5	Intercept of the S vs. $\ln k_0$ correlation
γ_6	Slope of the S vs. $\ln k_0$ correlation

correlation), the retention time of a sample of the given polymer of any mass and polydispersity can be predicted (within the range of masses covered by the model). The retention-time residuals calculated using the global model are given in Fig. 7 and should be compared with the retention-time residuals calculated using the non-global approach (i.e. S and $\ln k_0$ estimated for single standards) given in Fig. 3.

When the global approach is used, the residual error increases. This is because S and $\ln k_0$ estimates are no longer optimised one by one, but are calculated from the relationship between molar mass and the model. Some increase in the error of the prediction is therefore to be expected. However this is compensated by the enhanced versatility of the model. In general, the residuals remain reasonably low. Residuals were less than 60 s in all cases and less than 20 s in most cases (note that the retention times were long). The maximum relative residual was 3.2%, and the median relative residual was only 0.5%. The parameters of the global model are particular to a given polymer, in a given chromatographic system (in this case RPLC of polystyrene in THF–ACN). However, with further experimental work, other homologous series in other chromatographic systems can easily be characterised. As long as there is some way to describe the change in retention behaviour across a homologous series, it should be possible to find a global model. If a mass spectrometer can be coupled (off-line or on-line) to the separation, it is no longer necessary to use narrow dispersity standards of known molar mass, since the relationship between mass and retention time can be easily established from the MS data. This opens the way to model the retention behaviour of novel homopolymeric (and possibly co-polymeric) polydisperse samples. One important issue here is the number of experiments required to characterise a particular polymer. If the number of experiments that are

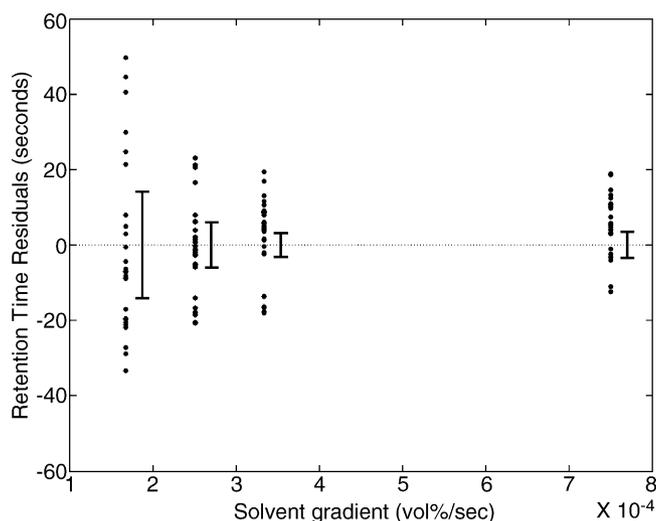


Fig. 7. Retention time residuals (dots) calculated using the global model. Solid lines correspond to the measurement error (as in Fig. 3). All experimental results are shown.

needed for each characterisation can be reduced (without any significant reduction in the accuracy of the model), then the application of retention models to polydisperse systems will become much more attractive to the polymer analyst.

3.6. Reducing the number of experiments required

Using the global model (function f), retention can be described as a three-dimensional surface, such as the one in Fig. 8 (the initial mobile phase composition A is kept constant). The surface of this function is quite smooth, indicating that the relationship between B , molar mass and retention time changes in a reasonably predictable fashion. This suggests that the number of data points required to define the surface can be decreased, without compromising the accuracy of the prediction. In the original procedure (calculating model parameters for individual standards), thirteen different standards were used and in each case four different gradient slopes were applied. Each measurement was performed in duplicate, giving over one hundred measurements. This is a significant amount of experimental work. If the number of experiments required could be reduced, many more systems could be characterised, with much lower costs (in terms of time and consumables).

The global model requires six parameters, γ_1 – γ_6 to predict the effect of molar mass on retention behaviour within the model. These parameters are taken from the slope and intercept values of the correlations, plus two experimental variables, i.e. the initial mobile phase composition A and the gradient slope B . The system parameters t_0 and t_D must also be known or measured. In principle therefore, only six measurements are required to calculate the parameters γ_1 – γ_6 of function f . When choosing the most-useful experiments for a reduced data set (i.e. the experiments that give the most information), the shape of the function in Fig. 8 must be considered. Measurements are required to span the surface (in terms of mass and B values), rather than focus on one

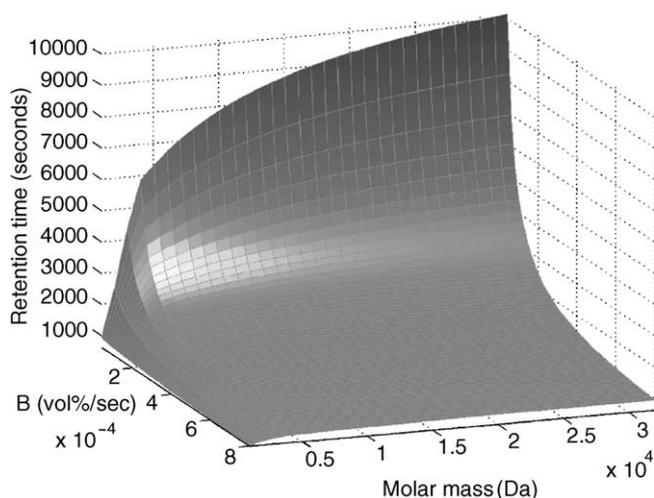


Fig. 8. Surface plot of the global model f .

Table 3
Details on the reduced data sets

Data set 2		Data set 3		Data set 4	
Masses	<i>B</i> values	Masses	<i>B</i> values	Masses	<i>B</i> values
1 700 × 2	All	1 700 × 2	All	1 700	0.000167; 0.00025
2 100 × 2		2 100 × 2		3 250	0.000167; 0.00025; 0.00033
3 250 × 2		3 250 × 2		4 000	0.000167
4 000 × 2		4 000 × 2		10 900	0.000167
4 000 × 2		10 900 × 2		39 200	0.00033
10 900 × 2		39 200 × 2		325 000	0.00025
39 200 × 2		325 000 × 2			
76 600 × 2					
325 000 × 2					

Table 4
Calculated parameters for the global model using different data sets

Data set	<i>N</i>	Parameters of the global model						Relative residuals (%)	
		γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	Median	Maximum
1	104	0.47	0.0018	0.0007	4000	5.20	2.00	0.5	3.2
2	72	0.41	0.0018	0.0007	4132	5.56	2.00	0.7	3.0
3	56	-0.56	0.0024	0.0008	2947	5.56	2.00	0.6	4.5
4	9	0.17	0.0021	0.0008	4000	6.82	1.98	0.7	3.1

particular area. The measurements that were used in each data set are given in Table 3.

Table 4 shows four different data sets that were used to predict the parameters γ_1 – γ_6 . Data set 1 uses all available (104) measurements, to calculate the parameters of function *f* (this corresponds to the residual plot in Fig. 7). Data sets 2–4 use a limited number of data points (see Table 3) for the determination of the parameters γ_1 – γ_6 . The quality of the fit for data sets 2–4 are compared by comparing the relative residual errors (maximum and median error) in the retention-time predictions.

Fig. 9 shows the residual errors for the final data set (nine measurements). It is clear that drastically reducing the num-

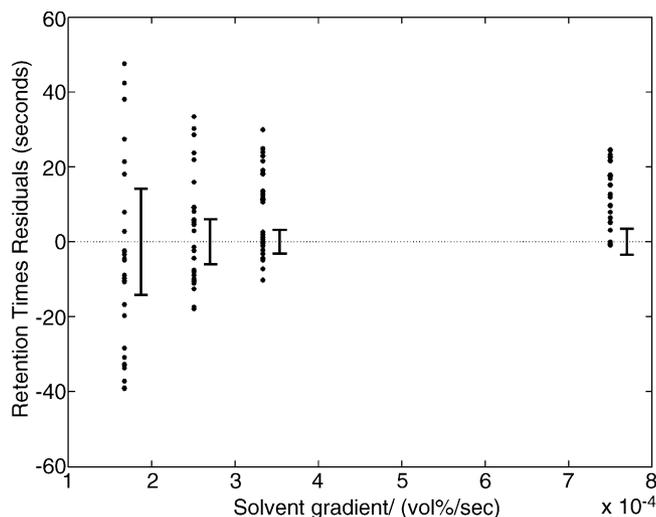


Fig. 9. Residual errors calculated using the global model and data set 4 to define the function *f* (see also Tables 3 and 4).

ber of measurements does not adversely affect the predictive quality of the model. The maximum relative error was as low when only nine measurements were used to determine the parameters of function *f*, as when all 104 measurements were included in the data set.

4. Conclusions

We have shown, through rigorous validation, that the RP model works very well for the characterisation of polystyrene in RPLC. We have also shown that by making the link between the model and molar mass, the retention models become more general and therefore, more useful predictive tools. A drastic reduction in the number of experiments that are necessary to characterise a polymer sample was possible using this generalised (or global) model, without jeopardising the quality of the prediction. This reduction in the experimental workload, along with the use of a mass spectrometer, opens the way for the characterisation of many more polydisperse macromolecular systems, including novel and unknown samples. To investigate the applicability of this approach to a wider range of characterisations, other polymer separations need to be studied, including some that require other predictive models, to determine if our approach remains valid and to help improve our general understanding of the retention behaviour of macromolecules.

References

- [1] P. Jandera, J. Churáček, Gradient Elution in Column Liquid Chromatography, Theory and Practice, Elsevier, Amsterdam, 1985.

- [2] L.R. Snyder, in: Cs. Horváth (Ed.), *HPLC—Advances and Perspectives*, Academic Press, New York, 1980.
- [3] P.J. Schoenmakers, H.A.H. Billiet, R. Tijssen, L. de Galan, *J. Chromatogr.* 149 (1978) 519.
- [4] L.R. Snyder, J.W. Dolan, J.R. Gant, *J. Chromatogr.* 165 (1979) 3.
- [5] J.W. Dolan, J.R. Gant, L.R. Snyder, *J. Chromatogr.* 165 (1979) 31.
- [6] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, *J. Chromatogr.* 185 (1979) 179.
- [7] P. Jandera, *J. Chromatogr. A* 965 (2002) 239.
- [8] G. Glöckner, *Gradient HPLC of Copolymers and Chromatographic Cross-Fractionation*, Springer, Berlin, 1991.
- [9] H. Pasch, B. Trathnigg, *HPLC of Polymers*, Springer, Berlin, 1998.
- [10] D.W. Armstrong, K.H. Bui, *Anal. Chem.* 54 (1982) 706.
- [11] H.J.A. Philipsen, Ph.D. Thesis, University of Eindhoven, 1998.
- [12] J.P. Larmann, J.J. DeStefano, A.P. Goldberg, R.W. Stout, L.R. Snyder, M.A. Stadalius, *J. Chromatogr.* 255 (1983) 163.
- [13] F. Fitzpatrick, R. Edam, P. Schoenmakers, *J. Chromatogr. A* 988 (2003) 53.
- [14] F. Fitzpatrick, B. Staal, P. Schoenmakers, *J. Chromatogr. A*, submitted for publication.
- [15] M.A. Quarry, M.A. Stadalius, T.H. Mourey, L.R. Snyder, *J. Chromatogr.* 358 (1986) 1.
- [16] M.A. Stadalius, M.A. Quarry, T.H. Mourey, L.R. Snyder, *J. Chromatogr.* 358 (1986) 17.
- [17] P.C. Hiemenz, *Polymer Chemistry: The Basic Concepts*, Marcel Dekker, 1984.
- [18] K. Valko, L.R. Snyder, J.L. Glajch, *J. Chromatogr. A* 656 (1993) 501.
- [19] P. Schoenmakers, F. Fitzpatrick, R. Grothey, *J. Chromatogr. A* 965 (2002) 93.
- [20] A.J.P. Martin, *Biochem. Soc. Symp. Lond.* 3 (1949) 4.
- [21] H.J.A. Philipsen, H.A. Claessens, H. Lind, B. Klumperman, A.L. German, *J. Chromatogr. A* 790 (1997) 101.
- [22] A. Skvortsov, B. Trathnigg, *J. Chromatogr. A* 1015 (2003) 31.
- [23] Microsoft Excel online help.
- [24] G.A.F. Seber, C.J. Wild, *Non Linear Regression*, Wiley, New York, 1989.
- [25] J. Shao, T. Dongsheng, *The Jackknife and Bootstrap (Springer Series in Statistics)*, Springer, New York, 1995.
- [26] C.W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes*, Cambridge University Press, Cambridge, 1992, pp. 408–412.