Journal of Chromatography A 1617 (2020) 460823

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Separation optimization in HPLC analysis implemented in R programming language

Ch. Zisi*, A. Pappa-Louisi, P. Nikitas

Department of Chemistry, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 20 October 2019
Revised 18 December 2019
Accepted 20 December 2019
Available online 24 December 2019

Keywords:
HPLC analysis
Isocratic and gradient elution
Computer assisted separation optimization and simulation
R programming language

ABSTRACT

A complete package of functions in the R-language has been written for professional separation optimization of complex mixtures of ionized and/or non-ionized solutes. The package includes functions for (a) base-line correction of experimentally recorded chromatograms, (b) modeling of chromatographic peak shapes and retention data, (c) prediction of the retention time of the test analytes and/or their chromatograms, and (d) separation optimization under either isocratic or single and/or double gradient elution conditions by changing the organic modifier(s) content and/or eluent pH. The optimization functions presented in this study offer two different modes for selection of optimal separation conditions: automatic and manual mode. In the automatic mode, the optimal separation conditions are determined by maximizing the resolution within separation time preset by the analyst. In the manual mode, the optimal separation conditions are selected via scatter or contour plots. The foreknowledge of the precise dependence of resolution and separation time upon one or two retention parameters of interest, provided by the proposed computer-assisted separation optimization method, gives chromatographers a feel of confidence for the selection of the optimal conditions for a desired separation. An illustrative video given in the Supplementary material may encourage a novice practitioner in R (software) programming language to follow the proposed separation optimization procedure in a real HPLC analysis.

1. Introduction

Optimization in separation sciences is an important demand from analysts who look for a desired resolution in minimum time. Reversed-phase liquid chromatography (RPLC) continues to be the most popular chromatographic technique in analytical laboratories. The goal of a separation optimization in RPLC is to obtain satisfactory separation conditions in the shortest time with only a few experiments. In simple separations, like those under isocratic conditions, the empirical optimization by trial and error may be a solution. However, in most cases of gradient elution, the large number of chromatographic parameters involved in the separation rules out the possibility of using empirical optimization and demand computer-based optimization procedures.

For this reason, several software packages have been marketed and can assist untrained users to set up separations. The most important of them are DryLab [1], PREOPT-W [2], OSIRIS [3], MICHROM [4], and ChromSword [5]. However, in most laboratories where RPLC is used daily, computer-assisted separation optimization is not used to improve separations. The lack of popularity of readily available multivariate optimization software may due to the fact that many chromatographers are uncomfortable with software tools, which are still rather expensive and not very user-friendly.

In our recently published work [6], a package of Excel VBA macros was developed for modeling and optimization of a mixture of analytes under multilinear single or double gradient elution conditions by changing the organic modifier(s) content and/or eluent pH. Moreover, a series of illustrative MS Excel spreadsheets were developed by authors to simulate the process of separation optimization under isocratic and simple gradient conditions depended only on organic modifier content [7]. Here, we present a free available R package that implements a computer-assisted separation optimization of a mixture of ionized and/or non-ionized solutes comprising a visualization of predicted chromatograms. In fact, this study is an upgraded combination of our previous work on separation optimization in HPLC analysis implemented in R programming language and software environment

R is a freely available open source programming language and software environment, which has become extremely popular among statisticians, bioinformaticians and academics because of its large user base and vast array of user-contributed packages

* Corresponding author.
E-mail address: chryzisi@chem.auth.gr (Ch. Zisi).

https://doi.org/10.1016/j.chroma.2019.460823
0021-9673/

designed for specific statistical problems. To date, there are some examples of using the R programming language in HPLC analysis [8–12]. However no R package for an automated separation optimization in HPLC analysis has been proposed, since most chromatographers may be unlikely to learn the basic functionality of R in order to take advantage of its power as a computational, statistical and graphical resource. In this study, we present a complete package of functions in the R-language for modeling both retention data and chromatographic peak shapes, for retention prediction and visualization of predicted chromatograms as well as for determination of the optimal separation conditions. To assess the performance of the R package, original (experimental) as well as artificial chromatographic data obtained under different conditions were analyzed and optimized.

2. Materials and methods

A computer-assisted separation optimization of a mixture of solutes comprising a visualization of predicted chromatograms involves the following steps: (1) an initial experimental study of the chromatographic behavior of solutes obtained by the least number of chromatographic runs adequately selected, (2) fitting of the experimental retention data of each solute to a retention model, (3) modeling of the peak shape of analytes in the form of a preferred function, (4) comparison of the predicted (simulated) chromatograms plotted under selected conditions with the corresponding experimental chromatograms recorded under the same conditions in order to test the accuracy of the optimization process, and (5) determination of the optimal separation conditions, i.e. the conditions that lead to the best separation of solutes under consideration. For using R programming language in the whole separation optimization procedure performed in the above stepwise fashion, ten R functions were written for modeling retention data and chromatographic peak shapes, for prediction retention and visualization of predicted chromatograms as well as for determination of the optimal separation conditions. The code of these functions is presented in the Supplementary *R_Functions.docx* file, whereas a summary of inputs and outputs of the different functions is given in Table 1. Moreover, the workflow of the implemented R functions is described in the Fig. 1.

The use of different functions written in R requires the installation of the R programming language to the host computer that can be downloaded from the official website http://www.r-project.org/. Once R has been installed, you can install additional packages/libraries as follows: Run R and the *R Console* should pop up. Then click on *Install package(s)* from the *Packages* menu at the top of the *R Console*. Alternatively, type in the *R Console* the command: *install.packages()*. A list of Uniform Resource Locators (URLs) appears to choose a location, a CRAN Mirror. After choosing a CRAN mirror, a list of available packages pops up to select the package you want to install. You can install one package per time. Install the packages *optimx, VGAM, plot3D*, and *numDeriv*.

To use the R functions described in the next Section, there are two options. 1)You may copy the code of a certain function provided in the Supplemental *R_Functions.docx* file, for example the code of *ifitk,* or the code of all functions and paste it to the command window of R (*R Console*). Then in this window type the function you are going to use. 2) Alternatively, you may save the code of a certain function or all functions as an *.RData* file. To do that, open a new *R Console* window, copy the code given in the Supplemental *R_Functions.docx* file, paste it to the *R Console*, and save it as for example *RChromOptim.RData* from *File > Save Workspace.* In order to use a certain function of the *RChromOptim.RData* file, you just open the *RChromOptim.RData* file with R and type the function as in the first option. Such an *RChromOptim.RData* file is given in the Supplementary Material along with thorough instructions of using all the functions comprised in the R package.

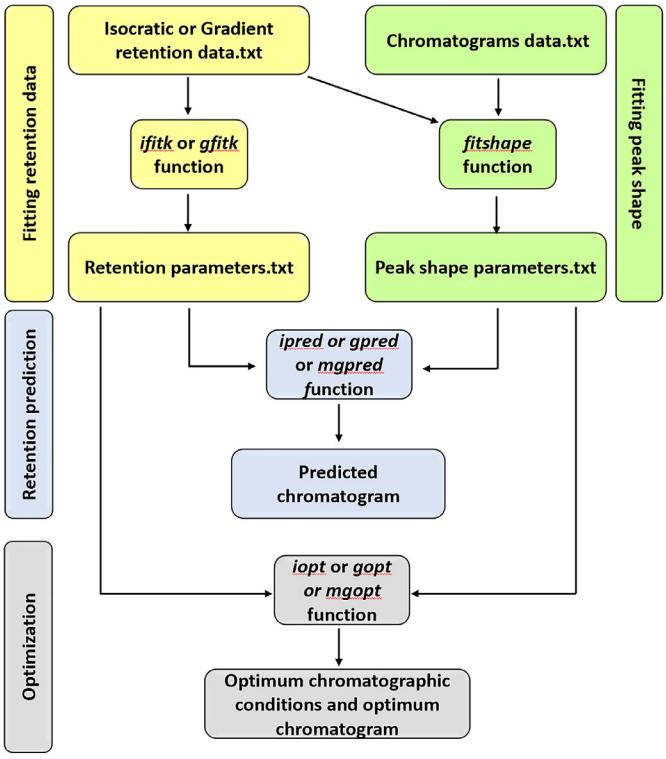

Fig. 1. Flowchart representing the procedures implemented in the whole proposed R package for separation optimization.

3. Software description: the proposed R functions and examples of their implementation in separation optimization procedure

3.1. Modeling retention data

Two R functions were written for fitting experimental retention data of solutes to a retention model; the first one, *ifitk*, determines retention model parameters from isocratic retention data, whereas the second one, *gfitk,* from simple/single or double monolinear gradient data. The retention models adopted in the present study are listed in Table 2 and they are the most important used in a reversed-phase type elution mode [6], although some of these models can also find application in other LC modes such as in HILIC [35,36] as well as in mixed-mode columns [37] Models 1–6 describe the effect of the organic modifier volume fraction, φ, on the solute retention factor, $k = (t_R - t_0)/t_0$, model 7 the effect of mobile phase pH on the retention of ionizable compounds (monoprotic acid or base), models 8 and 10 the simultaneous effect of eluent pH and organic modifier content on the retention of monoprotic acids or bases, and model 9 the combined effect of two organic modifier volume fractions, φ_1 and φ_2 on the solute retention. Note that *f* stands for φ in all R functions written for the proposed separation procedure.

The adjustable parameters of the selected retention model, depending on the parameter(s) that affect the elution procedure, i.e. organic modifier content(s) and/or eluent pH, are easily estimated in isocratic elution mode via the *ifitk* function. However, the determination of the retention parameters using gradient data is more demanding since it needs the solution of the fundamental gradient

Table 1
Inputs and outputs of different R functions written for the proposed separation optimization procedure.

Function name	What it does	Input	Output
ifitk	Fitting isocratic retention data	A *.txt* file containing isocratic retention data	Tables with a) retention model parameters, b) their standard deviations, c) their p-values, and d) the standard error of estimate as well as the sum of squared residuals. A table of the retention parameters may be saved as a *.txt* file.
gfitk	Fitting gradient retention data	A *.txt* file containing simple or double monolinear gradient retention data	The same with the above output
fitshape	Peak shape modeling	A *.txt* file with solutes retention data obtained under at least 2 or 3 similar isocratic or gradient conditions as well as *.txt* files with the chromatograms data recorded under the same conditions	A table of the estimated peak height and peak width parameters of tested solutes is saved as a *.txt* file.
ipred	Prediction/visualization of isocratic elution	A table of the retention model parameters comes from the output of the *ifitk* function including or not the output of the *fitshape* function in a *.txt* file. Optionally, a *.txt* file with the experimental chromatogram data recorded under conditions that are going to be predicted	The predicted values of solute retention times, the minimum R_s (or the minimum *dt*), a plot of the predicted chromatogram (or a plot of the predicted pseudo-chromatogram) with or without the corresponding experimental chromatogram. A table of the predicted chromatogram data may be saved as a *.txt* file
gpred	Prediction/visualization of single/double monolinear gradient elution	The same with the above input except that the table of the retention model parameters comes from the output of the *gfitk* function	The same with the above output plus the gradient profile of the gradient run.
mgpred	Prediction/visualization of single bilinear gradient elution	The same with the above input	The same with the above output
iopt	Optimization/visualization of isocratic elution	A table of the retention model parameters comes from the output of the *ifitk* function including or not the output of the *fitshape* function in a *.txt* file.	Scatter plots of R_s *(or dt)* and $t_{R,\max}$ vs. φ (or pH) as well as contour plots of *Rs* (or *dt)* and $t_R max$ vs. φ *and* pH *(or* φ_1 and φ_2) depending on the retention model. A table with the data of these plots along with the optimal isocratic conditions and optimal solute retention times. The table of plots data may be saved as *.txt* files. A plot of the optimum predicted chromatogram (or pseydo-chromatogram).
gopt	Optimization/visualization of single/double monolinear gradient elution	The same with the above input except that the table of the retention model parameters comes from the output of the *gfitk* function	Scatter plots of R_s *(or dt)* and $t_{R,\max}$ *x* vs. $t_{G,\varphi}$ (or $t_{G,\mathrm{pH}}$) as well as contour plots of R_s (or *dt)* and $t_{R,\max}$ vs. $t_{G,}$, *and* $t_{G,\mathrm{pH}}$ *(or* $t_{G,\varphi 1}$ and $t_{G,\varphi 2}$) depending on the retention model. A table with the data of these plots along with the optimal gradient conditions and optimal solute retention times. The table of plots data may be saved as *.txt* files. A plot of the optimum predicted gradient profile along with the optimal chromatogram (or pseydo-chromatogram).
mgopt	Optimization/visualization of single bilinear gradient elution	The same with the above input	The optimal gradient conditions, the optimal solute retention times and a plot of the optimun predicted gradient profile along with the optimal chromatogram (or pseydo-chromatogram).
ifitopt	Automatical (in one step) optimization/visualization of isocratic elution	A *.txt* file with solutes retention data obtained under isocratic conditions as well as *.txt* files with the chromatograms data recorded under the same conditions	A table of the estimated retention and peak shapes parameters of solutes along with optimal values of φ, R_s and solute retention times. The table of solutes parameters is saved as a *.txt* file. A plot of R_s and $t_{R,\max}$ vs. φ. A table with data of the above plot may be saved as *.txt* file. A plot of the optimal predicted chromatogram.

Key: .txt file is a tab-delimited. *txt* file; chromatogram(s) data are time-dependent detector response data; R_s is the resolution of the least resolved pair of adjacent solutes; *dt* is the difference of retention times of the least resolved pair of adjacent solutes; pseydo-chromatogram is a chromatogram with vertical lines instead of peaks at solute retention times; $t_{R,\max}$ is the retention time of the last eluted solute, $t_{G,\varphi}$, $t_{G,\mathrm{pH}}$, $t_{G,\varphi 1}$ and $t_{G,\varphi 2}$ is the gradient duration of the linear variation of φ, pH, φ_1 and φ_2, respectively.

Table 2
Retention models adopted in R functions used for separation optimization procedure under either isocratic or single and/or double gradient elution conditions.

No	Retention model	Used in elution mode	Refs.
1	$lnk = c_0 - c_1\varphi$	Isocratic dependent on φ, monolinear and bilinear φ-gradient	[13–16]
2	$lnk = c_0 - c_1 ln\varphi$		[15,17,18]
3	$lnk = c_0 + c_1\varphi + c_2\varphi^2$		[14,19–21]
4	$lnk = c_0 - c_1 ln(1 + c_2\varphi)$		[19,22]
5	$lnk = c_0 - 2ln(1 + c_1\varphi) - \frac{c_2\varphi}{1+c_1\varphi}$		[23]
6	$lnk = c_0 - \frac{c_2\varphi}{1+c_1\varphi}$		[13,19,24]
7	$k = \frac{k_0+k_1 10^{pH-pK}}{1+10^{pH-pK}}$	isocratic dependent on pH, monolinear and bilinear pH-gradient	[25–28]
8	$k = \frac{k_0(1+r10^{pH-pK})}{1+10^{pH-pK}}$ where $k_0 = c_0 - c_1\varphi$	isocratic dependent on both φ and pH as well as monolinear φ-gradient in different eluent pH	[29]
9	$lnk = c_0 + c_1\varphi_1 + c_2\varphi_2 + c_3\varphi_1^2 + c_4\varphi_2^2 + c_5\varphi_1\varphi_2$	isocratic dependent on both φ_1 and φ_2 as well as double mono-linear φ_1, φ_2-gradient	[30,31]
10	$k = \frac{k_0+k_1 10^{pH-pK}}{1+10^{pH-pK}}$ where $k_0 = c_0 - c_1\varphi$ and $k_1 = c_3 - c_4\varphi$	isocratic dependent on both φ and pH as well as double monolinear φ, pH -gradient	[32–34]

Key: c_0, c_1, c_2, c_3, c_4, c_5, pK, r are adjustable parameters; $r = k_0/k_1$, where k_0 and k_1 are the retention factors of the neutral and fully ionized species for monoprotic acids and vice versa for monoprotic bases.

elution equation

$$\int_0^{t_R} \frac{dt}{t_0 k} = 1 \tag{1}$$

where t_R is the solute elution time and t_0 is the column dead time. In the *gfitk* function, by default, an analytical solution is applied for the solution of Eq. (1) when the models 1 to 5 or 8 of Table 2 are used for fitting gradient data, while the approximate approach proposed by Nikitas-Pappa [6] is applied when the rest retention models in Table 2 are used. Fig. 2 depicts an example of adjustable parameter assayed using the *gfitk* function for fitting artificial φ gradient data given in the *g-ret.fit* spreadsheet of the *Data.xlsx* file in the Supplementary Material. Thorough instructions of using all the functions comprised in the R package are also given in the Supplementary Material.

3.2. Peak shape modeling

For modeling a symmetric Gaussian peak of a single analyte or a group of analytes that appear in a chromatogram obtained under either isocratic or gradient conditions the R function *fitshape* may be used. Assuming that the peak height, h, as well as the peak width parameter s (related with the peak width at base, w with $w = 4s/\sqrt{2}$) for each analyte recorded in different chromatograms depend primarily on the retention time, t_R , we can write a typical chromatographic Gaussian peak as [7,38]

$$y = h(t_R)\exp\left(\frac{-(t - t_R)^2}{s(t_R)^2}\right) \tag{2}$$

where y is the detector response at time t.

In fact, the *fitshape* function determines the dependence of each peak height and width upon its retention time via the following equations:

$$h(t_G) = h_0 + h_1 t_R + h_2 t_R^2 \tag{3}$$

$$\text{and } s(t_G) = s_0 + s_1 t_R \tag{4}$$

assuming a quadratic dependence of peak height on t_R but a linear dependence for peak width. However, the user of the R package has the choice to select a linear dependence of solute peak height upon t_R instead of the quadratic one, so that only the parameters h_0 and $h_1$are calculated and consequently fewer chromatographic runs are demanded for peak shape modeling. Fig. 3 shows a screenshot of the *fitshape* function concerning the estimation of peak shape parameters of a group of three solutes eluted under experimental isocratic conditions given in the *shape-fit* spreadsheet of the *Data.xlsx* file in the Supplementary Material. The final output of the *fitshape* function, which is a table of the estimated peak shape parameters of tested analytes, is depicted in Fig. 4.

3.3. Retention prediction and visualization of predicted chromatograms

After the estimation of the retention and peak shape parameters of all tested solutes, the values of these parameters are saved as a tab delimited *.txt* file and transferred into the R functions written for retention prediction and visualization of predicted chromatograms (see Table 1). The comparison of a real/experimental chromatogram with the predicted one presented in the *R Graphics Device* shows the degree of accuracy of the optimization process. In other words, a perfect similarity between a simulated chromatogram and the original/experimental one, recorded under the same conditions, shows the quality of modeling of both retention and peak shape data.

Three R functions were written for prediction retention data and visualization of predicted chromatograms obtained under conditions that are different than those used in the fitting procedure. The first function, *ipred,* predicts isocratic retention data and plots the corresponding simulated chromatograms obtained under isocratic conditions, the second one, *gpred,* predicts gradient retention data and plots the corresponding simulated chromatograms generated under simple or double monolinear gradient conditions, and the third function, *mgpred,* predicts gradient retention data and plots the corresponding simulated chromatograms created for simple bilinear gradient conditions. A comparison between the experimental/original and the calculated/predicted chromatogram can also be performed. An example of the *mgpred* function used to predict a chromatogram obtained under φ-bilinear gradient conditions, given in the *mg-predictions* spreadsheet of the *Data.xlsx* file in the Supplementary Material, is depicted in Fig. 5. This figure shows a perfect similarity between the simulated chromatogram created for the separation conditions tested and the original one, which shows the accuracy of the retention and peak shape adjustable parameters used in this procedure (see the *mg-predictions* spreadsheet of the *Data.xlsx* file in the Supplementary Material). Note that, in case no peak shape parameters are available, i.e. in

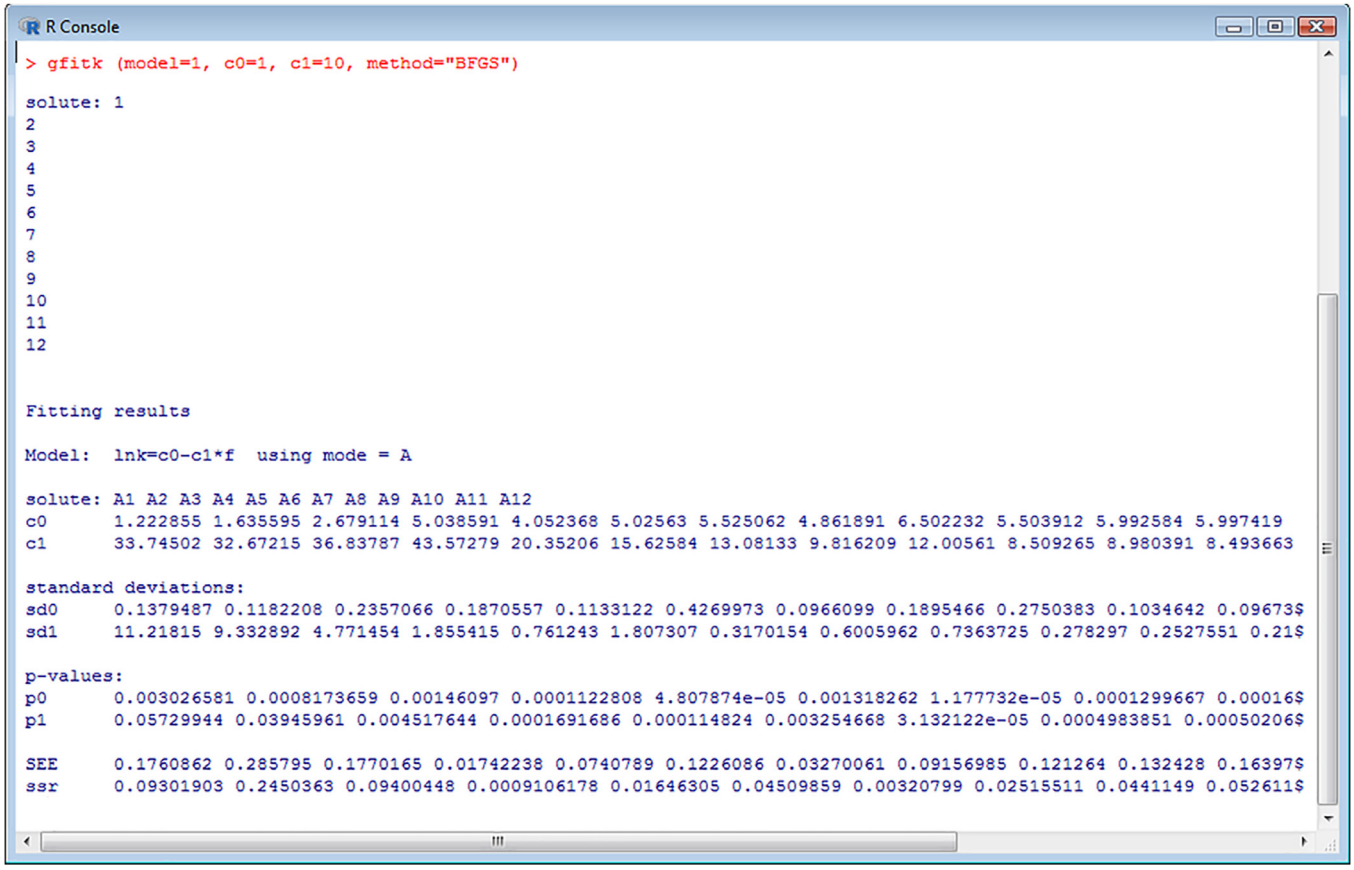

Fig. 2. Example output of the *gfitk* function for fitting φ gradient data (see the text for details) to model 1of Table 2: Solute adjustable parameters, their standard deviations, their *p*-values, the Standard Error of the Estimate (SSE) and the sum of squared residuals (ssr).

case only solute retention parameters of the *mg-predictions* spreadsheet of the *Data.xlsx* file in the Supplementary Material are used, a pseudo-chromatogram will be presented in the *R Graphics Device*. Even though, the comparison with the experimental chromatogram remains very informative for the accuracy of separation optimization procedure.

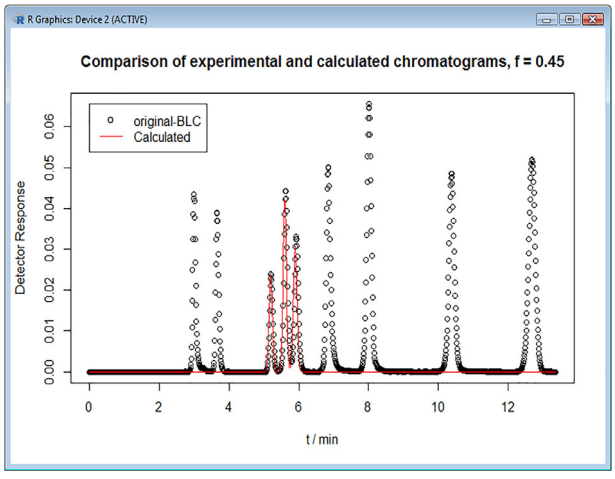

Fig. 3. A screenshot of the procedure of peak shape modeling by the *fitshape* function. A comparison of an experimental chromatogram used in this procedure with the calculated one is given in order to be checked the quality of the peak modeling. See the text and *R_Fuctions* in Supplementary Material for details. Original-BLC means baseline corrected experimental chromatogram. The symbol "f" stands instead of φ in all R package for organic modifier content.

3.4. Separation optimization

Having checked the accuracy of the estimated solute parameters, these parameters are transferred via a tab delimited *.txt* file into the R functions written for separation optimization (see Table 1). Three functions were written for the determination of optimal chromatographic conditions, i.e. the conditions that lead to the best separation of a test mixture within a desirable value of chromatographic time, t_{max}. The first one, *iopt*, optimizes isocratic separation conditions and simulates isocratic chromatograms. The second one, *gopt*, optimizes simple/single or double monolinear gradient conditions and simulates chromatograms generated under similar gradient profiles, whereas the third R-function, *mgopt*,

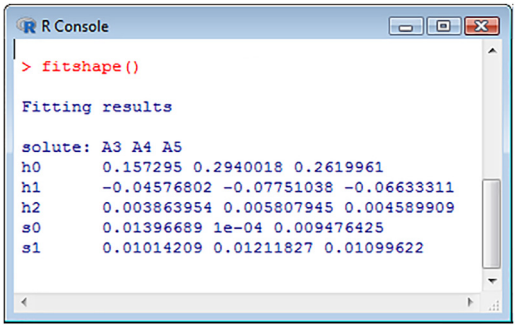

Fig. 4. The final output of the *fitshape* function in the procedure of peak shape modeling depicted in Fig. 2. A table of peak shape parameters of three solutes estimated by this R function using experimental isocratic retention data (see the text for details).

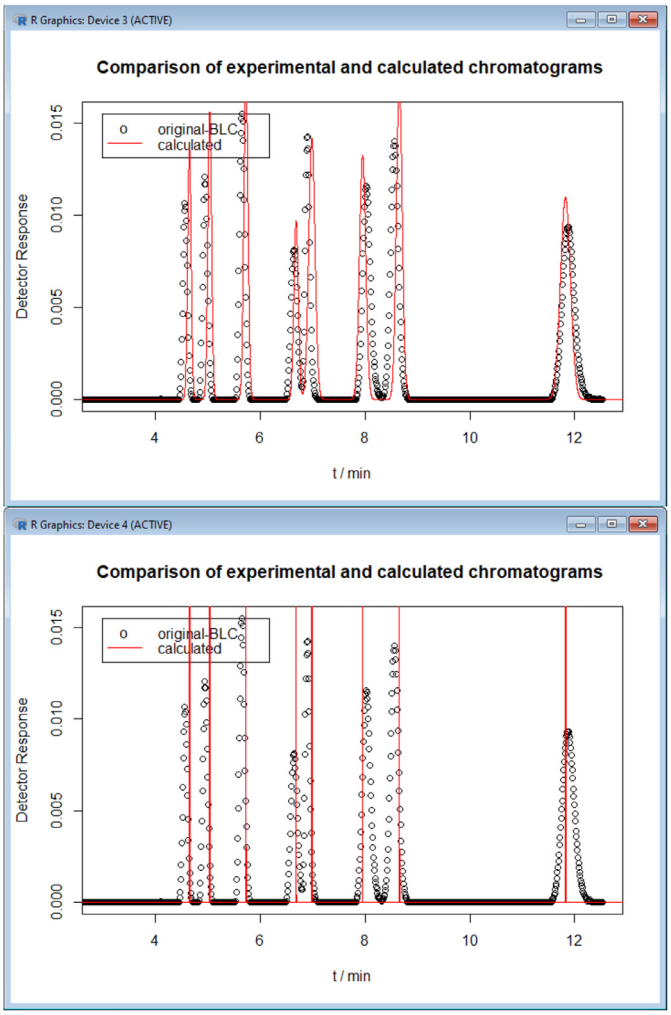

Fig 5. Example output (in the *R Graphics Device*) of the *mgped* function for prediction and visualization of φ-bilinear gradient elution using retention model parameters including (upper plot) or not peak shape parameters (lower plot), see the text for details.

optimizes simple bilinear gradient conditions and simulates chromatograms created for similar optimal gradient profiles.

Suppose, for example, that we run the *gopt* function using the solutes parameters for model 1 given in the *g-mg-optim* spreadsheet of the *Data.xlsx* file in the Supplementary Material. Then, the *gopt* function determines automatically the optimal φ-gradient conditions according to the default values of the arguments, i.e. this function, based on model 1, determines the optimal duration of monolinear variation of organic modifier content, $t_{G,\varphi}$, between $\varphi_{\min}$=0 and $\varphi_{\max}$=0.6, for the best separation in the preset separation time, t_{max} = 30 min. The estimated optimal φ-gradient conditions, i.e. $t_{G,\varphi}$ = 10 min, along with the corresponding gradient elution times of solutes, t_R, as well as the resolution of the least resolved pair of adjacent solutes, R_s, are presented in the *R Console*. Moreover, a table of $t_{G,\varphi}$, R_s and $t_{R,\max}$ values is included in the *R Console*, whereas the corresponding overlaid plots of R_s vs. t_G and $t_{R,\max}$ vs. t_G is presented in the *R Graphics Device*. Figs. 6 and 7 show these outputs of the *gopt* function in the *R Console* and in the *R Graphics Device*, respectively. The appreciation of scatter plots of R_s *and* $t_{R,\max}$ vs. $t_{G,\varphi}$ depicted in Fig. 7 enables a manual selection of optimal φ-gradient conditions.

Note that, in case two separation parameters are going to be optimized in an optimization procedure, a manual selection of optimal separation conditions is also possible but via contour plots. An example of the application of the *gopt* function in an optimization procedure of double φ *and* pH-monolinear gradients conditions using the solutes parameters for model 10 given in the *g-mg-optim* spreadsheet of the *Data.xlsx* file in the Supplementary Material is depicted in Fig. 8. In this case the parameters that should be optimized are the φ-gradient duration, $t_{G,\varphi}$ as well as the pH-gradient duration $t_{G,\mathrm{pH}}$. According to the default arguments (see supplemental *R_Instructions*), the *gopt* function automatically determines that the optimum duration of linear φ-variation between $\varphi_{\min}$ = 0.2 and $\varphi_{\max}$ = 0.7 is $t_{G,\varphi}$ = =2 min, whereas the optimum duration of linear pH-variation between pH_{min} = 2 and pH_{max} = 9 is $t_{G,\mathrm{pH}}$ = 9 min, which yields the best resolution, R_s = 1.739, within a desirable value of chromatographic time, t_{max} = 15 min. From the contour plots of Fig. 8, we observe that this optimal pair ($t_{G,\varphi}$, $t_{G,\mathrm{pH}}$) automatically assayed by the *gopt* function corresponds to a yellow spot at the left bottom of the R_s contour plot, which means that the selection of these optimal values for $t_{G,\varphi}$ and $t_{G,\mathrm{pH}}$ is not secure. However, in case we select $t_{G,\mathrm{pH}}$~20 min, then, irrespective of the $t_{G,\varphi}$ value (in the range between 25 and 35 min), we will obtain optimal resolution within a chromatographic time, t_{max} ~ 19 min. Consequently, the foreknowledge of the precise dependence of resolution and separation time upon one or two retention parameters of interest, provided by the proposed R package for separation optimization, gives chromatographers a feel of confidence for the selection of the optimal conditions for a desired separation.

3.4. Automatical (in one step) optimization and visualization of isocratic elution

An automatic function is also developed, the *ifitopt* function, which is partly a combination of *ifitk*, *fitshape* and *iopt* functions. This function optimizes the isocratic separation condition of a solutes mixture when the retention is described by one of the models 1 to 3 of Table 2, while the function has first determined the solutes retention and peak shape parameters, which are necessary for the optimization procedure.

4. Conclusions

Overall, the developed R software is designed as easy-handled and user-friendly tool for simulating and optimizing liquid chromatographic separations. In fact, this software is an upgraded version of our work on the issue of computer-aided separation optimization, which is implemented in R programming language. The proposed software offers two different modes for selection of the optimal separation conditions (automatic and manual mode), which gives chromatographers a feel of confidence for the selection of the optimal separation conditions. Moreover, as the peak shape modeling is also performed by this package, it provides to the analyst-user the opportunity to generate simulated chromatograms under selected chromatographic conditions. We demonstrate in this study several examples where the R package provides optimal isocratic, monolinear and/or bilinear separations condition governed by variations of φ and/or pH. On the other hand, this study may encourage the users to extend the capabilities of the proposed R software through adding new functionalities to this software such as for example the modeling of non-ideal Gausian peak shapes (tailors peaks) or the introduction of new retention models and/or gradient conditions. Finally, an illustrative video given in the Supplementary material may encourage a novice practitioner in R (software) programming language to follow the proposed separation optimization procedure in a real HPLC analysis of experimental data given in the *video* spreadsheet of the *Data.xlsx*. A flowchart representing a sum-up of the above separation optimization process is depicted in Fig. 9.

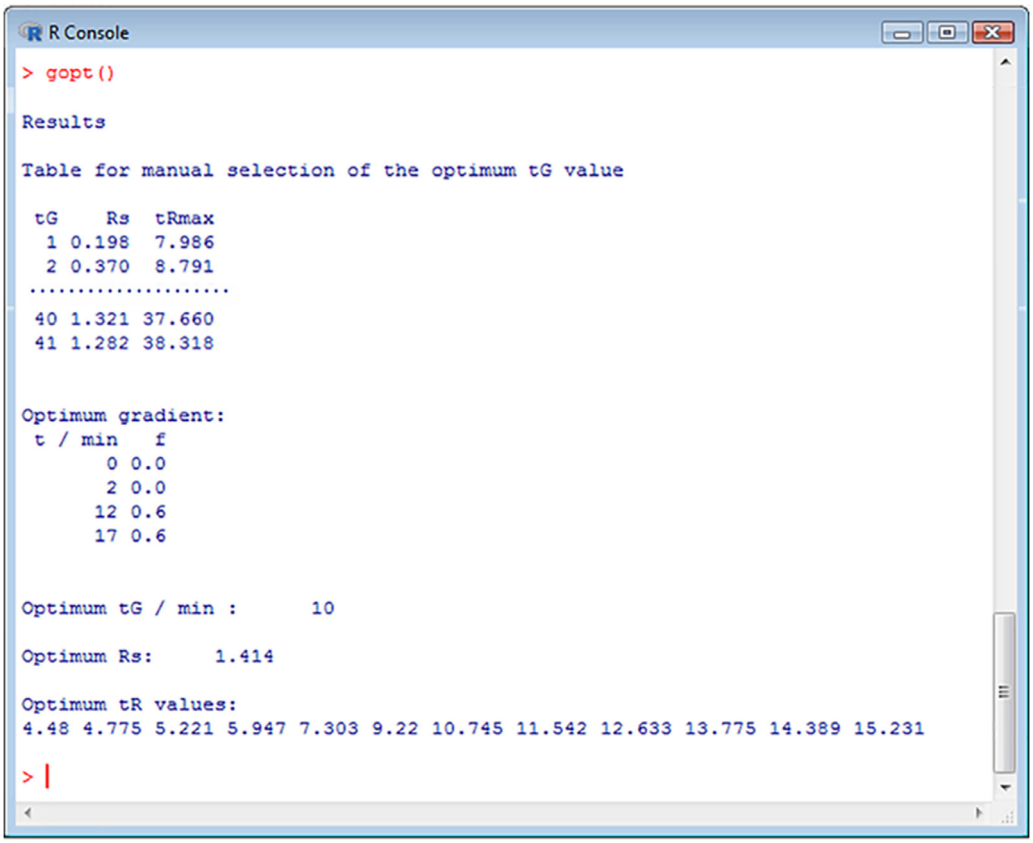

Fig. 6. Example output of the *gopt* function in the *R Console* (see the text for details).

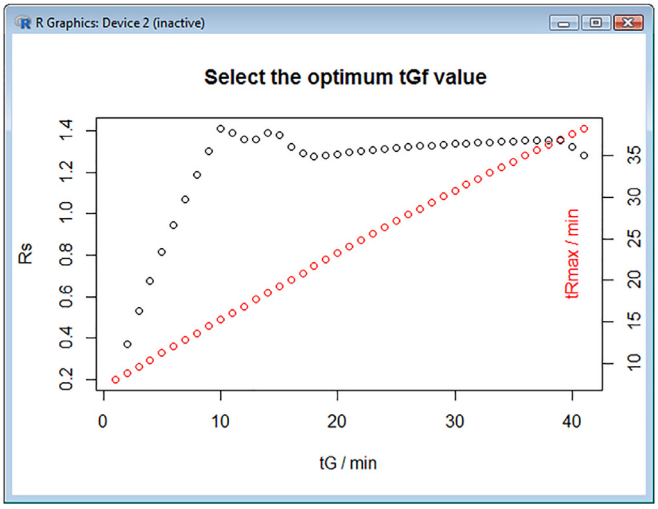

Fig. 7. Example output of the *gopt* function in the *R Graphics Device* (see the text for details). Scatter plots of R_s and $t_R max$ vs. $t_{G,\varphi}$ assayed by the *gopt* function in an optimization procedure of φ-monolinear gradients conditions (see the text for details).

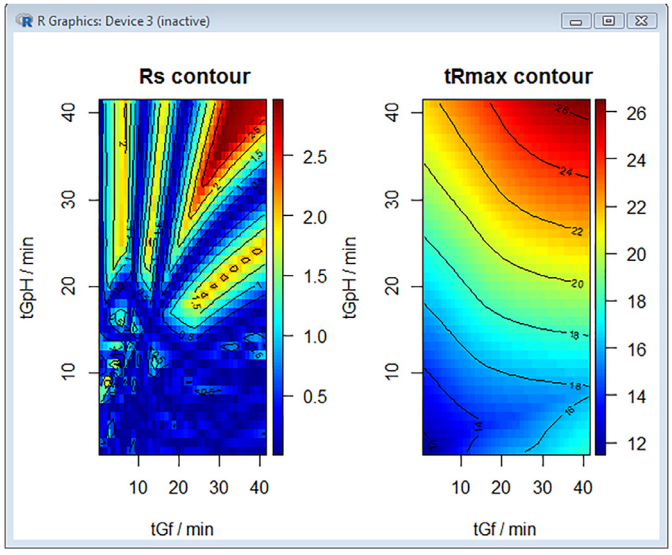

Fig. 8. Example output of the *gopt* function in the *R Graphics Device* (see the text for details). Contour plots of R_s and $t_{R,\max}$ vs. $t_{G,\varphi}$ and $t_{G,\mathrm{pH}}$ assayed by the *gopt* function in an optimization procedure of double φ and pH-monolinear gradients conditions (see the text for details).

Supplementary material

- Thorough Instructions for using the R package for separation optimization in the *R_Instructions.pdf file*.
- The code of R functions is presented in the *R_Functions.docx* file.
- The code of R functions is saved in *RChromOptim.RData* file for a comfortable use of the optimization software (ZIP).
- Original or artificial chromatographic data as well as solute parameters for implementation of the optimization software are given in the *Data.xlsx* file.
- The tab-delimited .txt files used as input in different examples of implementation of R functions described in the *R_Instructions.pdf* file as well as in the video are given in the folder "Text files" (ZIP)
- An Illustrative video supporting the user in following the proposed separation optimization procedure (.mp4)

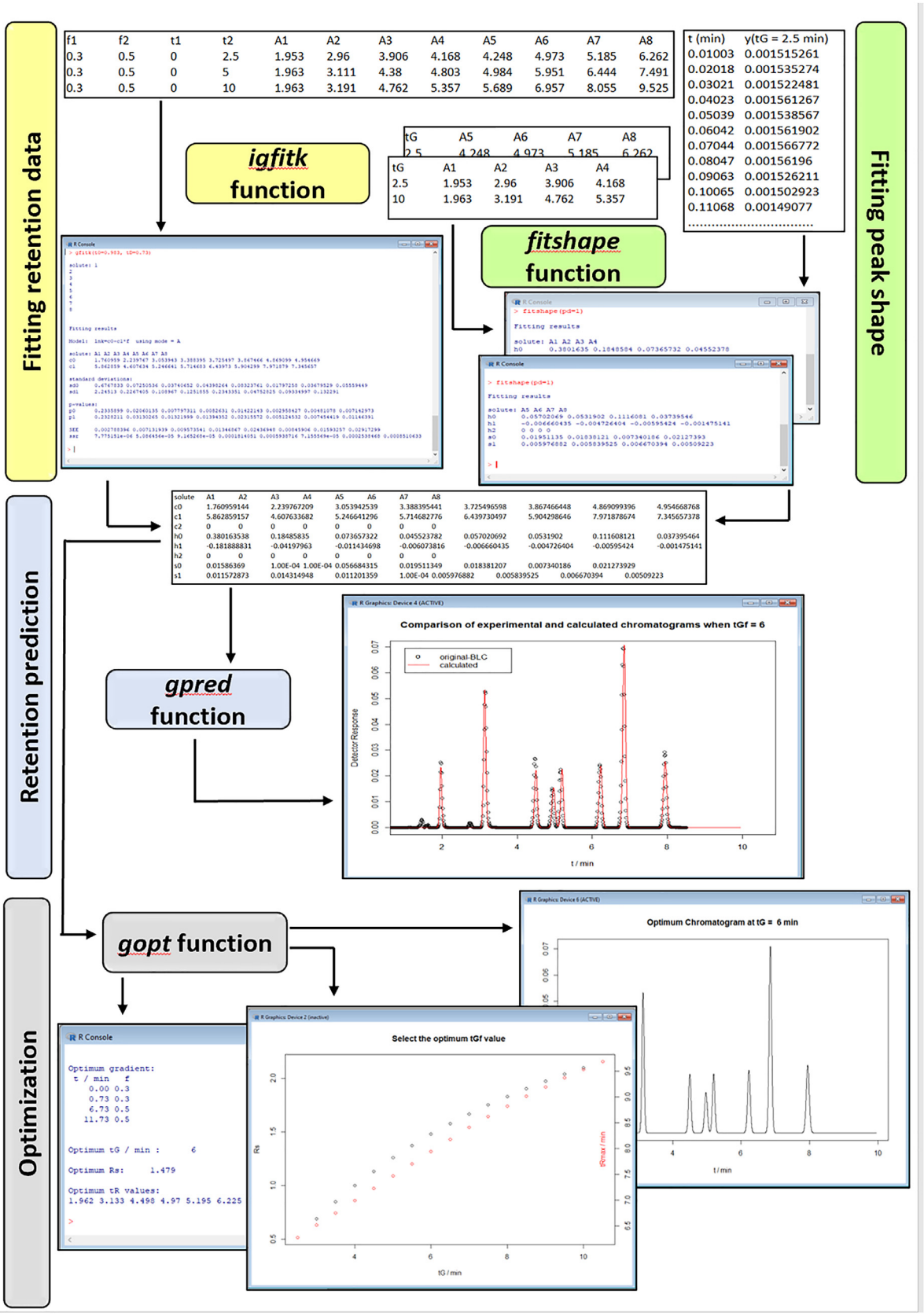

Fig. 9. Representative scheme of the operation procedure of the whole R package for separation optimization of experimental gradient data taken from Ref. [39].

Declaration of Competing Interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2019.460823.

References

[1] I. Molnar, Computerized design of separation strategies by reversed-phase liquid chromatography: development of DryLab software, J. Chromatogr. A 965 (2002) 175–194, doi:10.1016/S0021-9673(02)00731-8.

[2] R. Cela, M. Lores, PREOPT-W: A simulation program for off-line optimization of binary gradient separations in HPLC—II. Data management and miscellaneous aspects of use, Comput. Chem. 20 (1996) 193–202, doi:10.1016/0097-8485(95)00063-1.

[3] S. Heinisch, E. Lesellier, C. Podevin, J.L. Rocca, A. Tchapla, Computerized optimization of RP-HPLC separation with nonaqueous or partially aqueous mobile phases, Chromatographia 44 (1997) 529–537, doi:10.1007/BF02466747.

[4] S. Carda-Broch, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Evaluation of several global resolution functions for liquid chromatography, Anal. Chim. Acta 396 (1999) 61–74, doi:10.1016/S0003-2670(99)00368-2.

[5] W.-D. Beinert, R. Jack, V. Eckert, S. Galushko, V. Tanchuck, I. Shishkina, A program for automated HPLC method development, Am. Lab. 33 (2001) 14–15.

[6] S. Fasoula, Ch. Zisi, H. Gika, A. Pappa-Louisi, P. Nikitas, Retention prediction and separation optimization under multilinear gradient elution in liquid chromatography with Microsoft Excel macros, J.Chromatogr. A 1395 (2015) 109–115, doi:10.1016/j.chroma.2015.03.068.

[7] S. Fasoula, P. Nikitas, A. Pappa-Louisi, Teaching simulation and computer-aided separation optimization in liquid chromatography by means of illustrative Microsoft Excel spreadsheets, J. Chem. Educ. 94 (2017) 1167–1173, doi:10.1021/acs.jchemed.7b00108.

[8] M.D. Peris-Diaz, M.I. Alcoriza-Balaguer, J.C. Garcia- Canaveras, F. Santonja, E. Sentandreu, A. Lahoz, RpeakChrom, Novel R package for the automated characterization and optimization of column efficiency in high-performance liquid chromatography analysis, Electrophoresis 38 (2017) 2985–2995, doi:10.1002/elps.201700180.

[9] J. Wills, J. Edwards-Hicks, A.J. Finch, AssayR, A simple mass spectrometry software tool for targeted metabolic and stable isotope tracer analyses, Anal. Chem. 89 (2017) 9616–9619, doi:10.1021/acs.analchem.7b02401.

[10] D.T. Holmes, cp-R, an interface the R programming language for clinical laboratory method comparisons, Clin. Biochem. 48 (2015) 192–195, doi:10.1016/j.clinbiochem.2014.10.015.

[11] D. May, W. Law, M. Fitzgibbon, Q. Fang, M. McIntosh, Software platform for rapidly creating computational tools for mass spectrometry-based proteomics, J. Proteome Res. 8 (2009) 3212–3217, doi:10.1021/pr900169w.

[12] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Peak alignment using wavelet pattern matching and differential evolution, Talanta 83 (2011) 1108–1117, doi:10.1016/j.talanta.2010.08.008.

[13] P. Nikitas, A. Pappa-Louisi, Retention models for isocratic and gradient elution in reversed-phase liquid chromatography, J. Chromatogr. A 1216 (2009) 1737–1755, doi:10.1016/j.chroma.2008.09.051.

[14] P.J. Schoenmakers, H.A.H. Billiet, R. Tijssen, L. de Galan, Gradient selection in reversed-phase liquid chromatography, J. Chromatogr. 149 (1978) 519–537, doi:10.1016/S0021-9673(00)81008-0.

[15] P. Jandera, J. Churacek, Gradient elution in liquid chromatography: I. The influence of the composition of the mobile phase on the capacity ratio (retention volume, band width, and resolution) in isocratic elution – theoretical considerations, J. Chromatogr. 91 (569) (1974) 207–221, doi:10.1016/S0021-9673(01)97901-4.

[16] L.R. Snyder, J.W. Dolan, J.R. Gant, Gradient elution in high-performance liquid chromatography. 1. Theoretical basis for reversed-phase systems, J. Chromatogr. 165 (1979) 1–3, doi:10.1016/S0021-9673(00)85726-X.

[17] L.R. Snyder, Role of the solvent in liquid-solid chromatography, Rev. Anal. Chem. 46 (1974) 1384–1393, doi:10.1021/ac60347a052.

[18] P. Nikitas, A. Pappa-Louisi, P. Agrafiotou, A. Mansour, Multilinear gradient elution optimization in reversed-phase liquid chromatography based on logarithmic retention models: application to separation of a set of purines, pyrimidines and nucleosides, J. Chromatogr. A 1218 (2011) 5658–5663, doi:10.1016/j.chroma.2011.06.084.

[19] P. Nikitas, A. Pappa-Louisi, P. Agrafiotou, Effect of the organic modifier concentration on the retention in reversed-phase liquid chromatography: II. Tests using various simplified models, J. Chromatogr. A 946 (2002) 33–45, doi:10.1016/S0021-9673(01)01536-9.

[20] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, Description of solute retention over the full range of mobile phase compositions in reversed-phase liquid chromatography, J. Chromatogr. 282 (1983) 107–121, doi:10.1016/S0021-9673(00)91596-6.

[21] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, The solubility parameter as a tool in understanding liquid chromatography, Chromatographia 15 (1982) 205–214, doi:10.1007/BF02261542.

[22] F. Antia, Cs. Horvath, Dependence of retention on the organic modifier concentration and multicomponent adsorption behavior in reversed-phase chromatography, J. Chromatogr. 550 (1991) 411–424, doi:10.1016/S0021-9673(01)88553-8.

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

[24] U.D. Neue, C.H. Phoebe, K. Tran, Y.-F. Cheng, Z. Lu, Dependence of reversed-phase retention of ionizable analytes on pH, concentration of organic solvent and silanol activity, J. Chromatogr. A 925 (2001) 49–67, doi:10.1016/S0021-9673(01)01009-3.

[25] Cs. Horvath, W. Melander, I. Molnar, Liquid chromatography of ionogenic substances with nonpolar stationary phases, Anal. Chem. 49 (1977) 142–154, doi:10.1021/ac50009a044.

[26] R.M. Lopes-Marques, P.J. Schoenmakers, Modelling retention in reversed-phase liquid chromatography as a function of pH and solvent composition, J. Chromatogr. 592 (1992) 157–182, doi:10.1016/0021-9673(92)85084-7.

[27] P.J. Schoenmakers, R. Tijssen, Modelling retention of ionogenic solutes in liquid chromatography as a function of pH for optimization purposes, J. Chromatogr.A 656 (1993) 577–590, doi:10.1016/0021-9673(93)80820-X.

[28] P. Nikitas, A. Pappa-Louisi, New equations describing the combined effect of pH and organic modifier concentration on the retention in reversed-phase liquid chromatography, J. Chromatogr. A 971 (2002) 47–60, doi:10.1016/S0021-9673(02)00965-2.

[29] A. Andres, A. Tellez, M. Roses, E. Bosch, Chromatographic models to predict the elution of ionizable analytes by organic modifier gradient in reversed phase liquid chromatography, J. Chromatogr. A 1247 (2012) 71–80, doi:10.1016/j.chroma.2012.05.070.

[30] P.J. Schoenmakers, H.A.H. Billiet, L. de Galan, Systematic study of ternary solvent behavior in reversed-phase liquid chromatography, J. Chromatogr. 218 (1981) 261–284, doi:10.1016/S0021-9673(00)82060-9.

[31] A. Pappa-Louisi, P. Nikitas, M. Karageorgaki, Retention prediction in ternary solvent reversed-phase liquid chromatography systems based on the variation of retention with binary mobile phase composition, J. Chromatogr. A 1091 (2005) 21–31, doi:10.1016/j.chroma.2005.07.009.

[32] F. Gritti, N. Tanaka, G. Guiochon, Comparison of the fast gradient performance of new prototype silica monolithic columns and columns packed with fully porous and core–shell particles, J. Chromatogr. A 1236 (2012) 28–41, doi:10.1016/j.chroma.2012.02.046.

[33] R.M. Lopes Marques, P.J. Schoenmakers, C.B. Lucasius, L. Buyden, Modelling chromatographic behavior as a function of pH and solvent composition in RPLC, Chromatographia 36 (1993) 83–95, doi:10.1007/BF02263843.

[34] P.J. Schoenmakers, N. Mackie, R.M. Lopes Marques, Optimizing separations in reversed-phase liquid chromatography by varying pH and solvent composition, Chromatographia 35 (1993) 18–32, doi:10.1007/BF02278552.

[35] H. Gika, G. Theodoridis, F. Mattivi, U. Vrhovsek, A. Pappa-Louisi, Retention prediction of a set of amino acids under gradient elution conditions in hydrophilic interaction liquid chromatography, J. Sep. Sci. 35 (2012) 376–383, doi:10.1002/jssc.201100795.

[36] H. Gika, G. Theodoridis, F. Mattivi, U. Vrhovsek, A. Pappa-Louisi, Hydrophilic Interaction Ultra Performance Liquid Chromatography retention prediction under gradient elution. Anal. Bioanal, Chem 404 (2012) 701–709, doi:10.1007/s00216-012-6015-6.

[37] P. Balkatzopoulou, S. Fasoula, H. Gika, P. Nikitas, A. Pappa-Louisi, Retention prediction of highly polar ionizable solutes under gradient conditions on a mixed-mode reversed-phase and weak anion-exchange stationary phase, J. Chromatogr. A 1396 (2015) 72–76, doi:10.1016/j.chroma.2015.03.082.

[38] P. Nikitas, A. Pappa-Louisi, A. Papageorgiou, On the equations describing chromatographic peaks and the problem of the deconvolution of overlapped peaks, J. Chromatogr. A 912 (2001) 13–29, doi:10.1016/S0021-9673(01)00524-6.

[39] Ch. Zisi, A.-M. Mangipa, E. Boutou, A. Pappa-Louisi, Separation optimization of a mixture of ionized and non-ionized solutes under isocratic and gradient conditions in reversed-phase HPLC by means of microsoft excel spreadsheets, Separations 5 (2018) 19, doi:10.3390/separations5010019.