ABCS OF EDUCATION AND PROFESSIONAL DEVELOPMENT IN ANALYTICAL SCIENCE

Reduce cognitive overload with systemic cause-effect relation maps to teach and learn chromatography

Jérôme Randon[1] · Vincent Dugas[1] · Claire Demesmay[1]

Introduction

For complex sample analysis, separation of analytes from interfering compounds is mandatory for accurate quantification; this is the main goal of chromatography. Nowadays, reverse-phase liquid chromatography (RPLC) is very popular, linked on one hand to an increasing demand for analysis of samples whose constituents are not volatile (this is very often the case for compounds dedicated to pharmaceutical application) and on the other hand to the great robustness of the method and the wide range of solutes that can be separated compared to normal-phase chromatography, hydrophilic interaction chromatography or ion chromatography. Thus, HPLC/UHPLC columns with silica particles grafted with alkyl chain groups (often octadecyl, C_{18}) are the most widely used to date. Then, with an HPLC chromatograph and such C_{18} column, chemists have to design separation methods.

To select the relevant values of HPLC/UPLC experimental parameters, students (bachelor's or master's degree students) have to rely on fundamental concepts of chromatography, which are widely introduced in numerous textbooks. However, although concepts associated with chromatography are all introduced in textbooks, and all the elements of the chromatographic system are very well described, the methodology to design a chromatographic method is often missing.

Figure 1 shows the elements of the chromatographic system using concept map representation [1] with a focus

question dedicated to RPLC optimisation. In such map, links between operating parameters (blue: column characteristics; green: mobile phase) and quality criteria (orange: resolution and time for analysis) are completely missing. Without explicit cause-and-effect links between operating parameters and quality criteria, we observed a low success rate in method development. It also leads to significant loss of time, poor method robustness and consequently increased difficulty in transferring the method from an R&D lab to a QC lab.

If the knowledge of cause-and-effect links is mandatory to efficiently optimise analytical RPLC separations, the way to build these links is a very difficult task to perform during the teaching/learning process, even if mathematical relations are provided. Several suggestions for graphical tools have been proposed in the literature to facilitate learning, and a summary of these tools can be found in an article recently published in the *Journal of Chemical Education* [2]. From 2009, within this broad scope of graphical approaches, we used a pedagogical strategy we called "Systemic Cause-Effect Relation Map" (SCERM) in order to promote the development of skills related to RPLC method optimization and troubleshooting. In such kind of map or drawing, as for the concept map, concepts related to the studied process (RPLC in our case) are written on a white board: these concepts can be experimental parameters (such as mobile phase flow rate or column length), quality criteria (such as chromatographic resolution or analysis time) but also the so-called intermediate characteristics which cannot be directly controlled (such as mobile phase velocity or column efficiency). In SCERM, arrows between concepts are used to indicate an implication link between these concepts. The meaning of the arrow can be described as follows: if we change the value of a parameter located at the initial end of an arrow (if we change flow rate, for example), the value of the parameter located at the terminal end of the arrow (mobile phase velocity) will be

✉ Jérôme Randon
  randon@univ-lyon1.fr

[1] Université de Lyon, CNRS, Université Claude Bernard Lyon 1, Institut des Sciences Analytiques, UMR 5280, 5 rue de la Doua, 69100 Villeurbanne, France
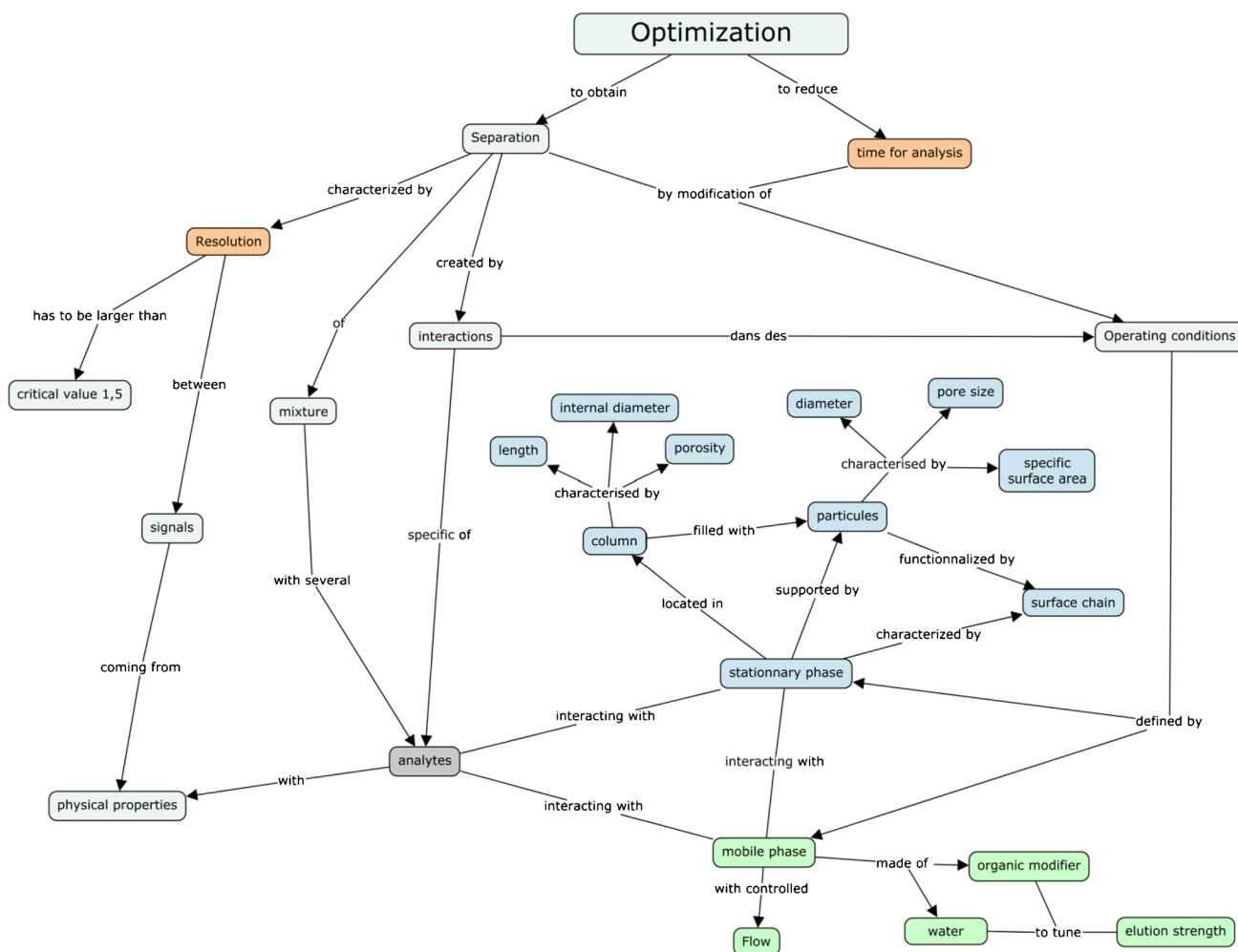
Fig. 1 Concept map dedicated to optimisation of reverse-phase liquid chromatography system

modified. Very often, such link represented by an arrow can also be described using a mathematical relation between parameters, and this relation will be associated with the arrow. Conversely, a mathematical relation can be represented by an arrow (or more often by a set of arrows). During the production of SCERM, students have to identify the causal links with physical meaning in order to draw arrows. However, to produce such map from their initial knowledge or set of documents, students are facing another issue related to the apparent multiplicity of relations/equations introduced through analytical lectures and textbooks: which relation is relevant and how to draw the arrows?

Relationships and cause and effect

To be in-line with the SCERM process when teaching chromatography (with bachelor's degree students, master's degree students, or chemists during lifelong learning), and so to reduce student cognitive overload [3], we have decided that any mathematical relation between quantities has to be exclusively introduced according to a unique convention associated with the sign "=". Whatever the mathematical relation introduced, the parameter situated to the left side of the equal sign "=" has to be the consequence of the parameters situated on the right side of the sign (this type of writing is also used in computer science when assigning the result of calculation to a variable).

For an advanced scientist, writing a mathematical relation does not involve a cause-and-effect link. It is just a relation between several parameters, and this relation can be used to find one unknown parameter when the other ones are known. An advanced scientist has a perfect understanding of implication links, but a student does not, and needs to learn them. The student may be completely confused and unable to grasp the underlying physical models unless we make them explicit, which is the purpose of this writing convention and SCERM representation.

To introduce the interest of such convention, let us consider three characteristic variables such as column length (L), mobile phase velocity (u) and hold-up time (t_m), and three different relations (a), (b) and (c) generated by rotating the variables:

$$u = \frac{L}{t_m} \quad (a) \quad L = u.t_m \quad (b) \quad t_m = \frac{L}{u} \quad (c)$$

Students are invited to find the relationship that provides a physical link (analogy with a car moving from one city to another can be useful with velocity, distance and time). Usually, relation (b) is immediately dropped out (the column is inside the oven and its length does not depend on mobile phase velocity and hold-up time!), but a discussion always follows about the other two relations. It is quite common that equation (a) can be considered the most meaningful by the students, although relation (c) is the one that should be considered as the operative one according to our convention: the hold-up time is the consequence of column length and mobile phase velocity (the time to go from one city to another depends on the distance and speed of the car, even if we can calculate the mean velocity from the distance and time spent to go from one city to another). A student's misrepresentation of relation (a) arises from the fact that mobile phase velocity can be determined by injecting an unretained compound which will exit the column at a measurable hold-up time, and using the column length in relation (a), the velocity phase velocity can then be obtained. In our pedagogical approach, equation (c), reported as 3 in Fig. 3, will be the only one written on our documents.

In a slightly different context, students also encounter difficulties associated with the multiplicity of relationships in which they have seen a parameter being displayed. If we consider, for example, the case of mobile phase velocity u, this parameter appears, among others, in relations (d) to (f):

$$u = \frac{L}{t_m} \quad (d) \quad u = B_0 \frac{\Delta P}{\eta.L} \quad (e) \quad u = \frac{F}{\varepsilon \frac{\pi.d_i^2}{4}} \quad (f)$$

where B_0 is column permeability; ΔP, pressure drop; η, viscosity; F, mobile phase flow rate; and d_i, column internal diameter. It takes some time for the students to move to a representation where the mobile phase velocity is a consequence of the mobile phase flow rate in relation to column section. Thus, relations (d) and (e) which are not associated with a causal link will never be used in our documents.

Let us now take another situation, efficiency (N), retention time (t_r) and standard deviation of the chromatographic peak (σ), often introduced using relationship (g):

$$N = \left(\frac{t_r}{\sigma}\right)^2 \quad (g)$$

This relationship is used to estimate the efficiency value from an experimental chromatogram. Very often, students think that the efficiency N is the consequence of the retention time and the standard deviation of the chromatographic peak, without realising that he/she is implicitly applying the convention described above (the quantity to the left of the "=" sign is a consequence of the parameters located on the right of this sign).

In many textbooks, we can also find the relation $H = L/N$ where L is the length of the column and H is the height equivalent to a theoretical plate. This relation could lead students to believe that H depends on L and N. But students are also facing the well-known Van Deemter's relation $H = A + B/u + C.u$ where u is the velocity of the mobile phase (A, B and C are respectively related to flow path anisotropy, longitudinal diffusion and mass transfer). Here again, there is something confusing according to the above convention: H cannot simultaneously depend on both (A, B, C, u) and (L, N). The H value depends on the quantities (A, B, C, u according to Van Deemter's model), and then efficiency N is a consequence of H and L (in our convention, $N = L/H$). So finally, relation (g) has to be written as (h).

$$\sigma = \frac{t_r}{\sqrt{N}} \quad (h)$$

The standard deviation of the chromatographic peak is a consequence of retention time and column efficiency! Why such a reversal in the student's mind? Simply because we can get the efficiency value from an experimentally obtained chromatographic peak according to relation (g).

We believe it is important to apply a consistent type of writing in all the documents we use with students. Later, students will be free to do any manipulation of the relationships, as they wish. The effort that we put into maintaining a consistent writing style during our lectures and teaching lab activities can make it easier for students to understand the physical system.

From equations to cause-effect map

After several lectures dedicated to chromatography at the bachelor or master level, we asked students to identify the key concepts related to RPLC in isocratic mode, and asked them to start to draw a map by locating the concepts in an A4 page (very often, we gave students a blank map without arrows such as Fig. 2 in order to help them to organise the different concepts).

In a second step, we asked each group of 3–4 students to draw the arrows that seem relevant to them with the associated mathematical relationships. But only one copy
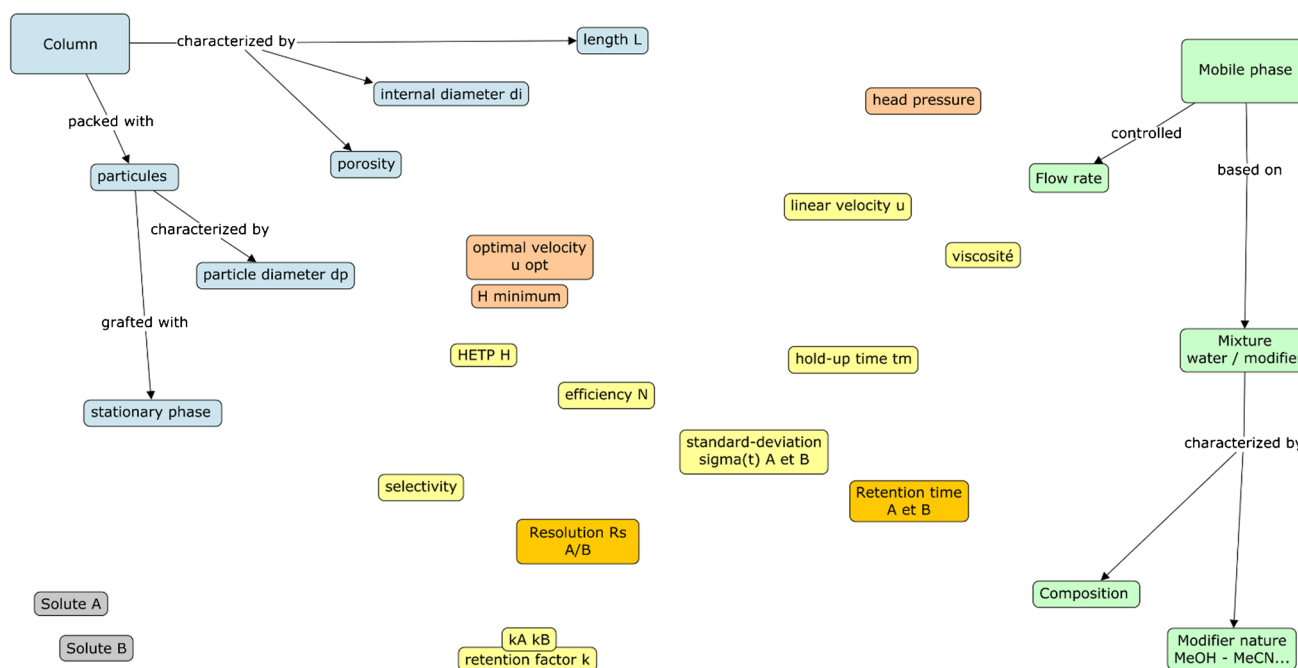
Fig. 2 Example of selected key concepts organised to facilitate expression of cause-effect relationships

of Fig. 2 is given to the group, and so students have to decide collectively which arrows will be represented. The meaning of the relationships they have chosen is then deeply discussed within the group and with the teacher if necessary.

In Fig. 3, we have summarised a restricted set of relationships associated with RPLC in isocratic mode, while in Fig. 3, we make the cause-effect links explicit by using arrows between the quantities. Of course, such a map is based on a limited collection of concepts and is not exhaustive: the parameters have been selected according to the specific goal we want to discuss with students. On the other hand, this map is only a representation of knowledge at a specific time during the learning process and will be further modified.

As an example of discussion, in Fig. 4, relation 10, the retention time appears as a consequence of hold-up time and retention factor. This may seem confusing to the student who has seen the relationship $k = t_r/t_m - 1$, sometimes taking this equation as the "definition" of retention factor, and therefore preferring to draw an arrow from t_r and t_m to k. Of course, solute retention factor can be obtained from the experimental chromatogram, but an increase in retention factor is not the consequence of an increase in retention time! It is because the retention factor has increased (e.g. the water content of the mobile phase has increased, favouring the partition equilibrium towards the stationary phase) that the retention time has increased.

Fundamental laws and models

In the set of relations reported in Figs. 3 and 4, some relations are fundamental relations of physics that are not open to discussion. For example, in relation (3), the time to cover a distance is a consequence of the distance to be covered (the length of the column) and the velocity.

On the other hand, relations (4) and (9) are relations associated with models. For example, in relation (4), we decided not to make explicit the definition of the term H (called the equivalent height on a theoretical plate, name coming from distillation theory): $H = \sigma_L^2/L$, which is difficult to manipulate for a first level of representation, and we have preferred to represent the quantity H as dependent on several phenomena: flow anisotropy (A), longitudinal diffusion (B) and mass transfer (C), in relation to the velocity of the mobile phase u. This remains a modelling choice, and therefore, other relationships known as Van Deemter's, Knox's, Golay's, Giddings's, etc., may be selected. Overall, it can be seen that the diameter of the particles makes a significant contribution to the value of H and that, subject to some approximations, it is possible to obtain a reasonable model for the quantities A, B and C and to obtain the relationships (6) and (5) which characterise the optimum speed associated with the minimum of the H term (and therefore with the maximum of the chromatographic efficiency according to relation 7).

Similarly, the definition of the retention factor $k = n_{stat}/n_{mob}$ (ratio of the amounts in each of the two phases) is not used in Fig. 2. Equation (4) corresponds to the so-called

Fig. 3 Selected set of relations displayed according to cause-effect convention

1 $u = \dfrac{D}{\pi \cdot d_i^2 \cdot \varepsilon}{4}$

2 $\Delta P = Cte. \dfrac{\eta \cdot L \cdot u}{d_p^2}$

3 $t_m = \dfrac{L}{u}$

4 $H = A + \dfrac{B}{u} + C \cdot u$
with $C = f(d_p^2)$

5 $u_{opt} \approx 5. \dfrac{D_m}{d_p}$
$D_m \approx 10.10^{-6} cm^2/s$

6 $H_{min} \approx 3. d_p$

7 $N = \dfrac{L}{H}$

8 $\sigma_{(t),i} = \dfrac{t_{r,i}}{\sqrt{N}}$

9$_{A\,et\,B}$ $\ln(k_i) = \ln(k_{w,i}) - S_i \cdot \varphi_{modif}$

10$_{A\,et\,B}$ $t_{r,i} = t_m(1 + k_i)$

11 $\alpha = \dfrac{k_2}{k_1}$

12 $Rs = \dfrac{t_{r,2} - t_{r,1}}{2.(\sigma_1 + \sigma_2)}$ $Rs = \dfrac{\sqrt{N}}{4} \dfrac{\alpha - 1}{\alpha} \dfrac{k_2}{1 + k_2}$

linear solvent-strength model (developed in detail in Snyder and Dolan [4]), which is very useful for the chromatographer and which assumes that the logarithm of the retention factor depends linearly on the content of organic modifier (strong solvent: methanol, acetonitrile, etc.). Although this model is broadly applicable to a large number of compounds

(a slight deviation is, however, often observed for acetonitrile), it is currently impossible to know a priori values of k_w and S for any solute for a specific organic modifier. In order to optimise the chromatographic separation, it is therefore necessary to experimentally determine these two quantities. By carrying out at least two experiments at two different
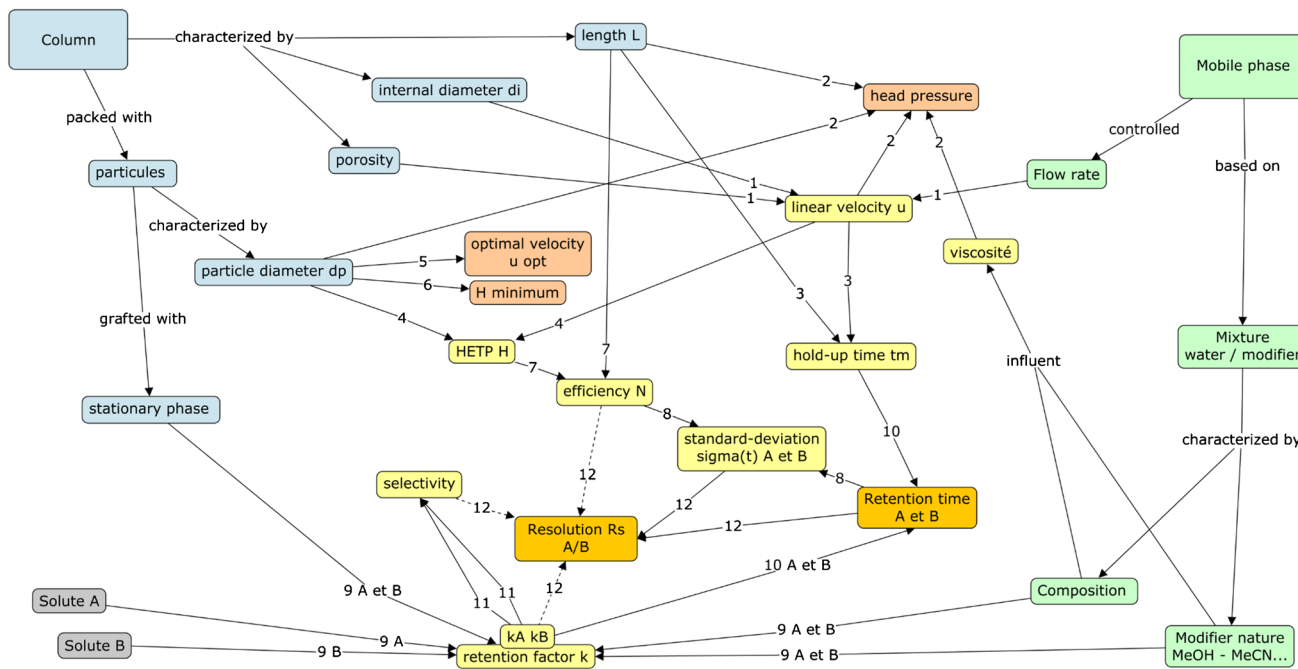
Fig. 4 Cause-and-effect relation map for isocratic separation in reverse-phase chromatography

contents of organic modifier, it is possible to obtain the retention factor of the compound for each organic content from the retention times observed by solving Eq. (10) and then to find the values of k_w and S, characteristics of the compound with the selected organic modifier by means of linear regression (example in Fig. 3).

A series of arrows around the quantity "resolution" in Fig. 2 might also seem confusing at first, but we expected to get as close as possible to the definition of resolution for Gaussian chromatographic peaks (left part of relation 12, solid arrows), but also to present the so-called Purnell relation (right part of relation 12, dotted arrows), which is a consequence of the definition of resolution and of relations 10 and 11. This relationship is very useful because it makes it easier to show the respective effects of varying the retention factor and the selectivity on the resolution.

As previously mentioned, a fully comprehensive map cannot exist. Each map is dedicated to a specific topic according to a focus question [1]. New maps will be drawn according to the targeted teaching goal, and we have also developed specific maps to address band broadening (from column and external effects), gas chromatography, capillary electrophoresis…

How can we take this map a step further?

Figures 3 and 4 can be used to design a rational optimization process of isocratic reverse-phase separation which can be implemented in a spreadsheet build by students themselves (a full instruction set is provided in the supplementary document). With a selected column/solvent pair, optimising the isocratic method requires the knowledge of retention models associated with the compounds to be separated. These models can only be experimentally determined, as we have discussed. After selecting a column, the chemist chooses a mobile phase velocity greater than the optimal velocity (estimated via Eq. 5) to save time without sacrificing too much efficiency. A first flow rate to be applied will then be calculated by solving Eq. 1 assuming a porosity of 65%. Injection of a non-retained compound, such as thiourea, which is very polar and therefore not retained on a C_{18} phase, and easily detected by UV, will provide access to the hold-up time of the system, and therefore, the true velocity can be determined by solving Eq. 3, and the true porosity of the column can be determined using Eq. 1.

It will then be possible to inject the sample at different organic content, identify the retention times associated with each compound—it is not necessary to know the compound, but rather to be able to recognise it on the different chromatograms, for example, through its spectrum—and determine k_w and S parameters of the retention model associated with each compound (Fig. 5).

This figure highlights that chromatographic retention (and so selectivity) is strongly dependent on methanol content, and thus, the methanol content must be chosen with great care. In the example discussed here, the separation of these three compounds will be very difficult around 60–65% methanol, and may even be impossible if working at 62% methanol when two of the compounds have identical retention factors (red and green curves intercept), and therefore have identical retention times (selectivity in this case is equal to 1), and regardless of the column efficiency, satisfactory resolution cannot be achieved (Rs = 0 if selectivity is 1, relation 12). The more compounds in the mixture, the greater the risk of co-elution, and this is how knowledge of retention models will guide the chemist in selecting appropriate experimental conditions.

The knowledge of retention models allows the chemist to estimate the minimum resolution that can be observed whatever the organic solvent content in the mobile phase. Indeed, for a selected composition, it is possible to calculate the value of the retention factor k for each compound, as well as the retention times and peak standard deviations with respect to the column and working conditions. Therefore, a graph—resolution versus solvent content—can be plotted (Fig. 6) to determine the solvent contents that enable the desired minimum resolution to be achieved.

Finally, a solvent content that leads to the shortest analysis time while maintaining satisfactory method robustness will be selected (a solvent content of 75% MeOH appears to be a relevant value in our example, as shown in Fig. 6). As retention models are not dependent on the geometry of the RPLC column or the mobile phase flow rate, it is also possible to test other experimental conditions and observe their consequences on the resulting chromatogram. For example, in Fig. 6, the mobile phase flow rate was changed to reach the optimal mobile phase velocity of the chromatographic column.

In some cases, we may not be able to separate the set of compounds using methanol. It will then be relevant to try another solvent such as acetonitrile, build the retention models that will be different due to the –C≡N group of the acetonitrile (compared to the –OH group of methanol) which generates other type of interactions between solutes, mobile and stationary phases—and here we see the interest of Snyder's classification [5]. With these new models, perhaps there will be a composition that allows for separation... or maybe not.

Simulations carried out by a tool such as the one presented in Fig. 6 (created by students according to instructions provided in the supplementary document) can also be used to "explore" the effects of modifications of one specific operating condition, one by one or simultaneously. However, one must be very careful and make students aware that we are, in this case, in the world of models that provide access to a simulation. Some experimental

Experimental column

Length (cm)	10	
Internal diameter (mm)	4,6	
Particule diameter (µm)	5	

Porosity ε 0,73

V_m (mL) 1,21

Flow rate (mL/min)	1,200	MeOH

F_{opt} (mL/min) 0,73

tm (min)	1,01
u (cm/s)	0,165
Estimated H (µm)	16,9

u_{opt} (cm/s) 0,1
H_{min} (µm) 15
N_{opt} 6667

Experimental data tr (min)

% MeOH	butylparaben	neburon	ethylbenzene
50	4,37	9,76	7,38
70	1,3	1,58	1,7

|	butylparaben	neburon	ethylbenzene
slope	-0,122	-0,137	-0,111
intercept	7,327	8,987	7,398
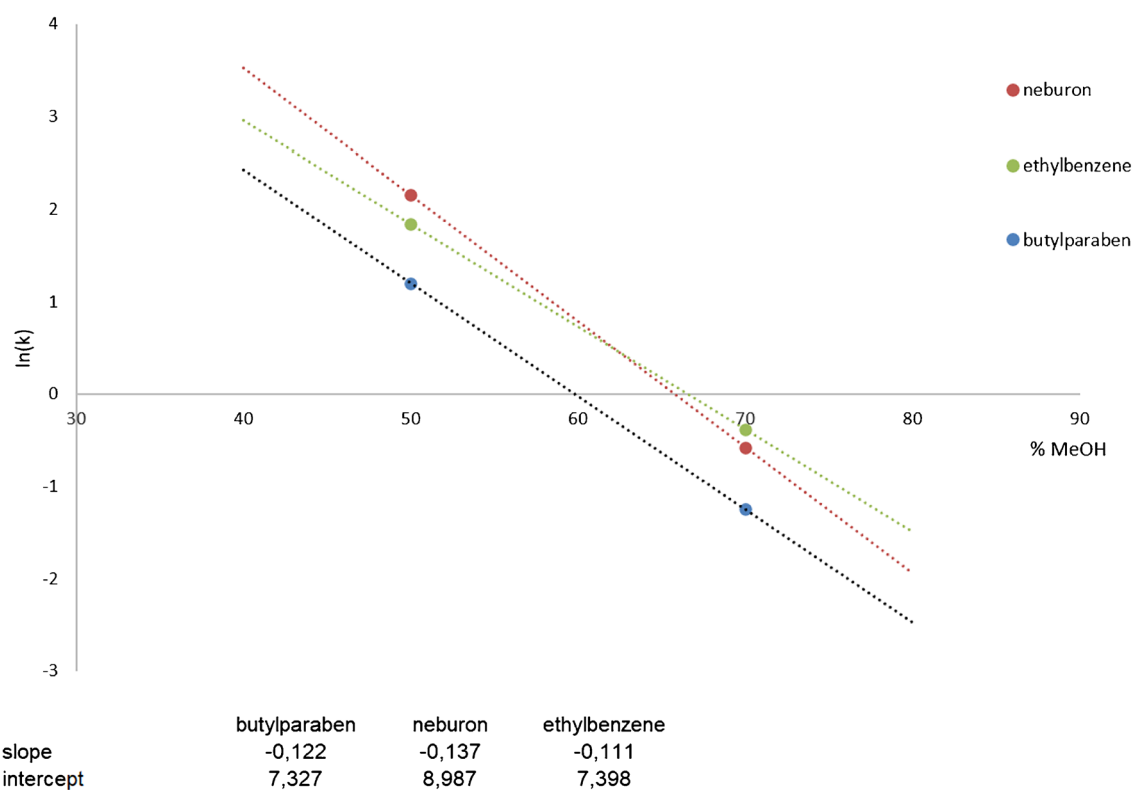
Fig. 5 Experimental retention model (logarithm of retention factor versus organic modifier content) obtained for ethylbenzene, butylparaben and neburon using C18 column with methanol as organic modifier (two experiments have been conducted using 70 and 50% of methanol in the mobile phase)

observations that can be made in the laboratory will not be in agreement with the simulation provided by the model—we previously mentioned the fact that the retention model with acetonitrile sometimes deviates from the linear model. In this case, it is not the experimental observations that should be rejected, but rather the model that is not comprehensive enough to describe all situations observed in the real world.

To highlight the limitations of simulation, it is relevant to let the student build the simulation tool themselves, which requires the student to identify the relationships which have to be implemented into the model. With the development of programming languages, and Python is very popular today, it is of course also possible to consider different levels of usage: from modelling $\ln(k) = f(\%\text{solvent})$ to chromatogram simulation.
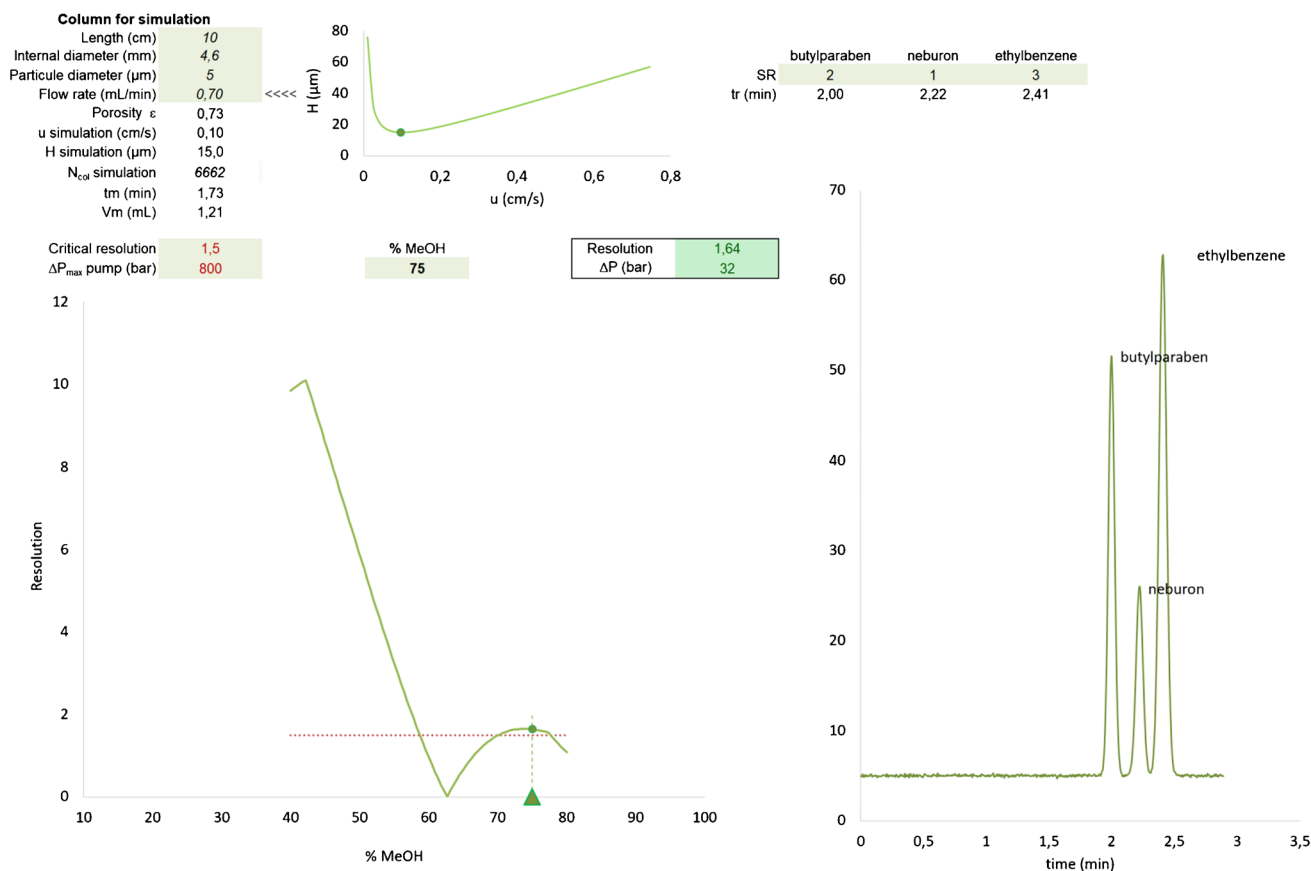
Fig. 6 Chromatographic optimisation curve: resolution versus organic modifier content for a virtual column. Expected chromatogram associated with the selected methanol content (here 75%)

Finally, troubleshooting in the chromatographic system can be very easily discussed from Fig. 3. For example, if the experimentally observed retention times of solutes increase, the arrows leading to "Retention times A and B" in Fig. 3 can be used to identify potential sources of drift.

– Has the hold-up time changed, caused by a change in mobile phase velocity created by a change in flow-rate? An experimental hold-up time could therefore be measured to test this hypothesis.
– Have the retention factors changed? Is the same modifier still being used? Has the percentage of modifier changed? Is the pump delivering the correct solvent composition?

Conclusion

Experts in chromatography will be surprised by the absence of a number of parameters in Fig. 3: temperature does not appear even though it modifies the partition coefficients; we do not talk about gradients even though they are widely used in the laboratory; we cannot see the consequences of external

dispersion effects that will be critical in UPLC... It was not our purpose at this stage, and students can gradually enrich the model when new concepts are introduced during lectures! As it is recommended for concept map [1], a new SCERM has to be built according to the *focus question* we want to address, a question that clearly specifies the problem or issue the map should help to resolve. In any case, the option of writing down the relationships between the variables can be retained to help the student understand the cause-and-effect relationships involved in chromatography. As a conclusion, we think that Systemic Cause-Effect Relation Map has several interesting characteristics in the teaching and learning process: SCERM is based on a systemic approach and extend concept maps, fishbone diagrams and causal loop diagrams; SCERM explicitly defines independent parameters, dependant criteria and cause to effect links; SCERM uses specific but common writing of equations. Based on this, we have observed during the last 14 years that SCERM enhances meaningful learning by explicitly defining the network of connections between concepts and experimental parameters. So, students use a more logical and systematic research process for solving problems during troubleshooting or method development.

Glossary

α	selectivity or separation factor
ε	porosity
η	viscosity
φ_{modif}	organic modifier content
σ	standard deviation of the peak, (t) time unit, (L) length unit
A	contribution of flow anisotropy
B	contribution of diffusion
C	contribution of mass transfer
d_i	column internal diameter
d_p	particle diameter
D_m	diffusion coefficient in mobile phase
F	flow rate
k	retention factor
k_w	extrapolated intercept of a plot of log k versus volume fraction of organic modifier
H	height equivalent to a theoretical plate
H_{min}	minimum value of height equivalent to a theoretical plate of a plot H versus u
i	analyte index: A or B
L	column length
N	efficiency
R_s	chromatographic resolution
u	mobile phase velocity
u_{opt}	optimum velocity in relation to H_{min}
S	slope of a plot of log k versus volume fraction of organic modifier
t_m	hold-up time
t_r	retention time

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. Novak JD, Cañas AJ. The theory underlying concept maps and how to construct and use them, Technical Report IHMC Cmap-Tools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition. 2008. Available at: http://cmap.ihmc.us/docs/pdf/TheoryUnderlyingConceptMaps.pdf.

2. Aubrecht KB, Dori YJ, Holme TA, Lavi R, Matlin SA, Orgill M, Skaza-Acosta H. Graphical tools for conceptualizing systems thinking in chemistry education. J. Chem. Educ. 2019;96:2888–290.

3. Johnstone AH. Chemistry teaching - science or alchemy? J Chem Educ. 1997;74(3):262. https://doi.org/10.1021/ed074p262.

4. High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model, Lloyd R. Snyder, John W. Dolan Ed, Wiley, ISBN: 978-0-471-70646-5. https://www.wiley.com/en-us/High+Performance+Gradient+Elution:+The+Practical+Application+of+the+Linear+Solvent+Strength+Model-p-9780471706465

5. Snyder L. Classification of the solvent properties of common liquids. J Chromatogr Sci. 1978;16:223–34.

Jérôme Randon works for the University Claude Bernard Lyon 1 (UCBL) at the Institute of Analytical Sciences (https://isa-lyon.fr) in the "Separation Techniques" team, which develops tools and miniaturised methods for chromatography and associated techniques. He is Full Professor at UCBL and particularly interested in gas chromatography using monolithic and micro structured columns using high-pressure conditions, as well as in the hyphenation of miniaturised detectors such as the nano gravimetric detector. He is also Director of the master's degree "Analysis and Control" (http://master-analyse-controle.univ-lyon1.fr/), a unique training degree on a European scale, from laboratory analysis to online industrial analysis, involved in two international consortia: EACH: Erasmus Mundus Excellence in Analytical Chemistry programme, and MSC: Measurement Science in Chemistry Summer School.

Vincent Dugas works for the University Claude Bernard Lyon 1 (UCBL) at the Institute of Analytical Sciences (https://isa-lyon.fr) in the "Separation Techniques" team, which develops tools and miniaturised methods for chromatography and associated techniques. He has been Associate Professor in Analytical Chemistry at UCBL since 2007. He is deeply involved in teaching the fundamentals of analytical chemistry at the undergraduate level and separation methods at the graduate level. He is an expert in the field of miniaturised analytical methods based on multi-dimensional separation tools.

Claire Demesmay also works for the University Claude Bernard Lyon 1 (UCBL) at the Institute of Analytical Sciences (https://isa-lyon.fr) in the "Separation Techniques" team, which develops tools and miniaturised methods for chromatography and associated techniques. She is Full Professor at UCBL and Team Leader of the teaching group dedicated to analytical chemistry at the undergraduate level. At postgraduate level, her teaching focuses on separation techniques from UPLC down to nanochromatography and electrophoresis. She is particularly interested in miniaturised separation methods using monolithic columns and in fully integrated analytical tools.