

Introduction to Chromatography

David T. Harvey

Introduction

This learning module provides an introduction to chromatography that is designed for an introductory course in analytical chemistry. The module consists of nine investigations:

- Investigation 1: A First Look at a Chromatogram
- Investigation 2: A Closer Look at the Separation
- Investigation 3: The Partition Coefficient and the Retention Factor
- Investigation 4: Theoretical Plates
- Investigation 5: A First Look at Chromatographic Resolution
- Investigation 6: A Closer Look at Chromatographic Resolution
- Investigation 7: Improving Resolution Through Column Efficiency
- Investigation 8: Improving Resolution Through Column Selectivity
- Investigation 9: Improving Resolution Through Column Retention

The learning module is programmed in R (www.r-project.org) using the **Shiny** package, which allows for interactive features. Each investigation includes a brief introduction, an explanation of the controls—sliders, radio buttons, and point selection—available to the user and the type of output produced by the underlying code. Each investigation also includes one or more questions to answer.

The purpose of this document is to provide instructors with additional background on the program’s features and data sets, to provide representative examples of the results students might generate, and to provide suggestions of possible ways to make use of the module’s investigations.

Some Background Details on the Learning Module’s Data

The chromatograms—both full and partial—are simulated using functions written in R. The chromatograms in Investigations 1-4 model the separation as a counter-current extraction using the approach outlined in the paper “Tutorial: Simulating Chromatography with Microsoft Excel Macros,” the full reference for which is Kadjo, A.; Dasgupta, P. K. *Analytica Chimica Acta* **2013**, *773*, 1-8. The reason for using this approach is that it makes it possible to visualize the separation on the column as it takes place.

The partial chromatograms in Investigations 5 and 6 are simulated as Gaussian functions with each peak defined by its retention time ($\bar{X} = t_r$) and its width at the baseline ($\sigma = w/4$). For the calculations in Investigation 6, the value of t_m is set to 1 min. The retention factor for the later eluting solute, k_B is used to calculate $t_{r,B}$ and then this value and α are used to calculate $t_{r,A}$. The width of each peak is calculated using the number of theoretical plates, N , and the analyte’s retention time.

The data in Investigation 7 is drawn from the paper “The Evaluation of the Parameters in the van Deemter Equation,” the full reference for which is Moody, H. W. *J. Chem. Educ.* **1982**, *59*, 290-291. Retention times and peak widths for a single analyte, 2-butanone, were measured using different flow rates of the carrier gas and the average height of a theoretical plate calculated and reported.

The data in Investigation 8 is drawn from the paper “Optimization of HPLC and GC Separations Using Response Surfaces: Three Experiments for the Instrumental Analysis Laboratory,” the full reference for which is Harvey, D. T.; Byerly, S.; Bowman, A.; and Tomlin, J. *J. Chem. Educ.* **1991**, *68*, 162-168. Retention times for each analyte are reported for the pure mobile phases—20% v/v methanol (M), 16% v/v acetonitrile (A), and 10% v/v tetrahydrofuran (T)—for three 50:50 binary mixtures (MA, MT, and TA), and for a

33.3 : 33.3 : 33.3 ternary mixture (MAT). An analyte's retention time for any other mobile phase combination is calculated using an empirical model built using this data and the equation

$$t_r = a_M X_M + a_T X_T + a_A X_A + a_{MA} X_M X_A + a_{MT} X_M X_T + a_{AT} X_A X_T + a_{MAT} X_M X_A X_T$$

where the X terms are the volume fractions of the three pure mobile phases and the a terms are the model's coefficients as determined from the standard runs.

The isocratic data in Investigation 9 is drawn from the paper “Mobile Phase Effects in Reversed-Phase Chromatography: I. Concomitant Dependence of Retention on Column Temperature and Eluent Composition,” the full reference for which is *J. Chromatogr.* **1979**, 185, 99-109. The retention times for gradient elutions were simulated using the method outlined in the supplementary materials to the paper “An Advanced, Interactive, High-Performance Liquid Chromatography Simulator and Instructor Resources,” the full reference to which is Boswell, P. G.; Stoll, D. R.; Carr, P. W.; Nagel, M. L.; Vitha, M. F; Mabbott, G. A. *J. Chem. Educ.* **2013**, 90, 198-202.

Investigation 1

This investigation presents students with a simple chromatogram and guides them in developing a common vocabulary for discussing chromatography and in developing an understanding of why the separation of a mixture is possible.

Question 1. Clicking and dragging over a peak returns the following results:

parameter	peak 1	peak 2	peak 3	peak 4
retention time (sec)	100	200	400	1000
peak height (a.u.)	1	2.82	1.15	0.42
peak width (sec)	NA	56.05	139.29	380.49
peak area (a.u. * sec)	1	100	100	100

Note: the first peak is too narrow to return a value for the peak width.

Question 2. Students should recognize that the first of the four peaks is very different from the other three peaks, particularly with respect to its peak width—to narrow to measure—and its peak area. They also should recognize that its retention time of 100 sec is equal to the column's length (100 mm) divided by the mobile phase's flow rate (1 mm/sec), suggesting that whatever contributes to the peak never interacts with the stationary phase; although the phrase “non-retained solute” may not occur to them, they should recognize what this term means when it appears in a later investigation and connect it back to their observation here.

Question 3. Although student's will not yet have command of specific vocabulary, such as partition coefficients and retention factors, they should recognize that an analyte can only progress through the column when it is in the mobile phase, that when it interacts with the stationary phase it is not moving with the mobile phase, and that it is reasonable to assume that different analytes interact with the stationary phase in different ways such that they need different amounts of time to pass through the column.

Question 4. Given that the peaks for the three analytes have identical areas, students should recognize that all three analytes have the same concentration.

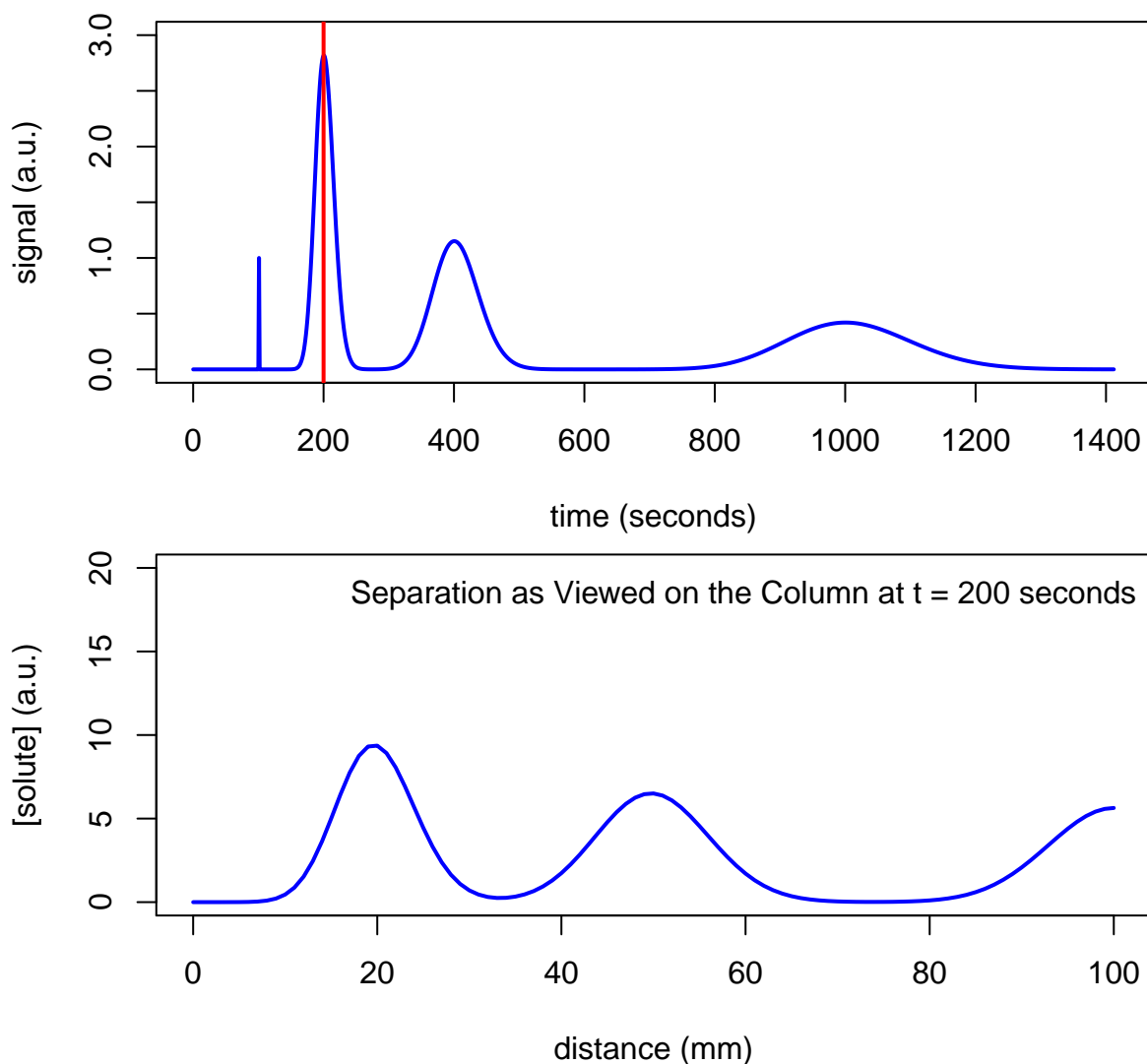
Question 5. Although they will not yet have command of the specific terms—such as longitudinal diffusion—to describe this, students should not be surprised that the longer it takes an analyte to pass through the column the more it tends to spread out in space. Given equal areas for the peaks, students should recognize that an increase in a peak's breadth requires a decrease in its height.

Investigation 2

The purpose of this investigation is to have students consider how a separation develops as the sample passes through the column by using a slider to view simultaneously the separation on the column and at the detector as a function of retention time. Using the available radio buttons, students can choose to examine the analytes individually or collectively, and to view the distribution of the analytes between the stationary phase and the mobile phase.

Question 1. Students should recognize that the analytes appear in reverse order because the analyte with shortest retention time in the detector view is the analyte that first reaches the end of the column in the column view.

Question 2. Students should note that the analyte's peak becomes broader in width and shorter in height as it passes through a column, and recognize that this happens for the same reasons they outlined in Question 5 of Investigation 1. Students also will see, more directly, how the analyte's elution from the column generates a signal at the detector; see, for example, these two views at 200 sec.



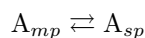
Question 3. Students should note that the individual concentration profiles are symmetrical in shape (although not necessarily equal in height), which should suggest to them that the analyte's movement between the two phases is an equilibrium process. If students have a hard time recognizing this or convincing themselves of this, they should resolve this as an issue when they work on Question 1 of Investigation 3.

Question 4. Students should see that the difference between the analytes in terms of their respective concentrations in the stationary phase and the mobile phase is consistent with their retention times; in particular, analyte A, which is equally soluble in the two phases elutes first, and analyte C, which is much more soluble in the stationary phase than in the mobile phase, elutes last.

Investigation 3

Following on Investigation 2, the purpose of this investigation is to make more concrete for students the analyte's equilibrium between the mobile phase and the stationary phase, and to add the retention factor to their list of an analyte's characteristic chromatographic properties.

Question 1. Although their choice of symbols might differ from those given here, students should be able to see that the equilibrium reaction is



where mp signifies the mobile phase and sp signifies the stationary phase. The equilibrium partition coefficient, K_A , is

$$K_A = \frac{[A]_{sp}}{[A]_{mp}}$$

Setting the slider to a distance of 75 mm shows that

$$[A]_{sp} = [A]_{mp} = 3.209$$

which makes $K_A = 1$; concentrations at other positions along the column give the same value for K_A . For the other two analytes, students should find that $K_B = 3$ and that $K_C = 9$

Question 2. Beginning with the equation for f_{mp} , students should recognize that dividing the numerator and the denominator by the analyte's formula weight, FW , converts f_{mp} from a mass ratio to a mole ratio

$$f_{mp} = \frac{(\text{mol A})_{mp}}{(\text{mol A})_{mp} + (\text{mol A})_{sp}}$$

which they can then rewrite as

$$f_{mp} = \frac{[A]_{mp} V_{mp}}{[A]_{mp} V_{mp} + [A]_{sp} V_{sp}}$$

where V_{mp} and V_{sp} are, respectively, the volumes of the mobile phase and the stationary phase. Factoring out $[A]_{mp} V_{mp}$ from the numerator and denominator

$$f_{mp} = \frac{1}{1 + \frac{[A]_{sp} V_{sp}}{[A]_{mp} V_{mp}}}$$

substituting in K_A for the ratio $[A]_{sp}/[A]_{mp}$, and simplifying leaves them with the desired final equation

$$f_{mp} = \frac{1}{1 + K_A \times \frac{V_{sp}}{V_{mp}}} = \frac{1}{1 + k_A}$$

Question 3. From Question 1 of Investigation 1, students know that $t_m = 100$ sec and that $t_{r,A} = 200$ sec; thus,

$$k_A = \frac{t_{r,A} - t_m}{t_m} = \frac{200 - 100}{100} = 1.00$$

The retention factors for the other two analytes are $k_B = 3.00$ and $k_C = 9.00$. Students should note that when $K = k$, as is the case here, that the volumes of the stationary phase and the volume of the mobile phase must be equal.

Investigation 4

This investigation uses the concept of a theoretical plate to begin focusing a student's attention on the processes that cause its elution band to broaden as it moves through the column. It is okay if a student does not appreciate fully the sources of band broadening at the end of this investigation as they will return to this topic in Investigation 7.

Question 1. The table below, which compares the characteristic values for analyte A for the two columns

parameter	$N = X$	$N = 2X$
retention time (sec)	200	200
peak height (a.u.)	2.82	3.98
peak width (sec)	56.05	39.915
peak area (a.u. * sec)	100	100

shows students that an increase in the number of theoretical plates decreases the width of an analyte's chromatographic peak and increases its height; it also shows that the number of theoretical plates does not affect the analyte's retention time or its peak area. Analytes B and C show similar results, although the measured retention time for analyte C is 1000 sec when $N = X$ and 999 sec when $N = 2X$.

Question 2. The simplest way for students to think about how the number of theoretical plates affects peak width is this: as it moves down the column, the number of theoretical plates over which the analyte is distributed is independent of the thickness of the plates; thus, if the analyte is distributed over y plates, those plates will encompass a smaller linear distance on the column and a smaller range of elution times at the detector when $N = 2X$ than when $N = X$. Because area is proportional to concentration, which is independent of N , a decrease in width must result in an increase in height if the area is to remain constant. Finally, an analyte's retention time is determined by the equilibrium constant—expressed as the partition coefficient, K , or the retention factor, k —that describes its relative solubility in the mobile phase and the stationary phase, which is not a function of N .

Question 3. Students may or may not recognize the contributions of longitudinal diffusion, multiple paths, and mass transfer—and almost certainly will not use these specific terms at this point. The intention here simply is to have them start thinking, in preparation for Investigation 7, about what happens to individual particles of the analyte as they pass through the column. Nevertheless, students should recognize that the broadening of a chromatographic peak as the analyte moves through the column is not unexpected.

Question 4. For the first chromatogram, we find that

$$N_C = 16 \left(\frac{1000 \text{ sec}}{380.49 \text{ sec}} \right)^2 = 110$$

and for the second chromatogram, we find that

$$N_C = 16 \left(\frac{999 \text{ sec}}{266.68 \text{ sec}} \right)^2 = 224$$

As expected, N_C essentially is $2\times$ greater for the second chromatogram. Note: the simulation itself uses $N = 100$ and $N = 200$; the difference between the actual and calculated values for N likely is the consequence of using such a small number of plates and the limitations this places on the binomial distribution to model a counter-current extraction and the Gaussian distribution assumed by the equation for N .

Investigation 5

With Investigation 5, students turn their attention to the resolution between two chromatographic peaks, with an initial focus on the factors that affect resolution and how we might calculate and report resolution; Investigation 6, which follows, takes an even closer look at resolution.

Question 1. Students should recognize that for a constant Δt_r , resolution is independent of the retention times of the individual analytes; thus, for example, when $w_A = 1.0$ and $w_B = 1.2$, Δt_r is 0.91 when the individual retention times are 8.0 min and 9.0 min, respectively, and when they are 7.0 min and 8.0 min. Increasing or decreasing the retention time of one analyte, however, increases or decreases the resolution.

Question 2. Students should recognize that smaller peak widths improve resolution. Students also should recognize that maintaining a constant total width, $w_A + w_B$, results in a constant resolution; thus, when $\Delta t_r = 1.0$, the resolution is 0.91 when $w_A = 1.0$ and $w_B = 1.2$, respectively, and when $w_A = 1.05$ and $w_B = 1.15$.

Question 3. With a little experimentation, students should recognize that resolution is independent of the concentrations of the two analytes.

Question 4. From the first two questions, students should recognize that resolution is directly proportional to Δt_r and that it is inversely proportional to the sum of w_A and w_B ; thus

$$R \propto \frac{\Delta t_r}{w_A + w_B}$$

By collecting data for several sets of conditions, students should be able to propose one of these final equations

$$R = \frac{2 \times \Delta t_r}{w_A + w_B} = \frac{2 \times (t_{r,B} - t_{r,A})}{w_A + w_B}$$

Question 5. Answers here will vary, but a resolution of 1.5 generally is accepted as baseline resolution. Students should recognize that any combination of retention times and peak widths that yields $R \geq 1.5$ will result in baseline resolution and that concentration is not important.

Question 6. Students should recall that the retention factor, k , is a function of retention time, and that the retention factor is proportional to the partition coefficient, K ; thus, they should recognize that they can increase Δt_r by finding a mobile phase/stationary phase combination that affects k for one analyte more than it affects k for the other analyte.

Question 7. Students should recall that increasing the number of theoretical plates results in smaller peak widths; however, beyond suggesting that they can increase N by using a longer column, students may not yet have the specific language in hand to address this in terms of the factors that contribute to band broadening; see, for example, the earlier comments in the suggested response to Question 3 of Investigation 4. Students will return to this topic in Investigation 7.

Investigation 6

Continuing from the previous investigation, which introduced resolution, students now explore an alternative equation for resolution, which defines it in terms of the number of theoretical plates, N , the relative retention factors of the two analytes, α , and the retention factor for the later eluting analyte, k_B . The investigations that follow this one provide a closer look at how these three factors are used to improve resolution.

Questions 1. Students should see that the equation's third term, $\frac{k_B}{k_B+1}$, represents the column's general retention of analytes because it accounts for the time needed to elute the last of the two analytes. Students also should recognize that the equation's second term, $\frac{\alpha-1}{\alpha}$, must represent the relative selectivity for the two analytes given that $\alpha = \frac{k_B}{k_A}$. Finally, students are left with the first term, $\frac{\sqrt{N}}{4}$, as being associated with the efficiency with which the analyte moves within the mobile phase and between the mobile phase and the stationary phase.

Question 2. Beginning with the equation

$$R = \frac{2 \times (t_{r,B} - t_{r,A})}{w_A + w_B}$$

students should recognize easily that

$$w_B + w_A = 2 \times w_B$$

which, on substituting back, gives

$$R = \frac{t_{r,B} - t_{r,A}}{w_B}$$

Next, from Investigation 3, students know that $k = \frac{(t_r - t_m)}{t_m}$, which they can solve for $t_{r,B}$ and for $t_{r,A}$; substituting back and simplifying leaves them with

$$R = \frac{k_B t_m - k_A t_m}{w_B}$$

From Investigation 4, students know that $N = 16 \left(\frac{t_{r,B}}{w_B} \right)^2$; solving for w_B , substituting back, and simplifying gives

$$R = \left(\frac{\sqrt{N}}{4} \right) \left(\frac{k_B t_m - k_A t_m}{t_{r,B}} \right)$$

Substituting $k_B t_m + t_m$ for $t_{r,B}$ and factoring out t_m leaves students with

$$R = \left(\frac{\sqrt{N}}{4} \right) \left(\frac{k_B - k_A}{k_B + 1} \right)$$

From the definition for α , students know that $k_A = \frac{k_B}{\alpha}$; substituting back, factoring out k_B from the numerator, and simplifying leaves

$$R = \left(\frac{\sqrt{N}}{4} \right) \left(\frac{k_B}{k_B + 1} \right) \left(1 - \frac{1}{\alpha} \right)$$

Recognizing that $1 - \frac{1}{\alpha}$ is equivalent to $\frac{\alpha-1}{\alpha}$, leaves students with the final equation

$$R = \left(\frac{\sqrt{N}}{4} \right) \left(\frac{k_B}{k_B + 1} \right) \left(\frac{\alpha - 1}{\alpha} \right)$$

Because the two poorly resolved analyte's elute at nearly the same time it is reasonable to assume that the factors affecting their peak width are similar; thus, the assumption that $w_A = w_B$ is reasonable.

Question 3. For k_B , students should recognize that changing its value from 3.0 (the default value)

$$\frac{k_B}{k_B + 1} = \frac{3}{3 + 1} = 0.75$$

to 40.0 (the maximum value)

$$\frac{k_B}{k_B + 1} = \frac{40}{40 + 1} = 0.98$$

will improve the resolution by a factor of about 30%, which likely is not sufficient to resolve the two analytes.

Using the same approach, for α , students should predict that changing its value from 1.1 (the default value) to 2.0 (the maximum value) will improve resolution by a factor of roughly 500%, which likely is sufficient to resolve the two analytes. For N , students should predict that changing its value from 1000 (the default value) to 6000 (the maximum value) by about 250%, which likely is nearly fully resolved.

By adjusting the sliders, students can confirm all three predictions.

Question 4. Answers here likely will vary, but students should recognize that to increase k_B without increasing α requires that all analytes remain on the column longer, which they can accomplish by decreasing the flow rate or choosing a stationary phase/mobile phase combination that affects all analytes equally. Students also should recognize that increasing α while maintaining a constant k_B will require finding a stationary phase/mobile phase combination that affects the retention time of just the earlier eluting analyte. Finally, students should recognize that using a longer column will increase the value of N .

Question 5. Answers here likely will vary, but students should recognize from a plot of resolution as a function of N that little improvement in resolution is possible for values of $k_B > 10$; students should recognize, as well, that an increase in k_B comes at the expense of a longer analysis time.

For α , students should recognize that increasing its value for one pair of analytes (say, analyte A and analyte B) comes at the expense of decreasing its value for another pair of analytes (say, analyte A and analyte C).

For N , students should recognize that improving resolution by using a longer column comes at the expense of a longer analysis time.

Investigation 7

In this investigation—the first of three that consider ways to improve resolution—students consider how to improve resolution by improving column efficiency.

Question 1. Students should recognize that increasing the length of the column increases the time it takes to elute the analytes.

Question 2. Students should recognize that the existence of different pathways through the column is independent of time and, therefore, independent of the mobile phase's flow rate; thus, they should find it easy to assign this as the A term in the van Deemter equation. The remaining two terms likely will require more thought on the part of students, but with some nudging where needed, students should recognize that

diffusion within the mobile phase (longitudinal diffusion) increases the longer the analyte remains in the column; thus, this factor is less important at high flow rates and is, therefore, inversely proportional to μ and is represented as the B term in the van Deemter equation. This leaves, of course, C as the term accounting for the kinetics of the analyte's movement between the stationary phase and the mobile phase; with some hints, as needed, students should be able to explain that a slow mobile phase flow rate ensures that the partitioning remains at equilibrium throughout the analyte's time in the column.

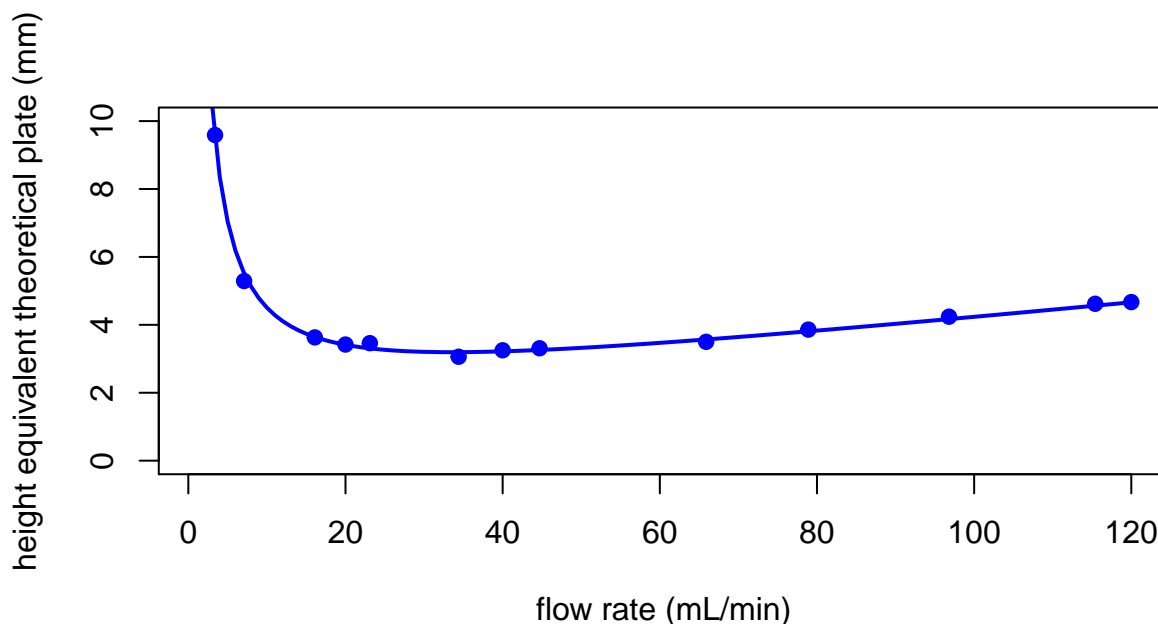
Question 3. For the multiple pathways term (A), students should have little difficulty recognizing that it will depend on particle size as larger particles will allow for a greater variety of path lengths. For the longitudinal diffusion term (B), students should recognize that it will depend on the analyte's diffusion coefficient in the mobile phase. Finally, for the mass transfer term (C), students should recognize that it will depend on the analyte's diffusion coefficients in the mobile phase and the stationary phase; they also should recognize, perhaps with some gentle guidance, that it depends on the thickness of the stationary phase, particle size, and the diameter of the column.

Question 4. By taking the derivative of the van Deemter equation with respect to the flow rate and setting it to zero

$$\frac{\delta H}{\delta u} = -\frac{B}{u^2} + C = 0$$

students can rearrange and solve for u to obtain $u = \sqrt{\frac{B}{C}}$. Substituting this back into the van Deemter equation and simplifying gives $H = A + 2\sqrt{BC}$.

Question 5. With some playing around, students should be able to find a good fit of the van Deemter equation to the data. Shown here is the result using $A = 1.61$ mm, $B = 26.6$ mm • mL/min, and $C = 0.0236$ mm • min/mL, with a total residual error of 0.12.



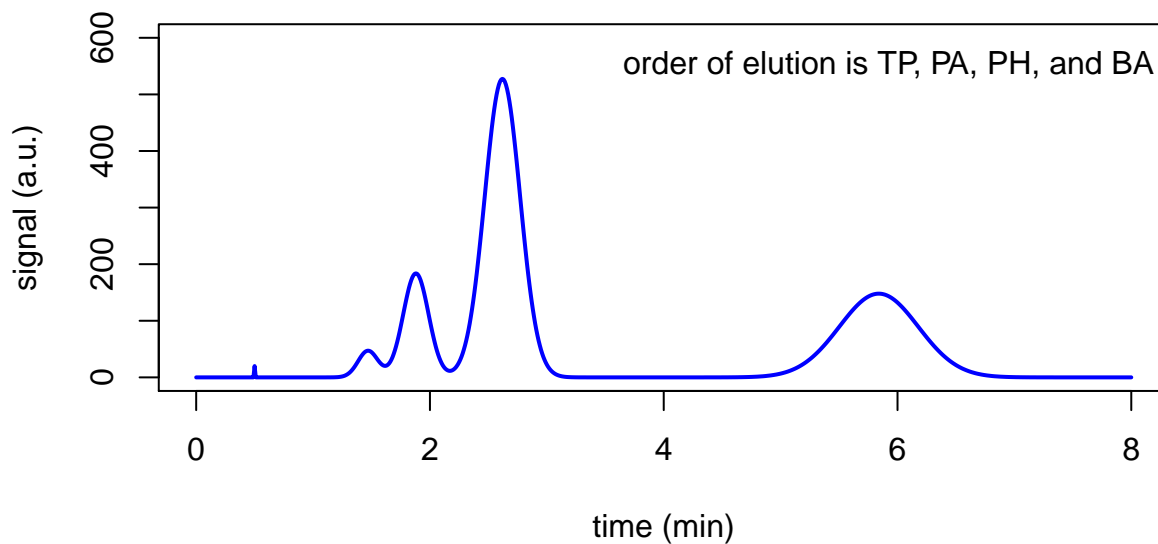
Investigation 8

Following on a consideration of improving resolution through column efficiency, in this investigation students explore how to improve resolution by controlling selectivity.

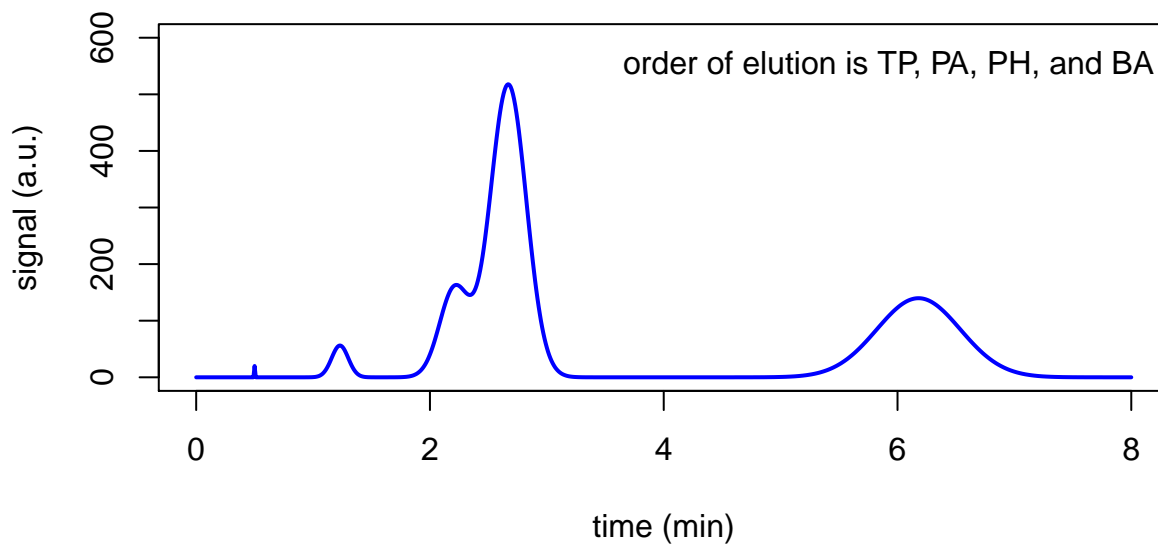
Question 1. Because the analytes will differ in their relative solubilities in the three different mobile phases, students can reasonably expect that changing the composition of the mobile phase will affect each analyte's retention times.

Question 2. The three chromatograms using the pure mobile phases are shown here; note that the order of elution is the same for all three chromatograms.

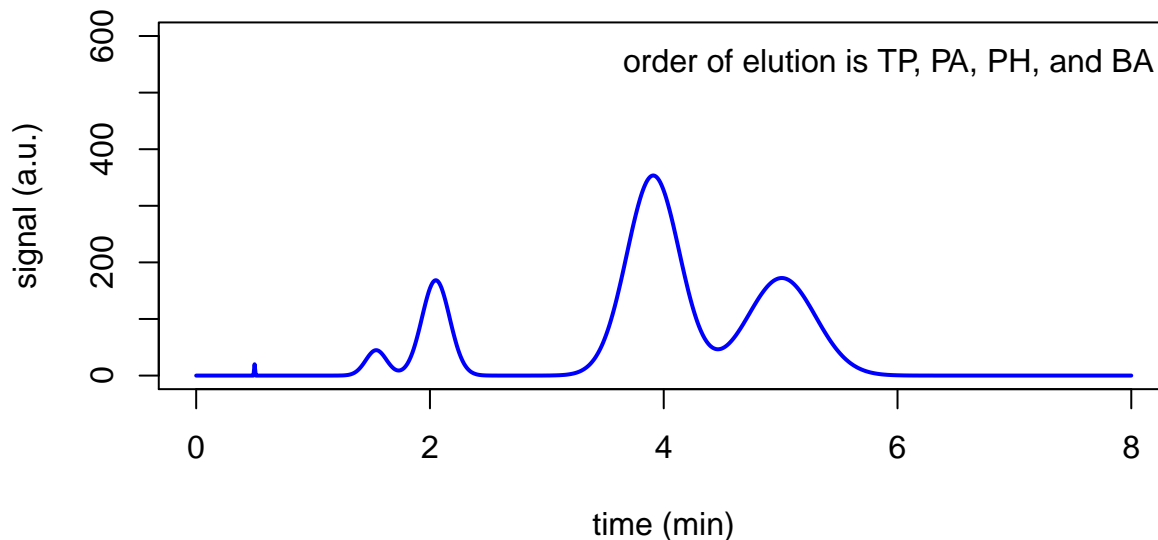
separation using 20% v/v methanol



separation using 16% v/v acetonitrile

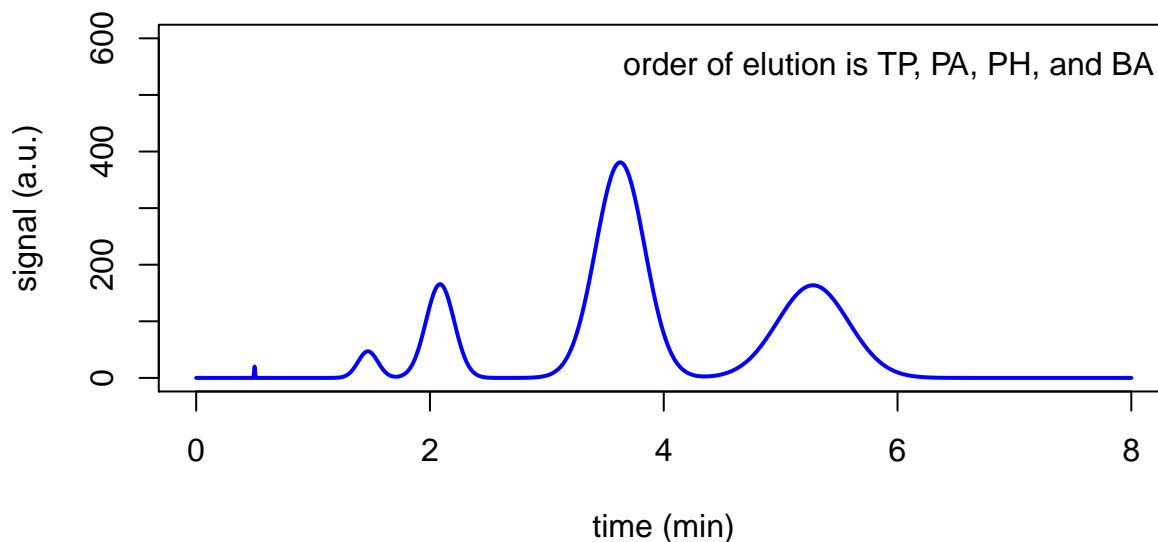


separation using 10% v/v tetrahydrofuran

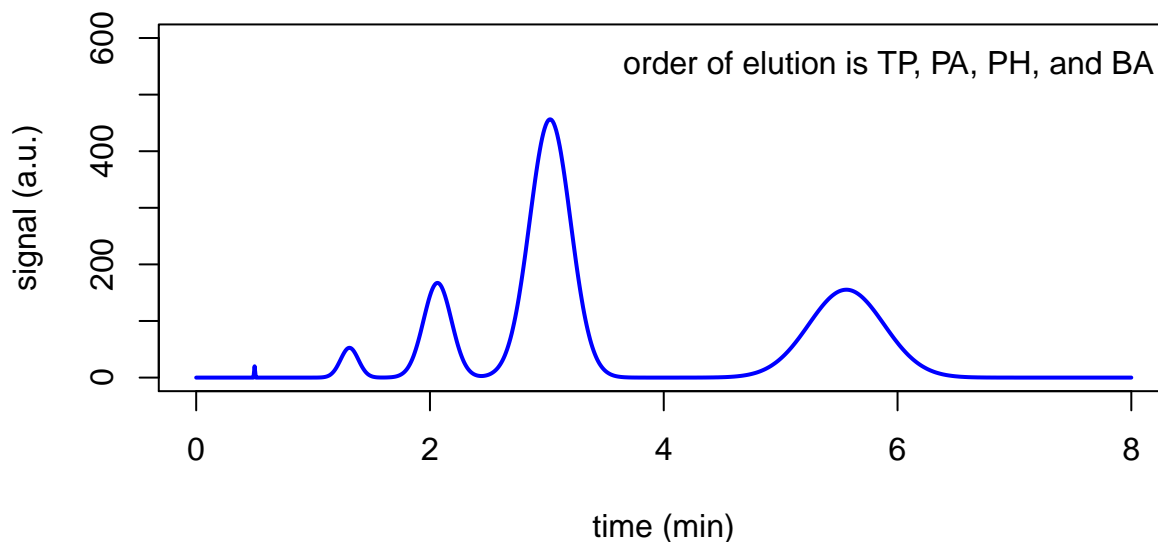


Students should recognize that the retention time for p-hydroxybenzoic acid shows the greatest sensitivity to the mobile phase's composition, particularly that for THF relative to MeOH and ACN. Benzoic acid shows the second greatest sensitivity to the mobile phase's composition. Terephthalic acid and p-aminobenzoic acid are much less sensitive to the mobile phase's composition.

Question 3. There are many possible mobile phase combinations that will work, one of which, as shown here, is 33% THF and 67% ACN.



In general, binary mixtures of THF and ACN from about 60% ACN to 85% ACN will work; near the middle of this range, some of the THF can be replaced with methanol. A separation using a mobile phase of 10% MeOH, 75% ACN, and 15% THF is shown here.

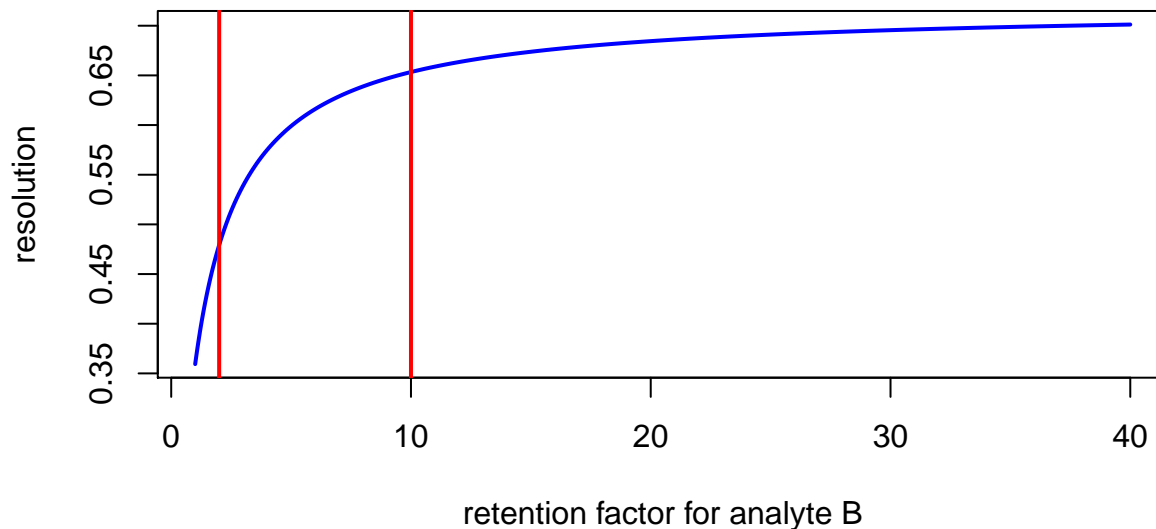


Question 4. Because we may not be able to control the composition of the mobile phase precisely, we want to use conditions where small variations in the composition of the mobile phase will not affect the quality of the separation.

Investigation 9

In this last investigation, students explore the control of overall retention as a means for improving resolution.

Question 1. The plot below shows the effect of k_B on resolution when α is 1.1 and when N is 1000 with the red lines set at $k_B = 2$ and at $k_B = 10$. Students should have no difficulty recognizing that the rate of improvement in resolution decreases quickly for $k_B > 10$.



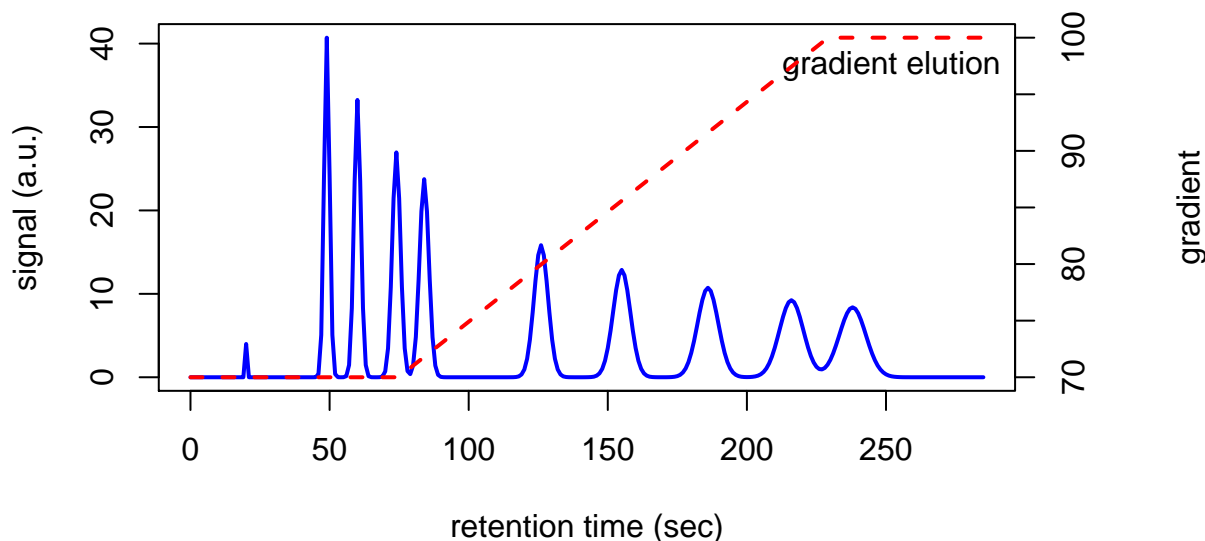
Question 2. Students should recall that $t_r = kt_m + t_m$, which gives a retention time of 60 sec for an analyte with a k_B of 2 and a retention time of 220 sec for an analyte with a k_B of 10.

When using 100% ACN, the last analyte in the mixture elutes with a retention time of approximately 45 sec and when using 70% ACN, the last analyte elutes with a retention time of approximately 430 sec.

Students should recognize that the general elution problem is this: conditions that favor a short analysis time, such as using 70% ACN, may resolve late eluting analytes but at the cost of a failure to resolve early eluting

solvents. Choosing conditions that will resolve the early eluting solvents comes at the cost of a substantial increase in analysis time.

Question 3. A baseline separation with the last analyte eluting within 220 sec is not possible when a linear gradient, although a baseline separation is possible with the last analyte eluting at 240 sec when applying the following gradient: 70% ACN from $t = 0$ to $t = 75$ sec; linear ramp to 100% ACN until $t = 230$ sec; hold at 100% until analysis complete.



A step gradient from 70% to 85% ACN at $t = 80$ sec, followed by a second step gradient from 85% ACN to 100% ACN at 160 sec provides baseline resolution of all analytes within a total run time of under 200 sec.

