An Approach to High Throughput Measurement of Accurate Retention Data in Liquid Chromatography

Dwight R. Stoll[a,*], Gudrun Kainz[a], Tina A. Dahlseid[a], Trevor Kempen[a], Tyler Brau[a], and Bob Pirok[a,b]

[a] *Gustavus Adolphus College, 800 W College Ave, St. Peter, MN 56082, USA*

[b] *University of Amsterdam, van 't Hoff Institute for Molecular Sciences, Analytical-Chemistry Group, Science Park 904, 1098 XH Amsterdam, the Netherlands*

[*] Corresponding Author. Tel: +1 507 933 6304

E-mail address: dstoll@gustavus.edu (Dwight Stoll)

Abstract

Efforts to model and simulate various aspects of liquid chromatography (LC) separations (e.g., retention, selectivity, peak capacity, injection breakthrough) depend on experimental retention measurements to use as the basis for the models and simulations. Often these modeling and simulation efforts are limited by datasets that are too small because of the cost (time and money) associated with making the measurements. Other groups have demonstrated improvements in throughput of LC separations by focusing on "overhead" associated with the instrument itself – for example, between-analysis software processing time, and autosampler motions. In this paper we explore the possibility of using columns with small volumes (i.e., 5 mm x 2.1 mm i.d.) compared to conventional columns (e.g., 100 mm x 2.1 mm i.d.) that are typically used for retention measurements. We find that isocratic retention factors calculated for columns with these dimensions are different by about 20%; we attribute this difference – which we interpret as an error in measurements based on data from the 5 mm column – to extra-column volume associated with inlet and outlet frits. Since retention factor is a thermodynamic property of the mobile/stationary phase system under study, it should be independent of the dimensions of the column that is used for the measurement. We propose using ratios of retention factors (i.e., selectivities) to translate retention measurements between columns of different dimensions, so that measurements made using small columns can be used to make predictions for separations that involve conventional columns. We find that this approach reduces the difference in retention factors (5 mm compared to 100 mm columns) from an average of 18% to an average absolute difference of 1.7% (all errors less than 8%). This approach will significantly increase the rate at which high quality retention data can be collected to thousands of measurements per instrument per day, which in turn will likely have a profound impact on the quality of models and simulations that can be developed for many aspects of LC separations.

Keywords: high throughput, retention, selectivity, isocratic, gradient, liquid chromatography, modeling, database

1. Introduction

Increasing complexity of challenges faced by separation scientists along with the ever-increasing drive for more efficient method development is fueling continuing interest in modeling and simulation of a variety of aspects of liquid phase separations [1–7]. For example, recent studies by different research groups have focused on aspects including the effect of the volume and composition of the injected sample on separation quality [2,4,5,8], the effect of temperature on analyte retention in reversed-phase liquid chromatography (RPLC) [9], the effect of pump non-idealities on the prediction of retention time when using gradient elution conditions [10], and resolution of difficult-to-separate mixtures by serially coupling columns with different selectivities [11,12]. Currently, these efforts depend on experimental data to build models that are accurate enough to be useful in method development. In our own work we are very interested in increasing the throughput of high quality measurements, both for the purpose of improving the accuracy of existing retention/selectivity models (e.g., HSM2 for RPLC [13]), and for opening new lines of investigation that would allow modeling aspects of LC separations that thus far have been relatively untouched, such as optimization of second dimension elution conditions in two-dimensional liquid chromatography [3].

There have been some substantial efforts at building retention databases for RPLC. However, to the best of our knowledge these efforts have been highly asymmetric in nature, focusing either on a single stationary phase chemistry, for example as in the work of Boswell et al. [14] and Weber et al. [9] , or many stationary phases, but a small number of test analytes (e.g., www.hplccolumns .org, [15]). Clearly, a public retention database that covered multiple stationary phase chemistries, and many test analytes, as well as other important variables such as temperature, mobile phase pH, and organic modifier would be highly useful to a wide range of users, but this would be a highly resource- and time-intensive effort using conventional approaches to retention time measurement. Other groups have demonstrated improvements in throughput of LC separations by focusing on "overhead" associated with the instrument itself. For example, the MISER approach introduced by Welch et al. eliminates between-analysis software processing time by making multiple injections during the course of data acquisition stored in a single datafile [16–18], however to the best of our knowledge this approach has not been used for building retention databases. Our view is that a primary reason that extensive, open retention databases do not currently exist is that acquiring

these data is very resource intensive (both in terms of instrument and person time). In principle, retention times acquired under gradient elution conditions can be used to extract retention model parameters that can then be used to predict retention under any isocratic or gradient elution conditions. This type of workflow is attractive because a generic set of broad gradient elution conditions can be used that are likely to work for most compounds, and gradient elution is generally good for dealing with mixtures of compounds spanning a range of properties. However, we have been unable to demonstrate that this can be done accurately in practice, and recently our own theoretical work has shown that at least part of the problem has to do with challenges encountered in fitting the data (i.e., lack of uniqueness of model solutions, and complex fitting landscapes) that are mathematical in nature and have nothing to do with the experiments, *per se* [19]. Thus, recently we have turned our attention mainly to using isocratic elution conditions to acquire retention information for modeling purposes.

In the current study described in this paper we focus on the fact that analyte retention is a thermodynamic property of the mobile and stationary phase conditions under study, and thus retention measurement should – in principle – be independent of the geometry of the columns and systems used to make the measurement. In turn, this suggests that accurate retention measurements should be possible with very short columns and correspondingly short retention times. We would like to be able to make the retention measurements needed to establish the dependence of retention on conditions (e.g., temperature and mobile phase pH, but especially mobile phase composition (organic/water) in RPLC) using an approach that is efficient (e.g., many compounds per day) and robust (so that non-experts can reliably compile large databases).

To this end, in this paper we describe an approach to determine isocratic retention factors of relatively small molecules (< 1000 Da) under reversed-phase conditions using small columns, recognizing that retention factor is a thermodynamic property of the mobile phase/stationary phase/analyte system that should be independent of column length. Using columns with dead times that are small (e.g., < 1 s) when using flow rates typical of analytical scale instruments (e.g., 1 mL/min) facilitates high throughput measurements. With this approach our aim is not to obtain the most accurate (i.e., thermodynamically correct) and precise determinations of retention factors for specific systems; rather, our primary aim is to enable compilation of large datasets (e.g., tens of thousands of measurements) of retention factors (i.e., over a large range in k for each system

studied, with values applicable to prediction of both isocratic and gradient elution separations – see Paths #1 and 2 in Fig. 1) with reasonable accuracy and precision, at low cost (i.e., with UV detection and minimal supervision of the measurement process by expert users).

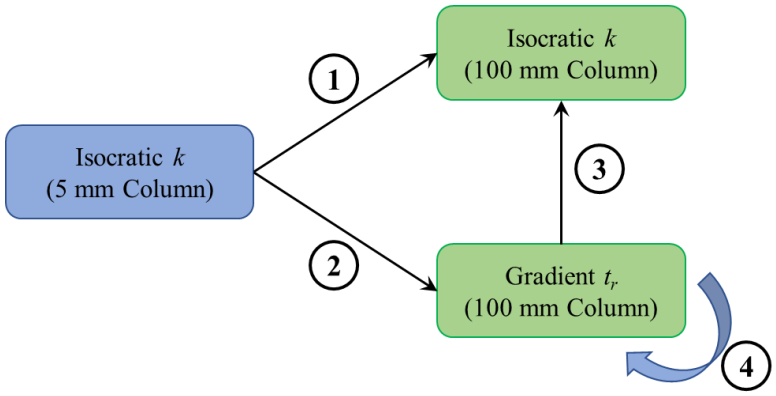


Figure 1. Schematic illustrating the different paths relating retention data collected or predicted under different conditions: 1) retention measurements made under isocratic conditions with a short column are used to predict retention that will be observed under isocratic conditions with a longer column; 2) retention measurements made under isocratic conditions with a short column are used to predict retention that will be observed under gradient elution conditions with a longer column; 3) retention measurements made under gradient elution conditions with a long column are used to predict retention that will be observed under isocratic elution conditions with a longer column; and 4) retention measurements made under gradient elution conditions with a long column are used to predict retention that will be observed under different gradient elution conditions with the same column.

2. Principles

2.1. Translation of measurements made using short columns to longer columns

In the experiments described below we have measured retention times for 13 test analytes using 5 mm or 100 mm long columns (both 2.1 mm i.d.). In this section we will refer to these generically as short (*S*) and long (*L*) columns. Ultimately our aim is to use retention data collected using the short columns to predict practical outcomes using longer columns typically used for analytical work (e.g., isocratic separations, gradient elution separations, selectivity comparisons, analyte

focusing, and breakthrough). The physical volumes outside of the stationary phase bed (*e.g.*, frits, flow distributors, and endfitting channels) that contribute to measured column dead volumes, but do not contribute to retention, can lead to errors in calculated retention factors. This problem becomes more serious as columns become short and the relative contribution of these unaccounted-for volumes becomes a larger fraction of the measured column dead volume. Our approach is to calculate selectivities – that is, ratios of retention factors measured using the small column - and use these to predict retention factors for long columns. This approach has the following steps:

Short Column (*S*)

1) Measure extra-column time ($t_{ex,S}$), column dead time ($t_{m,S}$), retention time for a reference compound (toluene in this work; $t_{r,ref,S}$), and retention time for analyte *i* ($t_{r,i,S}$).
2) Calculate retention factors for the reference compound ($k_{ref,S}$) and analyte *i* using Eq. 1. Note that the extra-column time t_{ex} must be subtracted from all instances of t_r and t_m to accurately calculate *k*:

$$k = \frac{\left(t_r - t_{ex}\right) - \left(t_m - t_{ex}\right)}{\left(t_m - t_{ex}\right)} = \frac{\left(t_r - t_m\right)}{t_m - t_{ex}} \quad (1)$$

3) Calculate selectivities using Eq. 2. Note that we define α_i here without regard to the relative magnitudes of k_i and k_{ref} (i.e., k_i is always in the numerator, even if it is smaller than k_{ref}). Although this is different from some uses of α that require $\alpha \geq 1$, we prefer the formulation defined here and shown in Eq. 2 for simplicity and efficiency:

$$\alpha_{i,S} = \frac{k_{i,S}}{k_{ref,S}} = \frac{t_{r,i,S} - t_{m,S}}{t_{r,ref,S} - t_{m,S}} \quad (2)$$

4) Assume $\alpha_{i,L} = \alpha_{i,S}$. Note that since each retention factor in the ratio of alpha is proportional to the product of the phase ratio and mobile-to-stationary phase transfer equilibrium constant, the phase ratio drops out of the expression because there can only be one phase ratio for a given column. Thus, while it is likely that the phase ratios are different for short and long columns, this does not matter to our approach because it drops out of the equation.

Long Column (*L*)

5) Measure extra-column time ($t_{ex,L}$), column dead time ($t_{m,L}$), retention time for a reference compound (toluene in this work; $t_{r,ref,L}$).
6) Calculate retention factor for the reference compound ($k_{ref,L}$) using Eq. 1.
7) Calculate retention factor for analyte *i* on the long column using Eq. 3:

$$k_{i,L} = \alpha_{i,L} \cdot k_{ref,L} = \alpha_{i,S} \cdot k_{ref,L} = \frac{k_{i,S}}{k_{ref,S}} \cdot k_{ref,L} \quad (3)$$

We emphasize here that this approach only requires the measurement of t_{ex}, t_m, and $t_{r,ref}$ for the long column to predict isocratic retention factors for any compound on the long column using retention measurements made using the short column.

2.2. Instrumental approach to high throughput measurements

When working with short columns like those used in this study, the actual separation times needed to acquire retention data over a large range in *k* are quite short. For example, the dead volume of 5 mm x 2.1 mm i.d. column packed with totally porous particles is about 10 µL (assuming a total porosity of 0.55, and neglecting frit volume). When used at a flow rate of 1.0 mL/min., the dead time is about 0.6 s. Even for a retention factor of 50, an analysis time of just 30 s is needed. When considering thousands of separations and such short analysis times, other factors associated with the measurement become significant, such as the time needed to draw a sample into an autosampler syringe for each analysis [20]. Faced with this reality, we developed the following instrumental approach for making high throughput retention measurements. The system, illustrated in Fig. 2, involves several conventional components: a binary UHPLC pump, autosampler, column thermostat compartment, and UV detector. Unique aspects of the configuration are: 1) the use of a four-port, two-position valve (this valve is normally used for 2D-LC applications) with two fixed volume internal loops for delivering small sample aliquots (about 150 nL) to the column; and 2) a low-pressure, single channel pump to push a sample stream from the autosampler to the injection valve. To acquire retention data across a range of mobile phase compositions, and thus retention factors, the following steps are followed. Figure 3 illustrates these steps and shows what the data string looks like at the detector for a complete data collection for one compound.

1) An aliquot of a sample containing the compound(s) of interest is drawn from a sample vial into the sampler needle and sample loop using the autosampler; in the work described here, this volume was 20 µL.

2) The sample is slowly displaced from the sample loop of the autosampler into the internal loop of the 4-port/2-position valve by the isocratic "flush pump". In this work the flow rate was typically about 1 µL/min.

3) After the internal loop of the 4-port/2-position valve has been filled with sample, the valve is switched, and data acquisition is initiated. The valve is switched an additional *m* times at time intervals that correspond to the desired analysis time. This provides *m* replicate injections of the sample at a given mobile phase composition.

4) The binary pump is then instructed to change to the next mobile phase composition, while continuing to switch the 4-port/2-position valve at regular intervals, all within the same data acquisition session. Data from the first injection after a change in mobile phase composition is ultimately discarded, and the time during this particular analysis is treated as an equilibration period. This leaves data *m*-1 replicate injections at each mobile phase composition. In the work described here *m*-1=5.

5) Step 4 is then repeated *n* times to acquire retention data for *n* different mobile phase compositions. This ultimately yields a datafile that contains *m* x *n* chromatograms that are parsed by simply dividing the entire data string into *m* x *n* equally-sized parts.

Figure 3 shows experimental data acquired using this process for the case where one compound is injected, thus we expect one peak per chromatogram. In this case $m = 3$ and $n = 4$, so we expect a total of $m \times n = 12$ peaks in the datafile. Starting from the left where the mobile phase is 50% ACN we see one peak that elutes early in the analysis interval. Moving to the right, as the % ACN is decreased, we see that the peak moves to the right (higher retention), as expected, with one peak per injection. When we get to 25% ACN, however, no peak is observed following the first injection. This is because the retention is too high for the peak to elute in the fixed analysis window of 30 s, and the peak actually elutes in the second analysis window after changing the mobile phase. During data processing we incorporate logic that checks to be sure we have exactly one peak per injection in the case where we have one compound per sample. If too few or too many peaks are observed, the datafile is inspected manually to make sure peak detection has worked

properly. In cases where multiple mobile phase compositions are used, a plot of ln(k) vs. % ACN is also constructed and visually inspected for discontinuities, which usually result from retention that is too high (i.e., like that shown in Fig. 3), and these data are then discarded.

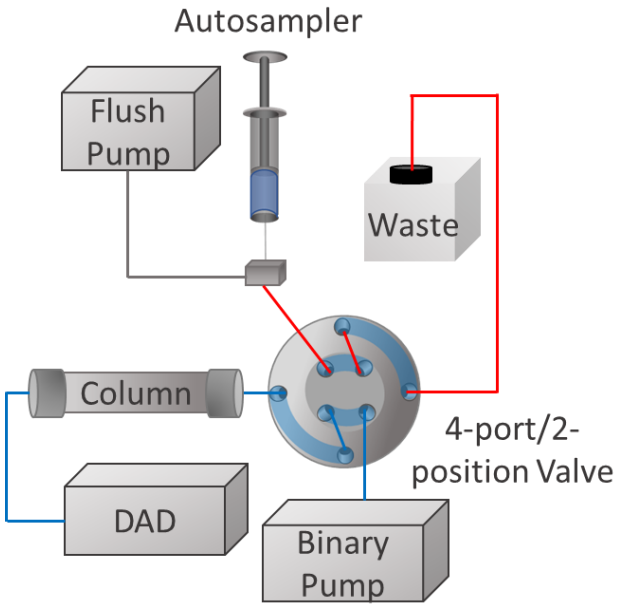


Figure 2. Illustration of the instrument setup used in this work.

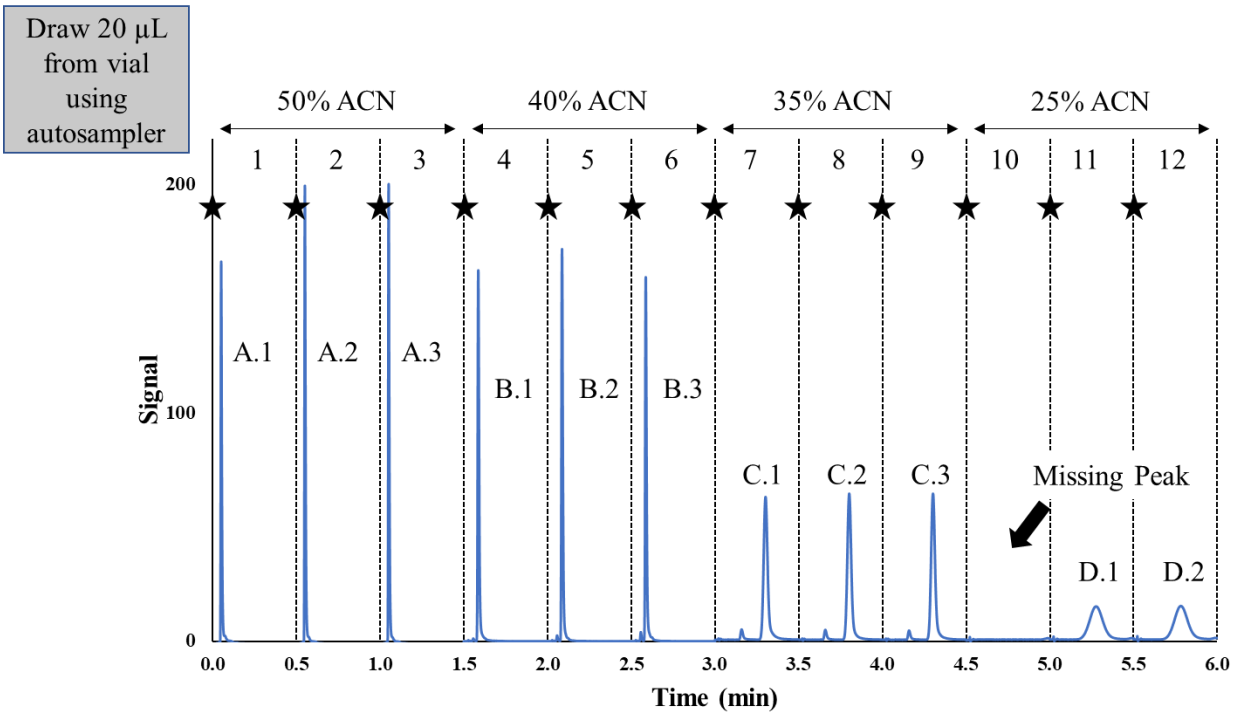


Figure 3. Illustration of the key steps in data acquisition and representative data for the case where retention data are acquired for a single compound in multiple mobile phases, but within a single datafile.

2.3. Effect of the measurement of "retention time" on apparent selectivity

One of the challenges encountered when working with low volume columns (e.g., the 5 mm x 2.1 mm i.d. columns used here) and conventional UHPLC instrumentation is that some degree of peak tailing due to extra-column flow paths is unavoidable. Peak tailing can also occur in short columns operated at high mobile phase velocities due to slow trans-column dispersion, and thus incomplete equilibration of the analyte zone across the column diameter [21]. In practice this means that the peaks observed with short columns tend to be more tailed than peaks observed for longer columns. This in turn can affect the apparent retention factors calculated from retention time as measured by the time corresponding to the peak apex. To quantify the magnitude of this effect, we carried out a simulation informed by realistic measures of the degree of peak tailing induced by extra-column flow paths between the point of sample injection and the point of detection (and the injector and detector elements themselves). The details associated with these simulations are described in detail and provided as Supplementary Information in Section S1. The important outcome from

these calculations is that the contribution to peak tailing from the instrument has a very small effect on the determination of alpha (k_i/k_{ref}) for long columns (e.g., 100 mm x 2.1 mm i.d.), but a practically significant effect on the determination of alpha for short columns (e.g., 5 mm x 2.1 mm i.d). The largest error in alpha introduced by using the time corresponding to the peak apex for the "retention time" over $0.1 < k < 50$ is less than 0.1% for the long column (see Fig. S3). However, errors on the order of 0.5% are possible for the short columns, and therefore we have chosen to use the first moment as the measure of "retention time" in all subsequent calculations of k and α going forward.

2.4. Determination of the first moment from raw data

To obtain accurate first moments to use as retention measurements for the calculation of retention factors, a curve-fitting strategy was applied to the raw chromatogram, and then the first moment of the resulting noise-free, fitted peak profile was calculated. The curve-fitting process was applied to a section of the chromatogram containing a peak. The time domain of this section was defined by $3.3 \cdot W_{0.5}$, where $W_{0.5}$ is the peak width at half-height, centered around the apex of the detected peak. This section is first baseline adjusted (i.e., to zero) and normalized such that the signal at the peak apex is 1. For curve fitting, a modified Pearson VII distribution [22], $f(t)$, was fit to the baseline-adjusted, normalized chromatographic peak:

$$f(t) = \left(1 + \frac{(t-\mu)^2}{M \cdot \left[\sigma + E \cdot (t-\mu) \right]^2} \right)^{-M} \tag{4}$$

where μ is the mean, σ the standard deviation, and E represents the asymmetry of the peak. M is correlated with the peak shape on a continuum from Chaucy ($M = 1$), to a modified Lorentzian, to a Gaussian as M approaches infinity (i.e., in practice $M > 10$) [23]. For the regression, the location of the apex of the peak (typically called the retention time), $W_{0.5}/2.35$, 0.15, and 5 were used as starting parameters for μ, σ, E and M, respectively. The latter two were determined earlier to be good estimates for most chromatographic peaks observed in practice [22]. While the algorithm was generally allowed to proceed for ten iterations, in most cases the residuals improved only marginally after four to five iterations.

Finally, the normalized first moment (m_1) of the peak (i.e., its center of gravity) was obtained by computing the first raw moment (M_1) and dividing it by the area of the peak (i.e., the zeroth moment, M_0), and used hereafter as the "retention time":

$$t_r = m_1 = \frac{M_1}{M_0} = \frac{\int_{-\infty}^{\infty} t \cdot f(t) \cdot dt}{\int_{-\infty}^{\infty} f(t) \cdot dt} \quad (5)$$

3. Experimental

3.1. Chemicals and columns

Acetonitrile, ammonium hydroxide (28-30%), formic acid, uracil, 5,5-diphenylhydantoin, acetophenone, benzonitrile, nortriptyline hydrochloride, amitriptyline hydrochloride, anisole, butyrophenone, n-butylbenzoic acid, toluene, ethylbenzene, mefenamic acid were obtained from Sigma-Aldrich and used as received. Alfa Aesar (Tewksbury, MA) was the supplier of trans-chalcone, p-nitrophenol was obtained from Eastman Kodak (Rochester, NY). Cis-chalcone was prepared by exposing a solution of trans-chalcone to sunlight, resulting in a solution enriched with the cis isomer. The cis isomer was purified by collecting the cis isomer fraction after separation on a C18 column. HPLC grade water was obtained from an in-house Milli-Q system (Burlington, MA). Stock solutions were prepared for each compound at 10 mg/mL stock using ACN as the diluent; in cases where the compound was not soluble in neat ACN, 50/50 ACN/water was used as the diluent. Analytical samples were prepared in 50/50 ACN/buffer, with analyte concentrations ranging from 0.1 to 2.5 mg/mL as needed to provide a peak height above 10 mAU at 254 nm. All measurements for short (5 mm) columns were made with one analyte per sample. Measurements with the long column (100 mm) were made with mixtures of analytes per sample (typically five analytes per mixture), except for the data shown in Fig. S7, where analytes were injected separately (i.e., one analyte per sample).

The columns were both from Agilent, packed with Zorbax SB-C18 particles (1.8 μm): 5 mm x 2.1 mm i.d. (p/n: 821725-902); 100 mm x 2.1 mm i.d. (p/n: 858700-902). Note that these two columns

were not prepared from the same batch of stationary phase particles, thus at least some differences in the selectivities of the two columns is to be expected (i.e., lot-to-lot variability) [24]. In a brief follow-up study, we did obtain a "matched pair" of short and long columns prepared from the same batch of stationary phase; the results from these measurements are discussed in Section S2.

3.2. Buffer preparation

Batches of 25 mM ammonium formate buffer pH 3.2 (105 mM with respect to formic acid) were prepared in two-liter portions using water, formic acid, and ammonia. To improve batch-to-batch repeatability of the buffer when using different lots of concentrated formic acid and ammonia, the weight percent of formic acid or ammonia as reported in the Certificate of Analysis (COA) for that material was used to calculate the mass of solution needed to achieve the desired concentration of the buffer components in the buffer solution. Each batch of buffer was prepared gravimetrically using a balance with a capacity of 4 kg, a 2-L glass bottle, and 1982.6 g of HPLC grade water. The mass of formic acid required to obtain a formal concentration of 105 mM was added, followed by the mass of ammonia required to obtain a formal concentration of 25 mM. Before and after the addition of ammonia, the bottle was shaken briefly by hand, and the solution was used for analysis without any further treatment.

3.3. Instrumentation and methods

The components of the system used for all retention measurements (short and long columns, isocratic and gradient elution) are illustrated in Fig. 2. All components were from Agilent Technologies with model numbers as follows: Flush pump, G5611A; Binary pump, G4220A; 4-port/2-position prototype valve, G5067-4236A; Thermostated column compartment, G1316C; Diode-array UV absorbance detector, G4212A (flow cell part number 4212-60008). The gradient delay volume between the mixing point of the binary pump and the inlet of the column was determined by installing a union in place of the analytical column and running a gradient from 5/95 to 95/5 B/A where A was 50/50 ACN/water and B was A spiked with 10 µg/mL uracil. Using this approach the delay volume was determined to be 46 µL. Column dead times ($t_{m,meas}$) and extra-column times (t_{ex}) were determined by injecting a 10 µg/mL sample of uracil in 50/50 ACN/water

into a mobile phase of 50/50 ACN/water at either 0.1 (long column) or 1.0 (short column) mL/min. We are well aware that this method does not produce the most accurate measure of the column dead time [25]; we use this approach in the interest of measurement throughput because it is straightforward to incorporate as part of the measurement workflow for other compounds. However, the magnitude of the error is about the same for the short and long columns, and thus much of the error cancels out in any comparison of selectivities for the two columns; when $k_{ref} \sim 5$, as in this work, the absolute error in α is about 1/5 of the error in t_m for $k > 1$. The system was controlled using Agilent OpenLAB CDS Chemstation Edition (Rev. C.01.07 [465]). Chromatographic conditions are given in the figure captions. Note that we deliberately chose flow rates of 0.1 and 1.0 mL/min for the long and short columns, respectively, to avoid significant effects of viscous heating and pressure on retention and selectivity (i.e., the column midpoint pressure is about 50 bar for both columns under these conditions).

4. Results and Discussion

4.1. Initial comparison of retention factors obtained from short and long columns

The magnitude of variation in typical sets of measurements (as measured by relative standard deviation, with n = 6) of t_{ex} and t_m were on the order of 0.25 and 0.05% for the 100 mm column (0.1 mL/min), and 0.45 and 0.30% for the 5 mm column (1.0 mL/min). The grey bars of Fig. 4 show the percent difference between retention factors calculated for short (5 mm) and long (100 mm) columns as described in Section 2.1 using Eq. 1. These differences are on the order of 18%. Given the excellent lot-to-lot reproducibility of modern stationary phases from main-line manufacturers it is highly unlikely that a difference of this magnitude can be explained by lot-to-lot variability, especially for the relatively simple molecules studied here. A likely explanation for the major differences in the retention factors determined for the two columns is that the volume of the inlet and outlet frits contributes to the measured column dead times to different extents, but cannot contribute to the actual retention time because there is no stationary phase in the frit. We note that several groups have studied the impact of analyte dispersion in the column endfittings and frits on peak width [26–30], however we are not aware of any thorough discussion of the volume associated with the endfittings and frits on apparent retention factors. Although it is

certainly true that these volumes must affect apparent retention factors, we initially were unsure if the magnitude of the effect could explain most of the differences observed in Fig. 4. The following theoretical calculations were used to produce the trend in Fig. 5, which ultimately shows that the effect of the frit volume on the apparent retention factor is indeed large enough to explain most of the differences shown in Fig. 4.

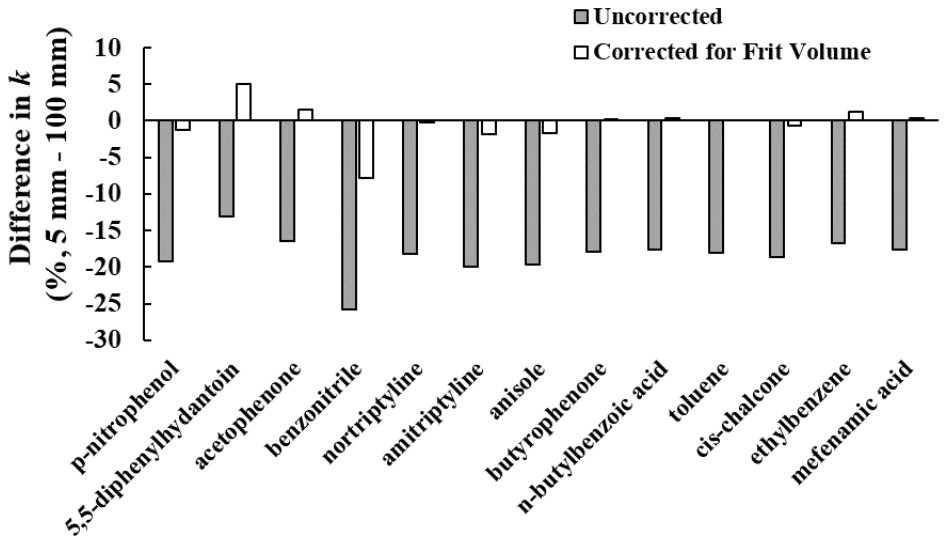


Figure 4. Percent differences in retention factors (k) calculated for short (5 mm) and long (100 mm) columns using retention measurements based in first moments and Eq. 1 (grey bars) or Eq. 8 (white bars) assuming a total frit volume of 2.4 µL. Chromatographic conditions: Flow rate, 1.0 (short) or 0.1 (long) mL/min.; Mobile phase, 50/50 ACN/25 mM ammonium formate in water, pH 3.2; Temperature, 40 °C.

We start by assuming a known retention factor of 1.00 for a hypothetical solute, dead volumes of 0.010 and 0.200 mL ($V_{m,col}$) for two columns that only vary in length (these are the approximate dead volumes of the 5 mm and 100 mm x 2.1 mm i.d. discussed in this paper), and a flow rate of 1.0 mL/min. We also assume that retention measurements are made on a system with an extra-column volume (V_{ex}) of 0.020 mL. We then choose a total frit volume for each column (*i.e.*, the sum of the volumes of the inlet and outlet frits; V_{frit}), calculate the retention factor that will be measured under these conditions for each column, and finally the difference between them. If the flow rate used for the two columns is the same, then we can convert all these volumes to times as in Eq. 6:

$$t_{m,col} = \frac{V_{m,col}}{F};\ t_{ex} = \frac{V_{ex}}{F};\ t_{frit} = \frac{V_{frit}}{F} \tag{6}$$

The measured dead time of the column ($t_{m,meas}$) will be the sum of all these times:

$$t_{m,meas} = t_{m,col} + t_{ex} + t_{frit} \tag{7}$$

The measured retention time can be calculated in a similar way, using the usual relationship between the retention factor, retention time, and dead time:

$$k_{col} = \frac{\left(t_{r,col} - t_{m,col}\right)}{t_{m,col}};\ t_{r,col} = t_{m,col}\left(1 + k_{col}\right) \tag{8}$$

$$t_{r,meas} = t_{r,col} + t_{ex} + t_{frit} \tag{9}$$

The retention factor determined from the measured retention, dead, and extra-column times (k_{exp}) can be calculated as usual (Eq. 1), but repeated here in explicit terms:

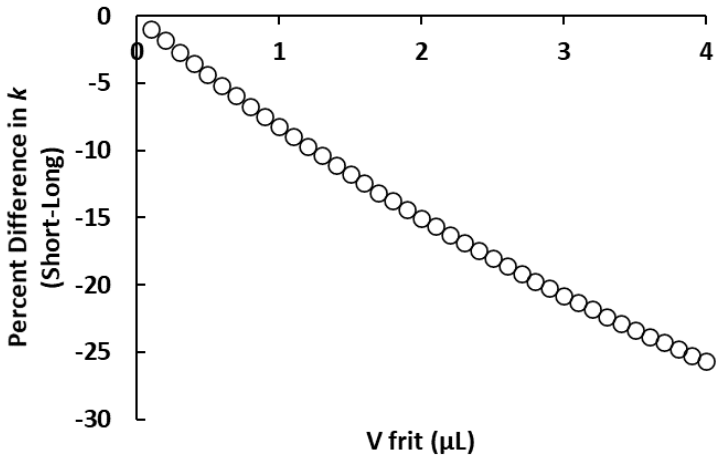
$$k_{\exp} = \frac{\left(t_{r,meas} - t_{ex}\right) - \left(t_{m,meas} - t_{ex}\right)}{\left(t_{m,meas} - t_{ex}\right)} \tag{10}$$

Substituting Eqs. 7 and 9 into Eq. 10 we find that all the experimental non-idealities (t_{ex}) cancel except for t_{frit}:

$$k_{\exp} = \frac{\left(t_{r,col}\right) - \left(t_{m,col}\right)}{\left(t_{m,col} + t_{frit}\right)} \tag{11}$$

Whenever t_{frit} and t_{ex} are both fixed and non-zero, but $t_{m,col}$ varies – as in the comparison of 5 and 100 mm columns in Fig. 4 – the calculated retention factors (k_{exp}) for the two columns will not be the same, and the apparent k value for the shorter column will always be smaller than that for the longer column. While there may be other reasons for differences in experimentally determined retention factors for columns of different lengths, the issue described here is purely physical in nature. The resulting differences in k_{exp} for the 5 and 100 mm columns as a function of frit volume for the conditions described here are shown in Fig. 5. If we assume for a moment that all the difference shown in the grey bars of Fig. 4 can be attributed to the unaccounted-for frit volume, then this relationship suggests that the total frit volume in these two columns is about 2.4 µL. This

391 value is entirely consistent with estimates of the interstitial volume of the frits provided by the
392 vendor in this case (*i.e.*, the estimated volume of each inlet and outlet frit is about 1.2 µL).



393

394 **Figure 5.** Theoretical percent difference in retention factors that arise from values calculated from
395 experimental measurements where the column frit volume (V_{frit}) is a significantly different fraction of the
396 measured dead volumes ($V_{m,meas}$) of columns of different lengths (S=short; L=long). It is assumed that the
397 columns are otherwise identical in terms of stationary phase chemistry and particle size. Other parameters:
398 Column volumes, 0.010 (short) and 0.200 (long) mL; Column diameters, 2.1 mm; Flow rate, 1.0 mL/min.;
399 Extra-column volume (V_{ex}), 0.020 mL.

400

401 One approach to deal with the major effect of the frit volume on the retention factors calculate for
402 short columns from experimental data is to add t_{frit} to t_{ex} when calculating k as in Eq. 10. Doing so
403 with our data yields the white bars in Fig. 4. Here we see that this removes most of the apparent
404 difference between the k values for the short and long columns, with the average difference close
405 to zero (-0.4%), rather than the average difference of -18% prior to the correction (grey bars). In
406 principle such a correction would be straightforward if the frit volumes were known, however
407 these numbers are not typically provided by column manufacturers, and are difficult to measure
408 accurately without dedicated equipment for doing so. Having first observed the major differences
409 in k_{exp} as in Fig. 4, and then realizing that most of this difference could be attributed to frit volumes
410 that are impractical to measure in practice, motivated us to pursue the use of experimentally
411 determined selectivities to translate retention measurements made with small columns to predict
412 retention in separations involving larger columns as outlined in Section 2.1.

4.2. Comparison of selectivities determined using short and long columns

Figure 6 shows the percent differences in α values ($\alpha = k_x/k_{toluene}$) calculated for the 5 and 100 mm columns using retention measurements based on first moments determined as described in Section 2.4. This particular plot is organized with the compounds listed from left to right in order of increasing retention. Although some of the differences are clearly different in magnitude from the others, there is no obvious dependence of the differences on analyte type (*i.e.*, acid, base, neutral). The average absolute difference in α is 1.7%. Although there is no clear pattern in the differences shown in Fig. 6 related to analyte chemistry, one might reasonably ask if the differences are retention dependent. Figure S4 shows the same differences as in Fig. 6, but plotted against retention factor. Here we see the overall trend that the absolute magnitude of the difference decreases with increasing retention factor, however the sign of the difference is not consistent at low retention. This is not unexpected considering that the relative variation in retention measurements increases as absolute retention decreases (i.e., when the absolute variation in retention measurement is nominally independent of *k*), however it suggests that it is important when using the measurement scheme proposed here that we focus primarily on retention factors above about 2. The percent relative standard deviations in retention factors used in the calculation of alpha values shown in Fig. 6 are shown as Supplemental Information in Fig. S5. The primary takeaway from Fig. 6 is that similar alpha values are obtained from the two columns that vary in volume by a factor of 20, despite the 18% differences in apparent retention factor values shown in Fig. 4. However, the α values for 5,5-diphenylhydantoin and benzonitrile differed by more than 2%. To check if small variations in mobile phase composition as a result of mobile phase preparation by the pump (i.e., mixing ACN/buffer mobile phase from neat ACN and buffer) influenced this comparison, we compared the α values obtained from retention measurements while letting the pump prepare the mobile phase (referred to here as 'machine-mixed) to those obtained with a pre-mixed mobile phase (both 50/50 v/v); this comparison is shown in Fig. S6. In the case of 5,5-diphenylhydantoin and benzonitrile we see that the errors are similar in magnitude, and have the same sign, which means that errors cannot be explained by small differences in mobile phase variation over the timescale of a retention measurement (i.e., a tens of seconds).

One possible cause of the larger differences in the alphas observed for 5,5-diphenylhydantoin and benzonitrile could be related to column-to-column variation in the stationary phase (*i.e.*, the 5 and 100 mm columns used for most of this work were packed with different manufacturing lots of stationary phase). To test this possibility, we repeated the comparison of α values for 5,5-diphenylhydantoin and benzonitrile using a 5 and 100 mm column pair that were packed from the same lot of packing material, using one analyte per injected sample for both columns. The resulting differences in alphas were -0.58 and 0.71%, respectively, as shown in Fig. S7, which are in line with the other small differences shown in Fig. 6.

Finally, we emphasize once more that we do not expect the approach described here to yield retention factors with the highest possible accuracy, in a thermodynamic sense. Determination of thermodynamically meaningful retention factors requires careful consideration of how the column dead time is measured [25,31], in addition to careful control of other parameters including the column temperature and mobile phase composition.

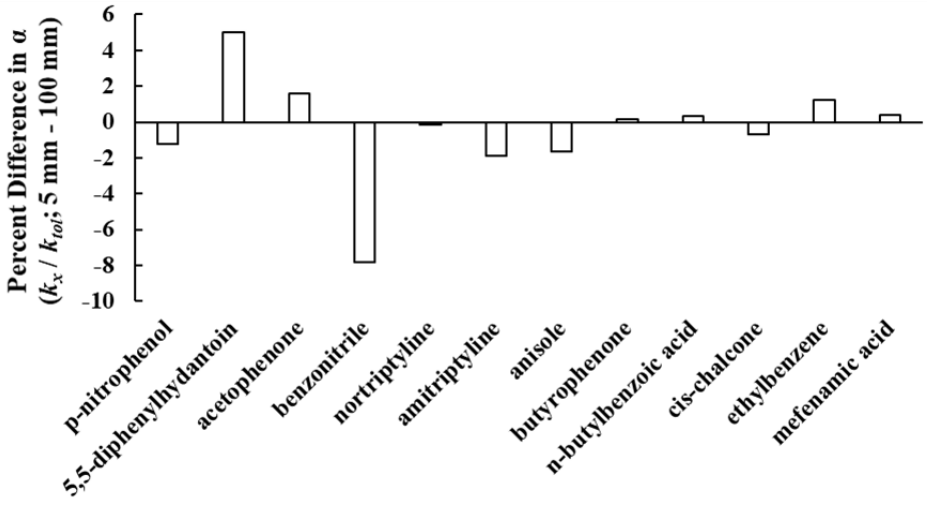


Figure 6. Percent differences in selectivities (α) calculated for 5 and 100 mm columns using retention measurements based on first moments and Eq. 2. Chromatographic conditions were as described in Fig. 4.

4.3. Prediction of isocratic retention factors using data from isocratic or gradient elution, and short or long columns

The preceding discussion has been focused on the prediction of isocratic retention factors for a long column using isocratic retention measurements made using a short column and the scheme outlined in Section 2.1. In principle, isocratic retention factors can also be predicted using retention measurements made under gradient elution conditions [19]. This would correspond to path #3 in Fig. 1. The white bars of Fig. 7 show the percent differences between isocratic retention factors predicted from retention measurements using gradient times of 10, 20, or 30 min with the 100 mm long column to isocratic retention factors calculated from isocratic retention measurements made using the same column. To make these predictions we first fit the gradient retention times to the non-linear Neue-Kuss model of the dependence of RP retention on volume fraction of organic modifier in the mobile phase (ϕ) [26]. The relationship between the effective gradient retention factor (k_{eff}) and ϕ for this model is shown in Eq. 12, where S_1, S_2, and k_w are the fitting parameters, and t_d and ϕ_i are the gradient delay time and the starting mobile phase composition used in the gradient, respectively. Before fitting the data k_{eff} was calculated using the first moment for the retention time (see Section 2.4), and Eq. 10. For the actual fitting of the retention data we used the *lsqnonlin* function in MATLAB as described in our recent publication on determination of retention model parameters [19].

$$k_{eff} = \frac{t_d}{t_m} + \frac{\dfrac{\phi_i + \dfrac{1+S_2\phi_i}{S_1}\ln\left\{\beta\cdot k_w\cdot S_1\left(t_m - \dfrac{t_d}{k_i}\right)\exp\left(\dfrac{-S_1\phi_i}{1+S_2\phi_i}\right)+1\right\}}{1-\dfrac{S_2\left(1+S_2\phi_i\right)}{S_1}\ln\left\{\beta\cdot k_w\cdot S_1\left(t_m - \dfrac{t_d}{k_i}\right)\exp\left(\dfrac{-S_1\phi_i}{1+S_2\phi_i}\right)+1\right\}} - \phi_i}{\beta\cdot t_m} \qquad (12)$$

Once the model parameters (S_1, S_2, k_w) have been determined via Eq. 12 or Eq. 13, isocratic retention factors can then be calculated for any mobile phase composition using Eq. 13.

$$\ln k = \ln k_w + 2\ln\left(1+S_2\phi\right) - \left[\frac{S_1\phi}{1+S_2\phi}\right] \qquad (13)$$

As shown by the white bars in Fig. 7, we see that the performance of these predictions (*i.e.*, prediction of isocratic k for the 100 mm column from gradient retention data obtained using the 100 mm column) is not good, with a maximum error of -35% for 5,5-diphenylhydantoin, and a mean error of -6.3% for all 13 probe compounds. On the other hand, the isocratic retention factors

predicted for the long column using isocratic measurements made using the short column (grey bars) are much better. In this case the maximum error is -7.8%, and the mean error is -0.4%. The poor performance of predicting isocratic *k* values from gradient retention times (i.e., Path #3 in Fig. 1) is not surprising, since such predictions involve a major extrapolation to a gradient slope of zero [27]. Nevertheless, this result adds to the value of the use of α values to translate retention values between short and long columns as described in Section 2.1. If one must choose between these two approaches to predict isocratic retention factors for a long column, this result shows that predicting isocratic retention factors for long columns using isocratic retention times measured using short columns is far more accurate than predicting isocratic retention factors from retention times measured under gradient elution conditions.

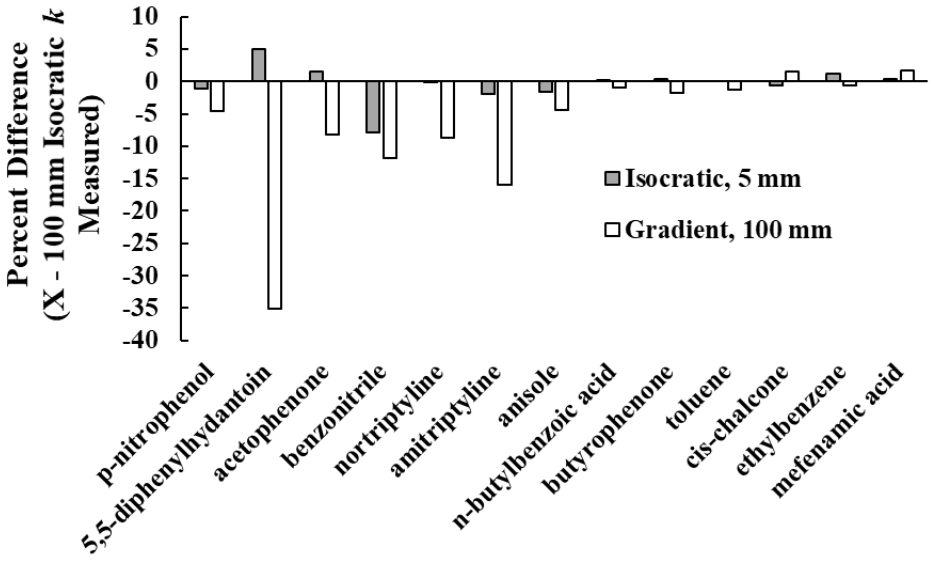


Figure 7. Percent differences between predicted and measured retention factors (*k*) for the 100 mm column. Grey bars show the difference between measured values (100 mm column) and values calculated from measurements using the 5 mm column but corrected using Eqs. 1-3. White bars show the differences between measured values (100 mm column) and values calculated by fitting gradient elution retention times to the Neue-Kuss model of reversed-phase retention as described in ref. 21. Chromatographic conditions for the isocratic measurements are the same as in Fig. 4. For the gradient measurements, gradient times of 10, 20, and 30 min. were used, with a gradient running from 5 to 60% ACN; other conditions were the same as in the isocratic experiments.

4.4.Prediction of gradient elution retention times using isocratic or gradient elution retention data, and short or long columns

In a final comparison we evaluated the ability to predict gradient elution retention times for a long column from either isocratic retention measurements made using the short column (Path #2 in Fig. 1), or retention measurements made under gradient elution conditions using the long column (Path #4 in Fig. 1). Path #2 requires that Neue-Kuss model parameters are first obtained by fitting isocratic k values determined for several different isocratic mobile phase compositions using Eq. 13. In this work we used k values (after translating measurements made using the 5 mm column to the 100 mm column as in Eq. 3) for five or six mobile phases (covering a range in k of about 1 to 10) to obtain retention model parameters for each compound. Then, the resulting model parameters can be used to predict a gradient elution retention time using Eq. 12.

Figure 8 shows the percent differences between gradient elution retention times predicted from Paths #1 and 4, and gradient elution retention times measured using the long column and a gradient time of 20 min. As shown by the white bars, the accuracy of prediction using the gradient elution retention times (Path #4 in Fig. 1) is incredibly good, with a maximum difference of -0.36% for 5,5-diphenylhydantoin, and a mean error of -0.03%. This is consistent with an extensive body or prior work showing similarly good results for this approach (e.g., see [27]). The grey bars in Fig. 8 show that the prediction of gradient elution retention times using isocratic measurements made with the short column (Path #2 in Fig. 1) is not nearly as good, but not terrible. Here the maximum error is -3.9%, with a mean error of -0.65%. Given that the errors for this approach increase with decreasing gradient elution retention time, it is conceivable that small errors that effectively cancel out in Path #4 (e.g., error in the determination of gradient delay volume, deviation of the solvent composition arriving at the column inlet from a simple linear gradient [10,28]) are exposed in Path #2.

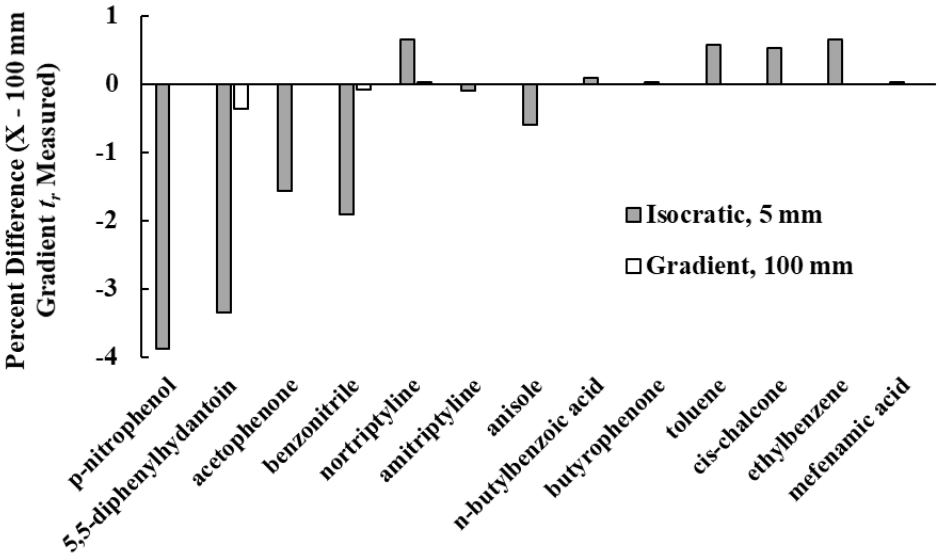


Figure 8. Percent differences between predicted and measured gradient elution retention times (t_r) for the 100 mm column. Grey bars show the difference between measured values (100 mm column) and values calculated from measurements using the 5 mm column but corrected using Eqs. 1-3. White bars show the differences between measured values (100 mm column) and values calculated by fitting gradient elution retention times to the Neue-Kuss model of reversed-phase retention as described in ref. 21. Chromatographic conditions for the isocratic measurements are the same as in Fig. 4. For the gradient measurements, gradient times of 10, 20, and 30 min. were used, with a gradient running from 5 to 60% ACN; other conditions were the same as in the isocratic experiments.

5. Conclusions

Accurate isocratic retention data are needed for a variety of applications of liquid chromatography ranging from fundamental research to practical method development. In this work we have explored an approach using low volume columns that minimizes the time needed for each retention measurement, thereby increasing the throughput of data collection for a single instrument. As the volume of the column used to make retention measurements is decreased, factors that are normally relatively inconsequential, such as inlet and outlet frit volumes, become more important and can compromise the accuracy of retention measurement. Fundamentally, retention is a thermodynamic property of the mobile and stationary phase combination under study, and should be nominally independent of column dimensions. We propose using measured selectivities (i.e., ratios of

retention factors) as column geometry-independent measures of retention that can be used to mitigate the effects of non-idealities such as frit volumes on retention measurements. After comparing measured retention data from short (5 mm) and long (100 mm) 2.1 mm i.d. reversed-phase columns, we have come to the following primary conclusions about difficulties associated with retention measurements from low volume columns and the benefits of our approach proposed here:

1) Errors in retention factors measured using the short (5 mm) column are on the order of 20% when compared to a long (100 mm) column packed with the same stationary phase. We attribute most of this difference to the volume of the inlet and outlet frits that contributes disproportionately to measured dead times and retention times.

2) Using the correction scheme based on selectivities, the apparent difference between the retention factors of 13 test analytes on the short and long columns can be reduced to an average absolute difference of 1.7% (all errors less than 8%).

3) The correction scheme described here should facilitate more rapid method development by collecting data needed to build retention models that can then be used to predict optimal separation conditions. The scheme should also enable building of large retention databases that can be used to deepen our understanding of retention in different separation modes (e.g., reversed-phase, ion-exchange, etc.), and support other aspects of method development, such as the effect of mobile phase mismatch in two-dimensional LC separations.

The approach demonstrated here has so far relied on single-channel UV detection, and one analyte per sample to facilitate data processing. With these parameters the approach can yield about 2,000 isocratic retention measurements per instrument per day (assuming an overhead of 25% of time dedicated to quality control and instrument overhead). Any effort to multiplex measurements by working with multiple analytes per sample – for example by using mass spectrometric detection or multi-channel UV detection – may further increase the throughput of retention measurement.

6. Acknowledgements

D.S., T.K., and T.D. have been supported by a grant from the National Science Foundation (CHE2003734). G.K.'s exchange semester at Gustavus Adolphus College was made possible by a scholarship from the Vienna-based Austrian Marshall Plan Foundation. All instrumentation and columns used in this work were provided by Agilent Technologies.

7. References

[1] M.J. den Uijl, P.J. Schoenmakers, B.W.J. Pirok, M.R. Bommel, Recent applications of retention modelling in liquid chromatography, J. Sep. Sci. 44 (2021) 88–114. https://doi.org/10.1002/jssc.202000905.

[2] D.R. Stoll, R.W. Sajulga, B.N. Voigt, E.J. Larson, L.N. Jeong, S.C. Rutan, Simulation of elution profiles in liquid chromatography − II: Investigation of injection volume overload under gradient elution conditions applied to second dimension separations in two-dimensional liquid chromatography, J. Chromatogr. A. 1523 (2017) 162–172. https://doi.org/10.1016/j.chroma.2017.07.041.

[3] D. Stoll R., B.W.J. Pirok, Perspectives on the use of retention modeling to streamline 2D-LC method development: Current state and future prospects, LC GC N. Am. 40 (2022) 30–34.

[4] S. Chapel, F. Rouvière, V. Peppermans, G. Desmet, S. Heinisch, A comprehensive study on the phenomenon of total breakthrough in liquid chromatography, J. Chromatogr. A. 1653 (2021) 462399. https://doi.org/10.1016/j.chroma.2021.462399.

[5] F. Gritti, M. Gilar, J. Hill, Mismatch between sample diluent and eluent: Maintaining integrity of gradient peaks using in silico approaches, J. Chromatogr. A. 1608 (2019) 460414. https://doi.org/10.1016/j.chroma.2019.460414.

[6] A.G. Usman, S. Işik, S.I. Abba, A Novel Multi-model Data-Driven Ensemble Technique for the Prediction of Retention Factor in HPLC Method Development, Chromatographia. 83 (2020) 933–945. https://doi.org/10.1007/s10337-020-03912-0.

[7] Z. Liu, J.P. Foley, Are two liquid chromatography columns in tandem better than one?: Answers from the hydrophobic subtraction model, Journal of Chromatography A. 1668 (2022) 462890. https://doi.org/10.1016/j.chroma.2022.462890.

[8] L.N. Jeong, R. Sajulga, S.G. Forte, D.R. Stoll, S.C. Rutan, Simulation of elution profiles in liquid chromatography—I: Gradient elution conditions, and with mismatched injection and mobile phase solvents, J. Chromatogr. A. 1457 (2016) 41–49. https://doi.org/10.1016/j.chroma.2016.06.016.

[9] A.R. Horner, R.E. Wilson, S.R. Groskreutz, B.E. Murray, S.G. Weber, Evaluation of three temperature- and mobile phase-dependent retention models for reversed-phase liquid chromatographic retention and apparent retention enthalpy, J. Chromatogr. A. 1589 (2019) 73–82. https://doi.org/10.1016/j.chroma.2018.12.055.

[10] T.S. Bos, L.E. Niezen, M.J. den Uijl, S.R.A. Molenaar, S. Lege, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Reducing the influence of geometry-induced gradient deformation in liquid chromatographic retention modelling, J. Chromatogr. A. 1635 (2021) 461714. https://doi.org/10.1016/j.chroma.2020.461714.

[11] I.A. Haidar Ahmad, A. Kiffer, R.C. Barrientos, G.L. Losacco, A. Singh, V. Shchurik, H. Wang, I. Mangion, E.L. Regalado, *In Silico* Method Development of Achiral and Chiral Tandem Column Reversed-phase Liquid Chromatography for Multicomponent Pharmaceutical Mixtures, Anal. Chem. (2022) acs.analchem.1c05551. https://doi.org/10.1021/acs.analchem.1c05551.

[12] Y. Mao, P.W. Carr, Adjusting Selectivity in Liquid Chromatography by Use of the Thermally Tuned Tandem Column Concept, Anal. Chem. 72 (2000) 110–118. https://doi.org/10.1021/ac990638x.

[13] D.R. Stoll, T.A. Dahlseid, S.C. Rutan, T. Taylor, J.M. Serret, Improvements in the predictive accuracy of the hydrophobic subtraction model of reversed-phase selectivity, J. Chromatogr. A. (2020) 461682. https://doi.org/10.1016/j.chroma.2020.461682.

[14] D. Abate-Pella, D.M. Freund, Y. Ma, Y. Simón-Manso, J. Hollender, C.D. Broeckling, D.V. Huhman, O.V. Krokhin, D.R. Stoll, A.D. Hegeman, T. Kind, O. Fiehn, E.L. Schymanski, J.E. Prenni, L.W. Sumner, P.G. Boswell, Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods, J. Chromatogr. A. 1412 (2015) 43–51. https://doi.org/10.1016/j.chroma.2015.07.108.

[15] J.W. Dolan, L.R. Snyder, The Hydrophobic-Subtraction Model for Reversed-Phase Liquid Chromatography: A Reprise, LCGC North America. 34 (2016) 730–741.

[16] C.J. Welch, X. Gong, W. Schafer, E.C. Pratt, T. Brkovic, Z. Pirzada, J.F. Cuff, B. Kosjek, MISER chromatography (multiple injections in a single experimental run): the chromatogram is the graph, Tetrahedron: Asymmetry. 21 (2010) 1674–1681. https://doi.org/10.1016/j.tetasy.2010.05.029.

[17] K.D. Berthelette, T.H. Walter, M. Gilar, F. Gritti, T.S. MacDonald, M. Soares, Evaluating MISER chromatography as a tool for characterizing HILIC column equilibration, J. Chromatogr. A. 1619 (2020) 460931. https://doi.org/10.1016/j.chroma.2020.460931.

[18] C.J. Welch, Are We Approaching a Speed Limit for the Chromatographic Separation of Enantiomers?, ACS Cent. Sci. 3 (2017) 823–829. https://doi.org/10.1021/acscentsci.7b00250.

[19] T. Brau, B. Pirok, S. Rutan, D. Stoll, Accuracy of retention model parameters obtained from retention data in liquid chromatography, J. Sep. Sci. (2022) jssc.202100911. https://doi.org/10.1002/jssc.202100911.

[20] M. Gilar, T.S. McDonald, F. Gritti, Impact of instrument and column parameters on high-throughput liquid chromatography performance, J. Chromatogr. A. 1523 (2017) 215–223. https://doi.org/10.1016/j.chroma.2017.07.035.

[21] A. Daneyko, D. Hlushkou, S. Khirevich, U. Tallarek, From random sphere packings to regular pillar arrays: Analysis of transverse dispersion, J. Chromatogr. A. 1257 (2012) 98–115. https://doi.org/10.1016/j.chroma.2012.08.024.

[22] B.W.J. Pirok, N. Abdulhussain, T. Brooijmans, T. Nabuurs, J. de Bont, M.A.J. Schellekens, R.A.H. Peters, P.J. Schoenmakers, Analysis of charged acrylic particles by on-line comprehensive two-dimensional liquid chromatography and automated data-processing, Anal. Chim. Acta. 1054 (2019) 184–192. https://doi.org/10.1016/j.aca.2018.12.059.

[23] B.W.J. Pirok, Westerhuis, J. A., Challenges in obtaining relevant information from one- and two-dimensional LC experiments, LC GC N. Am. (2020) 8–14.

[24] U.D. Neue, E. Serowik, P. Iraneta, B.A. Alden, T.H. Walter, Universal procedure for the assessment of the reproducibility and the classification of silica-based reversed-phase packings, Journal of Chromatography A. 849 (1999) 87–100. https://doi.org/10.1016/S0021-9673(99)00435-5.

[25] D. Cabooter, H. Song, D. Makey, D. Sadriaj, M. Dittmann, D. Stoll, G. Desmet, Measurement and modelling of the intra-particle diffusion and b-term in reversed-phase liquid chromatography, Journal of Chromatography A. 1637 (2021) 461852. https://doi.org/10.1016/j.chroma.2020.461852.

[26] U.D. Neue, H.-J. Kuss, Improved reversed-phase gradient retention modeling, J. Chromatogr., A. 1217 (2010) 3794–3803. https://doi.org/10.1016/j.chroma.2010.04.023.

[27] M.J. den Uijl, P.J. Schoenmakers, G.K. Schulte, D.R. Stoll, M.R. van Bommel, B.W.J. Pirok, Measuring and using scanning-gradient data for use in method optimization for liquid chromatography, Journal of Chromatography A. 1636 (2021) 461780. https://doi.org/10.1016/j.chroma.2020.461780.

[28] P.G. Boswell, J.R. Schellenberg, P.W. Carr, J.D. Cohen, A.D. Hegeman, A study on retention "projection" as a supplementary means for compound identification by liquid chromatography–mass spectrometry capable of predicting retention with different gradients, flow rates, and instruments, J. Chromatogr. A. 1218 (2011) 6732–6741. https://doi.org/10.1016/j.chroma.2011.07.105.

683