Computer-driven optimization of complex gradients in comprehensive two-dimensional liquid chromatography

Molenaar, S.R.A.; Bos, T.S.; Boelrijk, J.; Dahlseid, T.A.; Stoll, D.R.; Pirok, B.W.J.

Computer-driven optimization of complex gradients in comprehensive two-dimensional liquid chromatography

Stef R.A. Molenaar [a,b,#], Tijmen S. Bos [b,c,#], Jim Boelrijk [b,d,e,#], Tina A. Dahlseid [f], Dwight R. Stoll [f], Bob W.J. Pirok [a,b,*]

[a] van 't Hoff Institute for Molecular Sciences, Analytical Chemistry Group, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[b] Centre for Analytical Sciences Amsterdam (CASA), Amsterdam, The Netherlands
[c] Division of Bioanalytical Chemistry, Amsterdam Institute of Molecular and Life Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands
[d] AMLab, Informatics Institute, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[e] AI4Science Lab, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[f] Department of Chemistry, Gustavus Adolphus College, Saint Peter, MN 56082, United States

ABSTRACT

Method development in comprehensive two-dimensional liquid chromatography (LC × LC) is a complicated endeavor. The dependency between the two dimensions and the possibility of incorporating complex gradient profiles, such as multi-segmented gradients or shifting gradients, renders method development by "trial-and-error" time-consuming and highly dependent on user experience. In this work, an open-source algorithm for the automated and interpretive method development of complex gradients in LC × LC-mass spectrometry (MS) was developed. A workflow was designed to operate within a closed-loop that allowed direct interaction between the LC × LC-MS system and a data-processing computer which ran in an unsupervised and automated fashion. Obtaining accurate retention models in LC × LC is difficult due to the challenges associated with the exact determination of retention times, curve fitting because of the use of gradient elution, and gradient deformation. Thus, retention models were compared in terms of repeatability of determination. Additionally, the design of shifting gradients in the second dimension and the prediction of peak widths were investigated. The algorithm was tested on separations of a tryptic digest of a monoclonal antibody using an objective function that included the sum of resolutions and analysis time as quality descriptors. The algorithm was able to improve the separation relative to a generic starting method using these complex gradient profiles after only four method-development iterations (*i.e.,* sets of chromatographic conditions). Further iterations improved retention time and peak width predictions and thus the accuracy in the separations predicted by the algorithm.

1. Introduction

Comprehensive two-dimensional liquid chromatography (LC × LC) is a very powerful tool for the separation of complex samples [1,2]. When executed well, the addition of a second dimension (²D) separation can improve peak capacity and resolution significantly [3,4]. It is thus not surprising to see LC × LC being used for the analysis of a variety of different samples, for example, polymers [5], proteins [6,7], lipids [8], oil [9], bioactive compounds [10,11], etc. However, due to the complexity (*i.e.,* number of compounds) of the samples and the required systems, method development is increasingly more complicated with

every new advancement in the field of LC × LC [12,13]. First, the analyst must choose a suitable retention mechanism to use in the first dimension (¹D) separation, and then a compatible, and preferably complementary, ²D retention mechanism must be chosen [14,15]. Once the system parameters have been chosen, the method can be optimized using different programmed gradients [16,17]. Optimizing the gradient settings by "trial & error" however, can take months.

Optimization of elution conditions may be accelerated by computational methods [18]. Neural networks can be used for retention time predictions and optimization [19–23]. Other methods rely on direct optimization of "wet" gradient conditions using black-box approaches,

such as evolutionary algorithms or Bayesian optimization in LC [24,25], *in silico* LC and LC × LC [26,27], and gas chromatography [28]. Although these methods are powerful, they do not incorporate any known relationships between mobile-phase modifier fraction and retention factor as is done in retention modeling, and typically require a significantly increasing number of wet measurements as the number of adjustable parameters increases.

In contrast, empirical retention modeling allows the retention factor to be correlated to the mobile phase composition, and groups have investigated this for *e.g.,* reversed-phase LC [29–31], normal-phase LC [32], ion-exchange chromatography [33], hydrophilic interaction LC [34–37], and supercritical-fluid chromatography [38–41]. The advantage of empirical retention modeling is that typically only a minimum of two or three measurements are required to build a model. This makes retention modeling attractive for computer-aided method development strategies that can be used to optimize separations [42,43]. Some examples are Drylab [44,45] or PEWS [36] for one-dimensional (1D) LC, or PIOTR [46] and the recent open-access MOREPEAKS software [47] for LC and LC × LC.

What most of the neural network and method development concepts have in common is that retention times must be known for each compound under different conditions. When optimizing a complex separation involving a large number of compounds, it can be tedious to manually track compounds across each measurement, making such approaches impractical if the acquisition and processing of data cannot be automated [48]. Consequently, peak-tracking algorithms have been developed for both 1D [49–54] and 2D chromatography [55–58].

Recently, our groups worked on the automation of pre-processing, peak detection, peak tracking, and consequently, the computer-driven gradient optimization for an LC separation of a complex mixture of peptides without human interaction [59]. However, this work was limited to LC-MS. For LC × LC-MS, automated method-development is more complex. The number of parameters is much larger and the dependence of the two dimensions on each other renders method development significantly more complicated. This is especially acute when sophisticated gradient assemblies (*e.g.,* shifting gradients) are used in the second dimension [60]. Typically the slow speed of ^{2}D separations compared to ^{1}D peak widths results in a much lower number of data points in the first dimension compared to the second dimension. Consequently, retention prediction errors in the first dimension may cause an analyte to elute in a different ^{2}D gradient than was expected, causing prediction errors in the second dimension to propagate. Furthermore, establishing accurate retention models is more challenging in LC × LC. For example, fast gradients in the second dimension may induce gradient deformation, rendering the acquired retention models less accurate [61,62].

In this study, we developed an algorithm and workflow for the computer-driven optimization of mobile-phase gradients used in LC × LC-MS separations. Retention modeling, peak tracking, and optimization algorithms were developed to communicate with each other and the LC × LC-MS system directly. The repeatability of determining retention models is evaluated. To facilitate the efficacy of the methods proposed by the algorithm, equations for retention time modeling in shifting gradients and peak width predictions were derived. Finally, the workflow is tested on a complex sample consisting of a tryptic digest of an IgG1 monoclonal antibody (mAb) where the gradient profiles of both dimensions were optimized simultaneously without the need for intervention by an analyst.

2. Experimental

2.1. Chemicals

The sample consisted of a tryptic digest of an IgG1 mAb. Details related to the preparation of this sample were reported previously [63]. Ammonium hydroxide, ammonium bicarbonate, and acetonitrile were obtained from Sigma Aldrich (St. Louis, MO). Honeywell Research Chemicals (Muskegon, MI) was the supplier of formic acid. HPLC-grade water was obtained from an in-house Milli-Q system (Burlington, MA).

2.2. Chromatographic system

2.2.1. LC × LC-MS

The LC instrument was an Agilent Infinity II 2D-LC system consisting of two binary pumps (G7120A) with Jet Weaver V35 mixers (G7120-68135), an autosampler (G4226A), and two column ovens (G7116B). The active solvent modulation (ASM) valve interface (p/n: 5067-4266) used to connect the two dimensions was set up with two nominally identical 40 μL sample loops and a restriction capillary (170 × 0.12 mm, 1.9 μL) in order to obtain an ASM factor of 3. Dwell volumes were estimated at 0.225 mL in the first dimension and 0.070 mL plus 0.040 mL loop volume in the second dimension. The 2D-LC instrument was controlled by Agilent OpenLab CDS ChemStation Edition (C.01.10 [287]), with a 2D-LC add-on (rev. A.01.04 [033]). The mass spectrometer was a quadrupole time-of-flight (Q-TOF) instrument (G6545XT) from Agilent Technologies (Waldbronn, Germany) equipped with a Dual Agilent Jet Stream Electrospray Ionization (AJS ESI) source. A standard tuning compound mixture (Agilent, p/n: G1969-85000) was used to calibrate the mass analyzer. Hexakis (1H,1H,3H-perfluoropropoxy) phosphazene was used as a reference mass (*m/z* 922.0098) compound to calibrate mass spectra and was sprayed continuously into the electrospray source via a secondary reference nebulizer. The Q-TOF was controlled by Agilent MassHunter Workstation Data Acquisition.

2.2.2. LC columns

The column in the first dimension was an Agilent Poroshell HPH-C18 (2.1 × 150 mm, 1.9 μm), and the ^{2}D column was an Agilent Zorbax SB-C18 (2.1 × 30 mm, 3.5 μm). Column dead volumes were estimated based on column dimensions and a porosity of 0.7 at 0.365 mL in the first dimension and 0.069 mL in the second dimension.

2.3. Chromatographic conditions

2.3.1. First Dimension

Gradient elution was used with 10 mM ammonium bicarbonate pH 9.5 (Solvent A) and ACN (Solvent B). Three scanning gradients were used where the gradient profile remained constant (2-100% B) but the gradient time (t_G) was varied with times of 60, 45, and 30 min. The ^{1}D column temperature was 40°C, and the ^{1}D flowrate was 0.08 mL·min^{-1}. This flow rate was chosen to allow for an appropriate modulation time in the second dimension. For each analysis, 4 μL of mAb digest was injected (3 μg/mL).

2.3.2. Second Dimension

Gradient elution was used with 0.1% formic acid in water (Solvent A) and ACN (Solvent B). Three scanning gradients were used where the gradient profile remained constant (2-100% B), but the gradient time (t_G) was varied with times of 24, 18, and 9 s. Towards the end of the ^{1}D gradient, the modifier fraction might be high. Therefore, to improve efficiency, at the beginning of each ^{2}D cycle, the mobile phase was held at 2% B for 5.4 s to act as the diluent during ASM [64]. When using the initial set of scanning gradients, a modulation time of 30 s was used; in subsequent methods, a modulation time of 45 s was used with a 37 s gradient time. All other gradient profiles were calculated by the algorithm and uploaded to the LC × LC by an in-house C++ script. The ^{2}D column temperature was 60°C, and the ^{2}D flow rate was 2 mL·min^{-1}. The ^{2}D flow entering the MS nebulizer was reduced to approximately 0.3 mL·min^{-1} using a simple tee split and short, narrow restriction capillaries (75 μm i.d.) chosen to achieve the desired split ratio.

2.4. MS instrument and conditions

For MS detection, the Q-TOF mass spectrometer was operated in positive ion mode. MS settings were as follows: gas temperature, 320°C; drying gas, 8 L·min^{-1}; nebulizer, 35 psi; sheath gas temperature, 350°C; sheath gas flow, 11 L·min^{-1}; VCap, 3500 V; nozzle voltage, 1000 V; mass range, 200-2500 *m/z*; acquisition rate, 4 spectra·s^{-1}.

2.5. Software

Peak tracking was performed with a previously developed algorithm [56,58]. Scripts for retention modeling with the LSS and quadratic models for multi-segmented gradients and shifting gradient profiles were written in-house. Retention parameters were estimated using the *multistart* function in combination with the *fmincon* function with an optimality tolerance of 10^{-6} and 3000 maximum function evaluations. Optima were calculated using the *ga* (genetic algorithm) function within Matlab-software R2021b (Mathworks, Natick, MA, USA). Individual plate numbers per compound were estimated using the *fminsearch* function with an in-house script based on the peak compression model by Hao *et al.* [24] Overall system communications between the MS, LC × LC, and Matlab algorithms were performed using a script written in Python 3.8.12. Methods on the LC × LC were started using C++ code with Visual Studio 2022, similar to our previous work [59], as illustrated in Fig. 1. Raw MS data were converted into .mz5 format by ProteoWizard 3.0.22144 64-bit [65] using a threshold count most-intense (2000).

3. Results & discussion

3.1. Retention modeling

3.1.1. Challenges in establishing retention models

The methodology used in this study is based on the prediction of optimal separation conditions using retention models. These models relate retention with mobile phase conditions and theoretically enable the prediction of method conditions that will yield an optimal separation. Within the scientific community, there is discussion on how to accurately establish such retention models using data obtained using gradient elution. Quarry *et al.* and later Vivó-Truyols *et al.* showed that retention models obtained from isocratic data are not easily transferred to gradient elution [66,67]. The authors recommended a linear model with a normalized mobile-phase polarity parameter (*i.e.,* a parameter that incorporates the polarity of a mixture of mobile phases into one parameter) when switching between isocratic mode to gradient mode, and vice-versa [67]. However, this would require knowledge about the change in polarity by mixing the mobile phases. Based on their work, the authors recommended a non-linear model to construct retention models for isocratic elution and a linear model for gradient elution. Investigating how to construct these retention models, Den Uijl *et al.*

recently studied the use of what is known as scouting or scanning gradients in LC [68]. These are typically linear gradients with different gradient slopes to probe retention coefficients. As expected, the authors found that retention time predictions made within the gradient slope domain probed are more accurate than predictions outside of this domain, although retention time predictions in gradient programs in a similar modifier-range of the scanning gradients generally provide the most accurate results. This is in agreement with the work of Nikitas *et al.* [69]. However, for our workflow, an estimate of the final gradient shape cannot be made at the start of the optimization. Therefore, after the scanning gradients, retention models are iteratively updated with new gradient information to improve the retention models.

The determination of retention parameters is complicated by the indirect assessment of these parameters. In contrast to isocratic elution, the exact relationship between retention time and modifier fraction is not known. Therefore, fitting retention parameters to gradient elution data is dependent on curve fitting [70]. This is even more complicated by the physical characteristics of the system that are difficult to model. Especially for LC × LC separations, this may become problematic. The fast gradients in the second dimension can induce gradient deformation meaning that the difference between the solvent composition profile specified by the method is different from what is actually delivered to the column [61]. Thus, the fitted retention model may be inaccurate. Moreover, a limited number of data points are available in the ^{1}D axis of the 2D chromatogram due to the infrequent modulation of ^{1}D peaks. Consequently, the first peak-moment, or the retention time, of a peak may not be calculated accurately [71].

3.1.2. Selection of retention model

Before starting the computer-driven optimization, the robustness of two commonly used retention models for reversed-phase LC was compared: the quadratic retention model ($\ln k = \ln k_0 - S_1 \cdot \phi + S_2 \cdot \phi^2$) [29] and the linear solvent strength (LSS), or exponential, model ($\ln k = \ln k_0 - S \cdot \phi$) [72]. Where $\ln k$ is the natural logarithm of the retention factor, ϕ the used modifier fraction, k_0 the retention factor in pure weak solvent and S the parameters of change due to the modifier fraction. The frequently used Neue-Kuss model has been left out of this study. Even though it has been shown to describe isocratic retention curves well [31, 36], it also has been shown that the model doesn't perform well when fitting retention parameters to a limited number of gradient separations [35]. The quadratic retention model describes the actual retention curve more accurately using isocratic measurements, but the model has three parameters and thus needs at least three data points to be estimated with more data points increasing the accuracy of the model. The LSS model is a simpler model with only two parameters and consequently only needs a minimum of two data points to approximate retention behavior.

These points are important because the errors associated with incorrect retention models can propagate through the method development workflow, and thus it is important to investigate the robustness
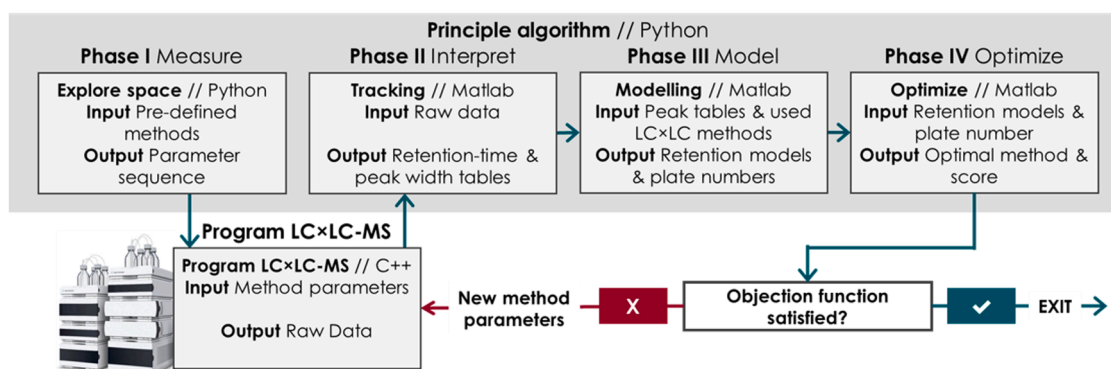

Fig. 1. Schematic overview of the workflow used for automation of method development.

of the retention models. Using eight 1D-LC chromatograms from previous work [59], retention parameters for both models were fitted 1000 times using a *multistart fmincon* fit with 20 different randomly selected starting points per iteration. Meaning that each of the 1000 iterations selects the best retention curve over 20 curve-fit routines that are initiated with different starting parameters. Fig. 2 shows results for the three analytes with the largest sum of standard deviations in the 1000 established retention parameters using the quadratic model (*i.e.* $\sigma_{\ln k_0} + \sigma_{S_1} + \sigma_{S_2}$), and are thus most difficult to fit. Both the 1000 determined retention curves of the quadratic model and the LSS model are plotted. Each retention curve is shown as slightly transparent. Thus, darker curves show more overlapping retention curves. The quadratic model (indicated with green) exhibits significant spread as shown in Figs. 2A and 2B, and it even features a bimodal distribution, centered around two local minima. This showed that the staring parameters for the quadratic model had a large impact on the determined retention curves. The LSS model (indicated in red) however, is more robust and shows all 1000 determined retention curves as one line on all three plots. As described above (Section 3.1.1), it is expected that the accurate assessment of retention parameters may be more difficult in LC × LC. Therefore, it was decided to continue with the LSS model in the current work.

3.2. Workflow in the computer-driven optimization of LC × LC

The computer-driven optimization was performed using two separate modules: 1) the LC × LC-MS system and 2) a PC that (*i*) tracked peaks over all available chromatograms [58], (*ii*) estimated retention parameters, (*iii*) calculated the next gradient steps, and (*iv*) submitted the optimized gradient profiles to the LC × LC-MS system. Before retention modeling, the LC × LC-MS system was tasked to perform three scanning gradients from $\varphi = 0.02$ to $\varphi = 1.00$ in 60, 45, and 30 min in the first dimension and in the second dimension from $\varphi = 0.02$ to $\varphi = 1.00$ in 24, 18 and 9 s, respectively.

As soon as data from two scanning gradients were recorded, sufficient data was available to approximate retention parameters using the LSS model and the algorithm was able to start an estimation for an optimal separation. Thus, after the second and all succeeding (scanning) measurements, the algorithm tracked peaks and fitted retention parameters using all available retention times and gradient profiles with the *fmincon* function in combination with the *multistart* function at 20 different randomly selected points. When retention models for all paired peaks were estimated, the column plate number per peak was estimated using the multilinear gradient peak compression model by Hao *et al.* [24] using the *fminsearch* function with the experimental peak widths and gradient conditions as inputs.

After estimating retention parameters and individual plate numbers, the algorithm was directed to calculate favorable (*i.e.,* maximizing resolution and minimizing analysis time) gradient programs in both dimensions simultaneously. The system was allowed to calculate ^{1}D gradient profiles with three gradient steps and in the second dimension the system calculated boundaries for a two-step shifting gradient (see Section 3.3). The resolution score (O_{R_s}) was estimated by calculating the normalized sum of resolutions of all peaks (*n*) with their nearest neighbors. The resolution metric described by Schure *et al.* [73] was used to estimate the resolution (R_S) (see Eq. 1) and the value was normalized using a maximum resolution (R_{opt}) of 2. Baseline-separated peaks are deemed optimal, thus the resolution score should not be increased for values above this baseline separation. Therefore, R_S values above the optimal resolution were capped to R_{opt}. The resolution with only a peak's nearest neighbor was chosen so as not to be biased toward peaks that were not in the vicinity of the peak. Afterward, the complement to 1 of the normalized R_S was used ($O_{R_s} = 1 - \frac{R_S}{R_{opt}}$), so that it could be minimized like the time score (O_t).

$$R_S = \sqrt{\frac{(^1t_2 - {}^1t_1)^2}{4(^1\sigma_1 + {}^1\sigma_2)^2} + \frac{(^2t_2 - {}^2t_1)^2}{4(^2\sigma_1 + {}^2\sigma_2)^2}} \tag{1}$$

Where $^1t_1$, $^1t_2$, $^2t_1$, and $^2t_2$ are the retention times in the first dimension of peaks #1 and #2 and the second dimension of peaks #1 and #2, respectively. And $^1\sigma_1$, $^1\sigma_2$, $^2\sigma_1$, and $^2\sigma_2$ are the corresponding standard deviations of the peaks.

The time score of a separation was calculated by normalizing the sum of the gradient times ($\sum t_G$) between 0 and the maximum allowed analysis time (t_{max}) (in our case a value of 90 min was selected). This means that the time score would be minimized if $\sum t_G = 0$. This, however, would result in no resolution (and thus a large O_{R_s} value). There is a trade-off between the two scores. This will theoretically result in a short analysis time with a high sum of resolutions if suitable weights are selected. The *ga* function was used to minimize a weighted performance score (O_{perf}) where the mean normalized resolution received a weight (w_{R_s}) of 0.8 and the time score received a weight (w_t) of 0.05, as indicated in Eq. 2. There are exceptions to Eq. 1. If the algorithm predicts that an analyte elutes later than t_{max} in the first dimension or later than the sum of the modulation time and the column dead time in the second dimension and wrap-around occurs, a performance score of 1 is outputted, meaning that the conditions are not viable.

$$O_{perf} = O_{R_s} * w_{R_s} + O_t * w_t = \left(1 - \frac{\sum R_S}{R_{opt} * n}\right) * w_{R_s} + \frac{\sum t_G}{t_{max}} * w_t \tag{2}$$

Performing peak tracking, determining retention parameters, and calculating optimal gradient profiles is computationally intensive, especially when dealing with large numbers of peaks encountered in complex samples. Nine method parameters are used for the shifting gradient, plus an additional two parameters per ^{1}D gradient, and a parameter for the initial mobile phase modifier fraction. This adds up to
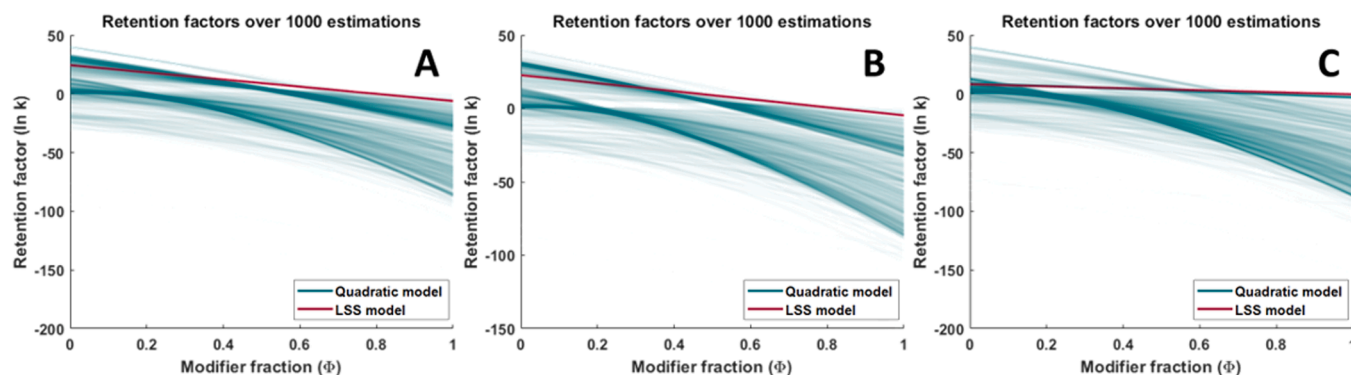

Fig. 2. Analysis of the spread of the three retention curves with the highest summed variance in the quadratic model over 1000 determinations of the retention parameters using a *multistart fmincon* fit with 20 different starting points on 1D-LC data. Plots A, B, and C are for the analytes with the highest, second highest, and third highest sum of standard deviations of the retention parameters in the quadratic model, respectively.

a total of 16 method parameters that need to be optimized. If all 16 parameters are only allowed three different values, this would already result in $3^{16} = 4.3 \cdot 10^7$ possible different combinations. In this work, a PC with 24 top-of-the-line commercially available Central Processing Units (CPU) was used. Although not all 24 CPUs could be used in parallel in each stage of the optimization protocol, an effort was made to perform calculations in parallel as often as possible. Despite this, the PC would often require one to two hours of calculation time before it determined a suitable gradient profile for a subsequent experiment. Thus, a crucial time-saving step was implemented.

As the LSS model only has two parameters, retention models could be estimated using only two scanning gradients. Therefore, when the third scanning gradient was being performed, the PC could already begin calculating a favorable (*i.e.,* minimizing O_{perf}) gradient profile for measurement #4. Upon completion of the third scanning gradient, the retention models were updated with the new information and while measurement #4 was being performed, the gradient profile for measurement #5 was calculated and so on, as shown in Fig. 3A. This effectively means that all calculations were one measurement behind and not all information that could be available was utilized. However, the LC × LC-MS system did not have to wait for the calculations to be completed for the (n-1)[th] analysis to start the nth analysis. The total time required for method development is thus reduced significantly compared to the case where both parts (algorithm and LC × LC-MS instrument) wait on each other (Fig. 3B).

3.3. Two-step shifting gradients in two-dimensional chromatography

In the simplest implementation of LC × LC, the gradient in the second dimension remains constant over the separation. In cases where the two separation mechanisms are not orthogonal, the 2D separations often benefit from a more advanced gradient that shifts over time to provide better usage of the 2D separation space [63,74,75]. The use of shifting gradients, however, will typically result in longer method-development times compared to conventional LC × LC [76]. Fig. 4 shows an example of a two-step shifting gradient in the second dimension. Retention modeling within a shifting gradient requires the knowledge of several parameters. Some of these parameters, such as the modulation time (t_{mod}) strongly depend on the [1]D parameters, and the maximum gradient time (^2t_G) is dependent on the modulation time. The algorithm can tune the [2]D parameters; the initial and final mobile phase modifier fractions (φ_{init} and φ_{final}) at the beginning, the intermediate points (t_{low} and t_{high}) and at the end of the shifting gradient program (t_{shift_end}). This effectively means that the algorithm needs to fit nine parameters used in the construction of the boundaries for the shifting gradient program.

To determine the experimental retention times in LC × LC, there is a need to consider if the [2]D retention times are before or after the [2]D column dead time, as the retention time can be misaligned by folding the chromatogram. When determining retention parameters using
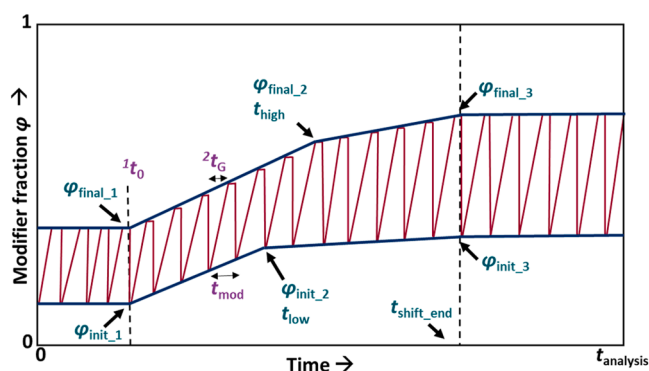

Fig. 4. Outline of a two-step shifting gradient for use in the second dimension. The [2]D gradient conditions change over time starting from the column dead time ($^1t_0$) of the first dimension to the shift-end time (t_{shift_end}). Purple values are [1]D dependent, green values can be chosen by the algorithm. ^2t_G indicates the gradient time in the second dimension, which is limited by the modulation time (t_{mod}). φ_{init_1}, φ_{final_1}, φ_{init_2}, φ_{final_2}, φ_{init_3} and φ_{final_3} indicate the initial and final modifier fractions at the beginning of the measurements, the intermediate times t_{low} and t_{high} and the final time t_{shift_end}, after which the gradient remains constant until the analysis time ($t_{analysis}$).

experimental retention data, first the retention times were determined for both dimensions. Then, the [2]D retention time was evaluated to see if it was smaller than the [2]D column dead time. If this was the case, the modulation time was added to the experimental [2]D retention time, and the retention time in the first dimension was reduced by the modulation time. Note that this does not mean that breakthrough had occurred, as the analyte still eluted during the programmed gradient and had not passed the dead time of the next modulation yet. Then the modulation time was subtracted from the [1]D retention time if it was added to the [2]D retention time.

Retention modeling in the second dimension requires the determination of the [1]D retention time. The modulation number in which the peak appeared was determined using the [1]D retention time. If the analyte eluted from the first dimension after t_{shift_end}, the final [2]D conditions were used for retention modeling in the second dimension. Otherwise, if the [1]D retention time was in between the start and end of the shifting profile, the [2]D gradient profile was determined using Eqs. 3 to 11:

$$N_{mod} = \left\lceil \frac{^1t_R}{t_{mod}} \right\rceil \tag{3}$$

$$S_{low\_1} = \frac{\varphi_{init\_2} - \varphi_{init\_1}}{t_{low} - t_0} \tag{4}$$

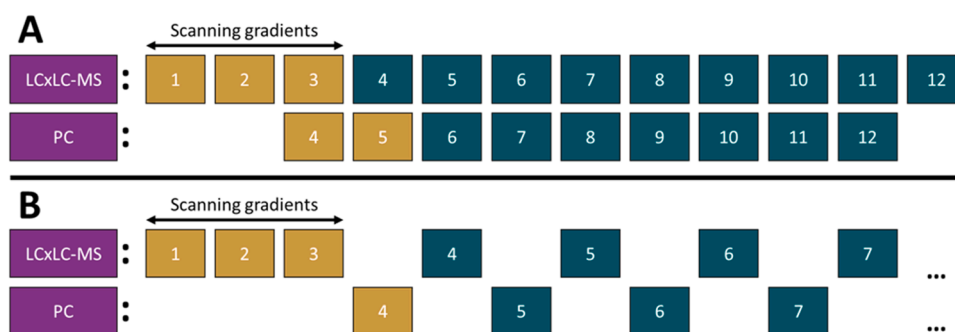$$S_{low\_2} = \frac{\varphi_{init\_3} - \varphi_{init\_2}}{t_{shift\_end} - t_0} \tag{5}$$

Fig. 3. Illustration of different ways of organizing measurements and calculations over time. Scanning gradients and calculations that only use data from these analyses are shown in yellow. Optimized methods that are calculated and future calculations that take these measurements into account are shown in green. **A)** While the LC × LC-MS is acquiring data, the next gradient profile is already being calculated by the PC. **B)** The LC × LC-MS and PC calculations wait for each other.

$$S_{\text{high\_1}} = \frac{\varphi_{\text{final\_2}} - \varphi_{\text{final\_1}}}{t_{\text{high}} - t_0} \tag{6}$$

$$S_{\text{high\_2}} = \frac{\varphi_{\text{final\_3}} - \varphi_{\text{final\_2}}}{t_{\text{shift\_end}} - t_0} \tag{7}$$

$$\varphi_{\text{start\_1}} = (N_{\text{mod}} * t_{\text{mod}} - t_0) * S_{\text{low\_1}} + \varphi_{\text{init\_1}} \tag{8}$$

$$\varphi_{\text{start\_2}} = (N_{\text{mod}} * t_{\text{mod}} - t_{\text{low}}) * S_{\text{low\_2}} + \varphi_{\text{init\_2}} \tag{9}$$

$$\varphi_{\text{end\_1}} = (N_{\text{mod}} * t_{\text{mod}} - t_0) * S_{\text{high\_1}} + \varphi_{\text{final\_1}} \tag{10}$$

$$\varphi_{\text{end\_2}} = (N_{\text{mod}} * t_{\text{mod}} - t_{\text{high}}) * S_{\text{high\_2}} + \varphi_{\text{final\_2}} \tag{11}$$

Where N_{mod} is the modulation number (rounded up to the nearest integer), ^1t_R is the first dimension retention time, S_{low} are the slopes of the initial modifier fraction, S_{high} are the slopes of the final modifier fraction and φ_{start} and φ_{end} are the start and end conditions of the selected ^{2}D gradient. The algorithm selects which start and end modifiers ($\varphi_{\text{start\_1}}$, $\varphi_{\text{start\_2}}$ or $\varphi_{\text{init\_3}}$ and $\varphi_{\text{end\_1}}$, $\varphi_{\text{end\_2}}$ or $\varphi_{\text{final\_3}}$) are needed depending on the ^1t_R.

3.4. Gradient profile boundaries

It is important that the algorithm calculates gradient profiles that will result in meaningful chromatography. Therefore, it is essential to select proper gradient profile boundaries for the *ga* function. To prevent stationary phase dewetting under fully aqueous solutions [77], the modifier fraction was constrained to values between 0.02 and 1 for both dimensions.

Furthermore, to ensure that analytes elute from the ^{1}D column a minimum final mobile-phase modifier fraction of 0.8 was used, which can be expanded to a final segment with a modifier fraction of 1. Negative gradient slopes generally do not provide suitable conditions and thus, constraints were set to ensure that the mobile phase modifier fraction always ended higher than the starting modifier fraction for each gradient segment ($\varphi_i < \varphi_{i+1}$). Also, to minimize the risk of excessive band broadening due to a sudden increase in modifier fraction (step-gradients [78,79]), a minimum ^{1}D gradient time (^1t_G) of 0.5 min was selected for each gradient segment. Lastly, to minimize unnecessary computations, the sum of gradient times was constrained to be less than the allowed analysis time ($\sum {}^1t_G < t_{\text{max}}$).

In the ^{2}D shifting gradient, the changes in gradient characteristics need to be aligned with the changes in analyte characteristics and this is vital to avoid peak splitting between modulations (*i.e.* eluting at ^{2}D retention times that differ too greatly), thus the boundaries cannot increase too fast. Therefore, the differences between t_0 and t_{low} /t_{high} and between t_{low} /t_{high} and $t_{\text{shift\_end}}$ were not allowed to be less than 10 modulations and the shifting gradient program was limited to end before the maximum allowed analysis time. Furthermore, to avoid negative gradient slopes, the upper boundary was constrained to be higher than the lower boundary ($\varphi_{\text{init}} < \varphi_{\text{final}}$).
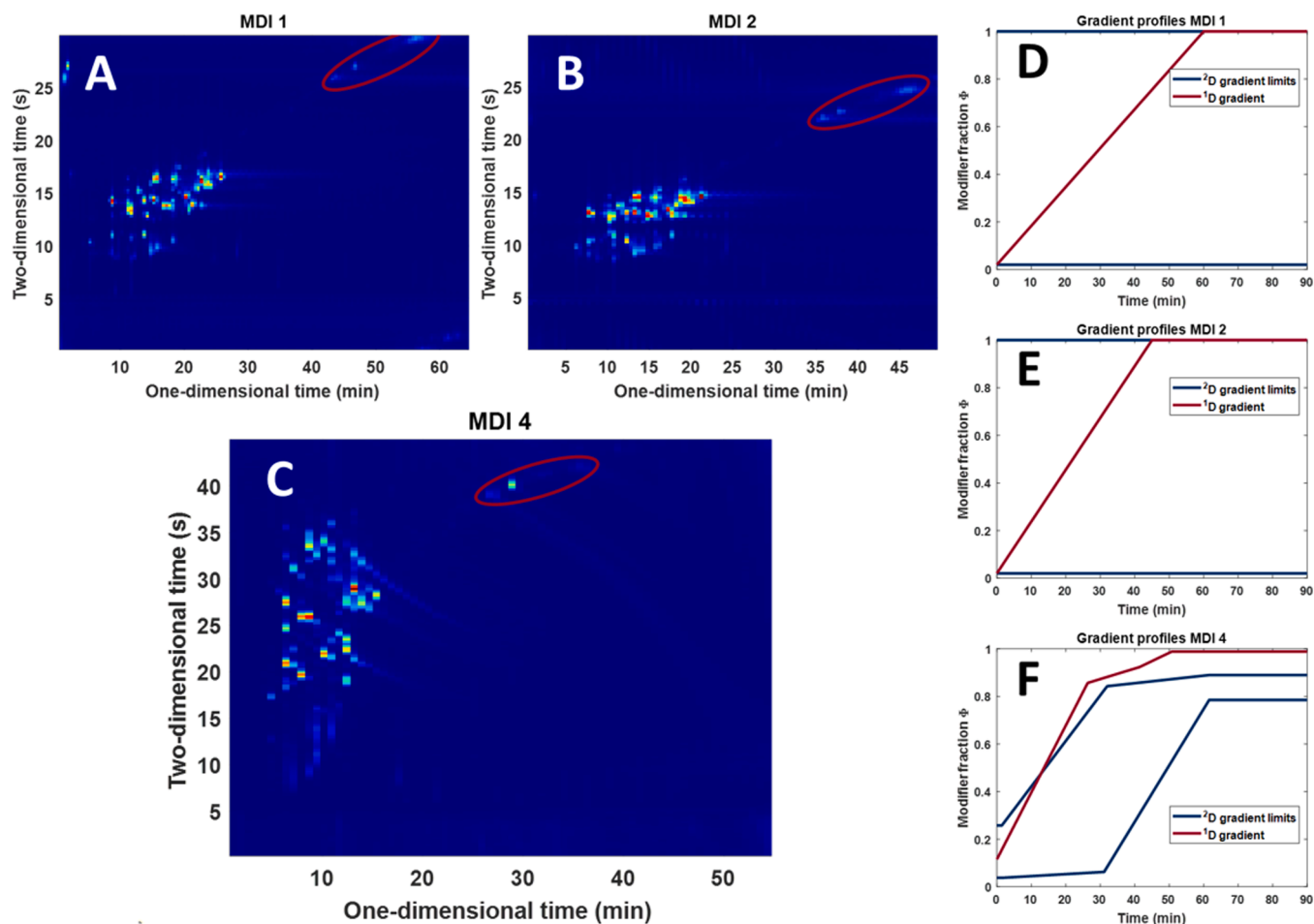

Fig. 5. Chromatograms (A-C) from the first two scanning methods and the first calculated method, and gradient profiles used in each method (D-F). The ^{2}D gradient profiles in D and E are identical, but use different ^{1}D and ^{2}D gradient times. The analytes in the red oval indicate analytes with very steep retention slopes. **A)** LC × LC-MS chromatogram from MDI #1. **B)** LC × LC-MS chromatogram from MDI #2. **C)** LC × LC-MS chromatogram from MDI #4. **D)** Gradient program used in MDI #1; $^2t_G = 24$ s **E)** Gradient program used in MDI #2; $^2t_G = 18$ s. **F)** Gradient program used in MDI #4.

3.5. Computer-driven separation of a monoclonal antibody digest

To test the algorithm, the system was allowed 10 method development iterations (MDI) for the separation of the mAb digest, of which the first three were scanning methods as described in Section 3.2. Fig. 5 shows that the scanning methods did not provide much separation of the peptides (Figs. 5A, B). However, the separation in MDI #4 (Fig. 5C), where the method parameters were calculated based on results from the first two scanning methods, already showed significant improvement in terms of resolution and usage of the separation space. The gradient profiles that were used are shown in Figs. 5D-F. The subsequent methods showed an incremental improvement (see Supplementary Material S-1 Figures S1-S10), however, MDI #5 and #6 used a much wider shifting gradient in the second dimension. Investigation into this phenomenon showed that some noise was detected and tracked in MDI #3. Retention models that were influenced by this noise produced results indicating that there were analytes eluting at a high mobile phase modifier fraction, and this increased the modifier range of the shifting gradient. After MDI #5, these peaks were classified as noise and the algorithm continued to produce narrower shifting gradients.

Furthermore, the optimization of the shifting gradient was hampered by analytes with very steep retention slopes (indicated by the red ovals in Fig. 5A-C). Effectively meaning that these analytes would only elute at high modifier fractions. The shifting gradient therefore needed to end at a high modifier fraction, while the rest of the analytes would elute with a lower modifier fraction. Furthermore, the maximum allowed analysis time was set at 90 min, however, the algorithm often determined analysis times of around 60 min. This meant that the benefit of an increase in average resolution did not outweigh the time cost associated with increasing the resolution. An analyst can decide if an increase in resolution is needed and adjust the weight of the time score in Eq. 2 accordingly. Another notable observation is that the algorithm focused more on the ^{2}D separation compared to the ^{1}D separation. This is not surprising since the highest increase in resolution could be gained by adjusting the shifting gradient. However, it was expected that the ^{1}D gradient profile would increase the modifier fraction quickly after the elution of the smaller compounds to elute the analytes with steep retention slopes faster. This was not the case. Providing the *ga* with a larger population of potential solutions could potentially increase the probability of arriving at a more optimal ^{1}D separation. Or allowing the algorithm more ^{1}D gradient segments might provide more flexibility in the ^{1}D gradient profile. This, of course, would increase the number of computations required since the genetic algorithm might require a significant increase in function evaluations with an increased number of variables to arrive at an optimum (due to the curse of dimensionality) [80,81]. However, this might not improve the usage of the separation space, as the shifting gradients are dependent on analyte elution from the first dimension. A larger number of ^{1}D gradient segments could also require more steps within the ^{2}D shifting gradient to fully benefit from the additional ^{1}D separation.

3.6. Accuracy of retention time predictions

The accuracy of the retention time predictions during the computer-driven method optimization strongly influences the success of the method development algorithm. Retention time prediction errors are highly probable due to the difficulty of assessing ^{1}D retention times, gradient deformation in the second dimension due to fast gradients, and the dependencies between the two dimensions. The robustness of retention models was discussed in Section 3.1 in a general sense, but because of the dependence of the two dimensions, prediction errors in the second dimension of LC × LC can be even larger. In the simplest implementation of LC × LC, the ^{2}D gradients do not change, and thus prediction errors in the second dimension are independent of ^{1}D retention times. Using shifting gradients, however, prediction errors in the first dimension cause an analyte to elute during a different

modulation period from what is expected, and thus with different gradient parameters. This causes a larger deviation between predicted and experimental ^{2}D retention times. Retention time predictions in all MDIs were calculated using Eq. 12:

$$\varepsilon_\% = \frac{t_{R,pred} - t_{R,exp}}{t_{R,exp}} * 100\% \tag{12}$$

where $\varepsilon_\%$ is the prediction error in percentage, and $t_{R,pred}$ and $t_{R,exp}$ are the predicted and experimental retention times, respectively. Fig. 6 shows the retention time and peak width prediction errors per MDI. Both the absolute prediction error (to show accuracy) and the non-absolute prediction errors are shown (to show over-/underestimation). Here it can be seen that the means and relative standard deviations of the prediction errors are larger in the second dimension than they are in the first dimension (Fig. 6A). Furthermore, the ^{2}D retention times are typically overestimated (Fig. 6C), but these overestimations decrease when the ^{1}D retention times predictions are more accurate. In addition to the reasons given above, prediction errors can have other causes. For example, the system dwell volumes and column dead volumes were estimated based on specifications (*e.g.,* LC pump series and column dimensions) from the suppliers. Experimental determination of these parameters could increase the accuracy of calculated retention factors and in turn more accurate retention time predictions for method parameters proposed by the algorithm. In this work, we chose to work with estimates to increase ease of use and quickly start the automated method development. However, with more method iterations, the retention time prediction errors became smaller. This was expected as more reliable retention parameters can be estimated when more data points are available. Histograms of the spread of retention time prediction errors for both dimensions are shown in Supplementary Material Section S-2, Fig. S-11.

The errors in the prediction of peak widths were larger than errors in the prediction of retention time, as shown in Fig. 6B, and typically overestimated (Fig. 6D). This could similarly be due to all the above-discussed reasons, or for example, mass overload or extra-column dispersion. Additionally, small retention time deviations in the first dimension can lead to compounds eluting in a ^{2}D gradient that is different from what is expected based on the predictions of the algorithm. The speed of elution, and thus the peak width, is dependent on the retention factor at elution. Thus, an accurate prediction of the modifier fraction at elution and the assessment of the retention models are crucial. It has been shown that calculated compression factors (*i.e.,* a ratio of the peak width under gradient conditions compared to isocratic elution) differ from the experimental values up to a ratio of 1.8 [78]. However, Vaast *et al.* concluded this difference is due to the incorrect determination of the retention factor at elution using linear retention models (*e.g.,* the LSS model). They stated that using non-linear retention models, such as the quadratic model or the Neue-Kuss model, can greatly improve the prediction of the compression factors [82] and thus peak widths. At this moment it is unclear which of these effects contributes to the observed deviations and further investigations will be required.

4. Conclusions

Computer-driven optimization of complex gradients in LC × LC was successfully demonstrated. The current workflow and all code are available in the Supplementary Materials. Using only two scanning gradients, a large improvement in resolution and usage of the available separation space was already realized using the first method recommended by the algorithm by using so-called shifting gradients. Subsequent MDIs showed an improvement in the accuracy of the predicted chromatograms. However, larger errors in the prediction of retention time and peak width were observed in the second dimension compared to the ^{1}D prediction errors. Retention time deviations in the first dimension can lead to a difference between the ^{2}D conditions used for
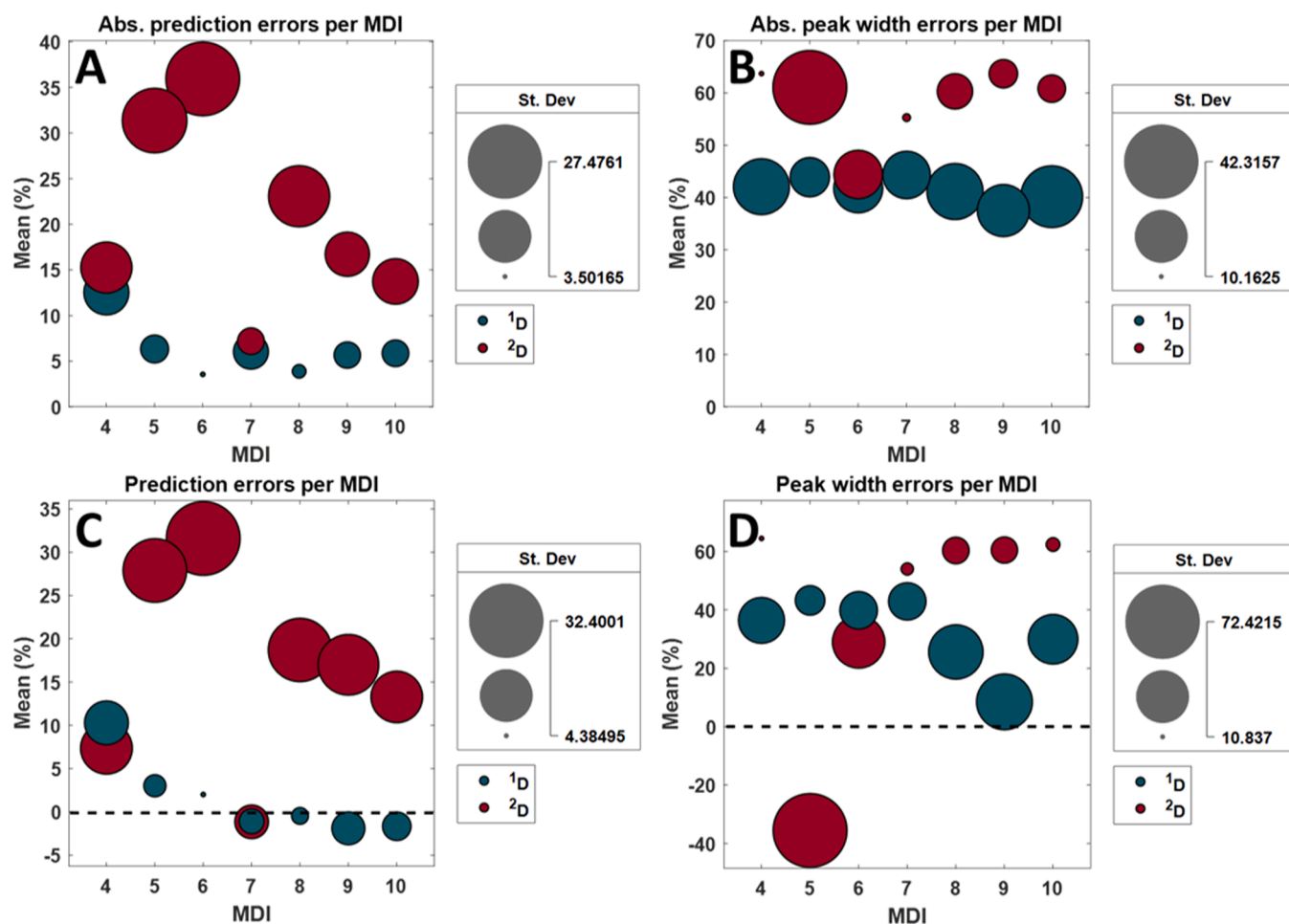
Fig. 6. Errors in the prediction of retention time and peak width for each MDI. Mean ^{1}D prediction errors are shown in green, and mean ^{2}D prediction errors in red. The size of the bubble corresponds to the relative standard deviation of the prediction errors. (Larger bubbles mean a larger spread in prediction errors) **A)** Absolute retention time prediction errors. **B)** Absolute peak width prediction errors. **C)** Retention time prediction errors. **D)** Peak width prediction errors.

prediction and those experienced by analytes during an experimental separation, leading to larger prediction errors for the second dimension. Furthermore, due to fast ^{2}D gradients, the programmed gradient may deform [61], also affecting the accuracy of predictions. Monitoring the actual gradient shape and using retention modeling on the actual gradient shape is expected to produce more accurate results. This, however, requires additional computational resources as the (now single) gradients need to be modeled into multiple small linear gradients to approximate the actual shape. Furthermore, the interaction by the analyst with the performance of the algorithm was limited. System and column constants, such as the dwell volumes and column dead volumes, were estimated in this work. Accurate determination of these constants is expected to improve retention modeling and thus performance.

Peak widths were generally overestimated in predictions made by the algorithm. The above-mentioned reasons for the retention time prediction errors are also likely to affect peak width predictions. Nonetheless, it has been shown that theoretical values for compression factors typically differ greatly from the experimental values [78]. Multiple causes for this phenomenon have been provided in literature, for example, extra-column volume [83], variation in plate height with mobile phase composition [84], and viscous fingering [79]. More recently, it was concluded that the incorrect determination of the retention factor at elution by using linear retention models is likely the cause for this [82]. Thus it is expected that by using a non-linear retention model the peak width predictions will improve. However, in this work, it has been shown that the non-linear quadratic retention model is not as robust as the LSS model for difficult-to-model analytes. In

the current work, the peak widths are overestimated and thus the algorithm underestimates the actual performance of the LC × LC separation (*i.e.*, broader peaks equal lower resolution). It could be argued though, that this is preferred over the other way around. It would be worse if the algorithm predicts a "good" separation, but the separation itself is of bad quality, although this would impede the total analysis time slightly. Nevertheless, it can be debatable what is preferred and other retention models can be incorporated within the current workflow. Hao *et al.* have derived the gradient compression equations for the quadratic model [24], therefore these can be incorporated if preferred.

The algorithm mainly focused on the ^{2}D separation. This is not surprising as the largest increase in resolution was expected to be obtained by the shifting gradient. Still, there was room for improvement in the first dimension. An analyst would see that a slow gradient at the start of the separation would result in a better separation for the early eluting analytes. Following this up with a steep gradient would result in the elution of the late eluting analytes and reduce the total analysis time. A larger population size for the *ga* or more gradient segments might aid the ga with the improvement of the ^{1}D separation. This, however, will come at a greater computational cost. The current *ga* was trained with 16 variables, of which nine belonged to the shifting gradient program, two parameters per ^{1}D gradient (*i.e.*, gradient time and final modifier), and the initial modifier fraction of the first dimension. Adding more ^{1}D gradient segments will increase the total variables by two per gradient segment, potentially requiring a significant increase in function evaluations for the number of variables that need to be determined. Increasing the population size/number of generations in the genetic algorithm will

most likely result in longer calculation times than measurement times. This means the LC × LC-MS system will have to wait for new instructions from the algorithm before continuing its measurements. The current workflow minimized the total optimization time by allowing computations while a separation was being performed. This decreased the total needed time by approximately half. It is debatable whether it would be worth the extra computational investment (and thus let the LC × LC-MS system wait) for an improvement in [1]D separation.

The current work focused on maximizing the resolution of each analyte with respect to its nearest neighbor, while simultaneously minimizing the total analysis time using pre-defined weights in the objective function. However, studying other objective functions could be very fruitful. While many so-called chromatographic response functions (CRF), or chromatographic objective functions, exist for 1D chromatography [85], not so many CRFs exist for 2D chromatography [86–89]. Most of these CRFs use the number of experimentally detected peaks as a quality descriptor. However, when using retention modeling to guide the optimization process, the number of peaks that are modeled is known. This does not mean that all experimental peaks are indeed modeled, but the number of detected peaks is hence not a suitable quality descriptor for the genetic algorithm. Using the current workflow, it would be highly interesting to study the effect of different quality descriptors and their performance on 2D separations. Likewise, it would be interesting to study the effect of using different weights for both the resolution and analysis time terms, as these might impact the found optima and might be dependent on the nature of the studied sample. Here, Pareto front-based multi-objective optimization strategies might also be interesting venues for future research [25,90].

CRediT authorship contribution statement

Stef R.A. Molenaar: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Tijmen S. Bos:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – review & editing, Visualization. **Jim Boelrijk:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – review & editing, Visualization. **Tina A. Dahlseid:** Methodology, Validation, Investigation, Writing – review & editing. **Dwight R. Stoll:** Methodology, Validation, Resources, Writing – review & editing, Project administration, Funding acquisition. **Bob W.J. Pirok:** Conceptualization, Methodology, Resources, Data curation, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2023.464306.

References

[1] F. Erni, R.W. Frei, Two-dimensional column liquid chromatographic technique for resolution of complex mixtures, J Chromatogr A 149 (1978) 561–569, https://doi.org/10.1016/S0021-9673(00)81011-0.

[2] P. Jandera, T. Hájek, P. Česla, V. Škeříková, Advantages of two-dimensional liquid chromatography in the analysis of complex samples, Chemija 22 (2011) 149–154.

[3] A.S. Cohen, R.M. Schure, Multidimensional Liquid Chromatography: Theory and Applications in Industrial Chemistry and the Life Sciences, Wiley & Sons, New York, 2008.

[4] X. Li, D.R. Stoll, P.W. Carr, Equation for peak capacity estimation in two-dimensional liquid chromatography, Anal Chem 81 (2009) 845–850, https://doi.org/10.1021/ac801772u.

[5] A. van der Horst, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography of polymers, J Chromatogr A 1000 (2003) 693–709, https://doi.org/10.1016/S0021-9673(03)00495-3.

[6] R. Xiang, Y. Shi, D.A. Dillon, B. Negin, C. Horváth, J.A. Wilkins, 2D LC/MS analysis of membrane proteins from breast cancer cell lines MCF7 and BT474, J Proteome Res 3 (2004) 1278–1283, https://doi.org/10.1021/pr049852e.

[7] E. Nägele, M. Vollmer, P. Hörth, C. Vad, 2D-LC/MS techniques for the identification of proteins in highly complex mixtures, Expert Rev Proteomics 1 (2004) 37–46, https://doi.org/10.1586/14789450.1.1.37.

[8] H. Nie, R. Liu, Y. Yang, Y. Bai, Y. Guan, D. Qian, T. Wang, H. Liu, Lipid profiling of rat peritoneal surface layers by online normal- and reversed-phase 2D LC QToF-MS, J Lipid Res 51 (2010) 2833–2844, https://doi.org/10.1194/jlr.D007567.

[9] F.T. van Beek, R. Edam, B.W.J. Pirok, W.J.L. Genuit, P.J. Schoenmakers, Comprehensive two-dimensional liquid chromatography of heavy oil, J Chromatogr A 1564 (2018) 110–119, https://doi.org/10.1016/j.chroma.2018.06.001.

[10] F. Cacciola, F. Rigano, P. Dugo, L. Mondello, Comprehensive two-dimensional liquid chromatography as a powerful tool for the analysis of food and food products, TrAC - Trends in Analytical Chemistry 127 (2020), https://doi.org/10.1016/j.trac.2020.115894.

[11] F. Cacciola, P. Donato, D. Sciarrone, P. Dugo, L. Mondello, Comprehensive liquid chromatography and other liquid-based comprehensive techniques coupled to mass spectrometry in food analysis, Anal Chem 89 (2017) 414–429, https://doi.org/10.1021/acs.analchem.6b04370.

[12] B.W.J. Pirok, D.R. Stoll, P.J. Schoenmakers, Recent Developments in Two-Dimensional Liquid Chromatography: Fundamental Improvements for Practical Applications, Anal Chem 91 (2019) 240–263, https://doi.org/10.1021/acs.analchem.8b04841.

[13] R.S. van den Hurk, M. Pursch, D.R. Stoll, B.W.J. Pirok, Recent trends in two-dimensional liquid chromatography, TrAC Trends in Analytical Chemistry (2023), 117166, https://doi.org/10.1016/j.trac.2023.117166.

[14] B.W.J. Pirok, A.F.G. Gargano, P.J. Schoenmakers, Optimizing separations in online comprehensive two-dimensional liquid chromatography, J Sep Sci 41 (2018) 68–98, https://doi.org/10.1002/jssc.201700863.

[15] B.W.J. Pirok, P.J. Schoenmakers, Practical approaches to overcome the challenges of comprehensive two-dimensional liquid chromatography, LC-GC Europe 31 (2018) 242–249.

[16] P. Jandera, P. Česla, T. Hájek, G. Vohralík, K. Vyňuchalová, J. Fischer, Optimization of separation in two-dimensional high-performance liquid chromatography by adjusting phase system selectivity and using programmed elution techniques, J Chromatogr A 1189 (2008) 207–220, https://doi.org/10.1016/j.chroma.2007.11.053.

[17] P. Česla, T. Hájek, P. Jandera, Optimization of two-dimensional gradient liquid chromatography separations, J Chromatogr A 1216 (2009) 3443–3457, https://doi.org/10.1016/j.chroma.2008.08.111.

[18] G.B. van Henten, T.S. Bos, B.W.J. Pirok, Approaches to Accelerate Liquid Chromatography Method Development in the Laboratory Using Chemometrics and Machine Learning, LCGC Europe 36 (2023) 202–209, https://doi.org/10.56530/lcgc.eu.rh7676j5.

[19] S. Agatonovic-Kustrin, M. Zecevic, L. Zivanovic, I.G. Tucker, Application of neural networks for response surface modeling in HPLC optimization, Anal Chim Acta 364 (1998) 265–273, https://doi.org/10.1016/S0003-2670(98)00121-4.

[20] M.C. García-Alvarez-Coque, J.R. Torres-Lapasió, J.J. Baeza-Baeza, Modelling of retention behaviour of solutes in micellar liquid chromatography, J Chromatogr A 780 (1997) 129–148, https://doi.org/10.1016/S0021-9673(97)00051-4.

[21] E. Marengo, V. Gianotti, S. Angioi, M.C. Gennaro, Optimization by experimental design and artificial neural networks of the ion-interaction reversed-phase liquid chromatographic separation of twenty cosmetic preservatives, J Chromatogr A 1029 (2004) 57–65, https://doi.org/10.1016/j.chroma.2003.12.044.

[22] H.J. Metting, P.M.J. Coenegracht, Neural networks in high-performance liquid chromatography optimization: Response surface modeling, J Chromatogr A 728 (1996) 47–53, https://doi.org/10.1016/0021-9673(96)82447-2.

[23] A. Malenović, B. Jancic-Stojanovic, N. Kostić, D. Ivanović, M. Medenica, Optimization of artificial neural networks for modeling of atorvastatin and its impurities retention in micellar liquid chromatography, Chromatographia 73 (2011) 993–998, https://doi.org/10.1007/s10337-011-1994-6.

[24] W. Hao, B. Li, Y. Deng, Q. Chen, L. Liu, Q. Shen, Computer aided optimization of multilinear gradient elution in liquid chromatography, J Chromatogr A 1635 (2021), https://doi.org/10.1016/j.chroma.2020.461754.

[25] J. Boelrijk, B. Ensing, P. Forré, B.W.J. Pirok, Closed-loop automatic gradient design for liquid chromatography using Bayesian optimization, Anal Chim Acta 1242 (2023), 340789, https://doi.org/10.1016/j.aca.2023.340789.

[26] B. Huygens, K. Efthymiadis, A. Nowé, G. Desmet, Application of evolutionary algorithms to optimise one- and two-dimensional gradient chromatographic separations, J Chromatogr A 1628 (2020), 461435, https://doi.org/10.1016/j.chroma.2020.461435.

[27] J. Boelrijk, B. Pirok, B. Ensing, P. Forré, Bayesian optimization of comprehensive two-dimensional liquid chromatography separations, J Chromatogr A 1659 (2021), 462628, https://doi.org/10.1016/j.chroma.2021.462628.

[28] S. O'Hagan, W.B. Dunn, M. Brown, J.D. Knowles, D.B. Kell, Closed-Loop, Multiobjective Optimization of Analytical Instrumentation: Gas Chromatography/ Time-of-Flight Mass Spectrometry of the Metabolomes of Human Serum and of Yeast Fermentations, Anal Chem 77 (2005) 290–303, https://doi.org/10.1021/ac049146x.

[29] 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, https://doi.org/10.1016/S0021-9673(00)81008-0.

[30] U.D. Neue, Nonlinear retention relationships in reversed-phase chromatography, Chromatographia 63 (2006) 45–53, https://doi.org/10.1365/s10337-006-0718-9.

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

[32] P. Jandera, M. Holčapek, L. Kolářová, Retention mechanism, isocratic and gradient-elution separation and characterization of (co)polymers in normal-phase and reversed-phase high-performance liquid chromatography, J Chromatogr A 869 (2000) 65–84, https://doi.org/10.1016/S0021-9673(99)01216-9.

[33] C.M. Roth, K.K. Unger, A.M. Lenhoff, Mechanistic model of retention in protein ion-exchange chromatography, J Chromatogr A 726 (1996) 45–56, https://doi.org/10.1016/0021-9673(95)01043-2.

[34] A.E. Karatapanis, Y.C. Fiamegos, C.D. Stalikas, A revisit to the retention mechanism of hydrophilic interaction liquid chromatography using model organic compounds, J Chromatogr A 1218 (2011) 2871–2879, https://doi.org/10.1016/j.chroma.2011.02.069.

[35] B.W.J. Pirok, S.R.A. Molenaar, R.E. van Outersterp, P.J. Schoenmakers, Applicability of retention modelling in hydrophilic-interaction liquid chromatography for algorithmic optimization programs with gradient-scanning techniques, J Chromatogr A 1530 (2017) 104–111, https://doi.org/10.1016/j.chroma.2017.11.017.

[36] E. Tyteca, A. Périat, S. Rudaz, G. Desmet, D. Guillarme, Retention modeling and method development in hydrophilic interaction chromatography, J Chromatogr A 1337 (2014) 116–127, https://doi.org/10.1016/j.chroma.2014.02.032.

[37] G. Jin, Z. Guo, F. Zhang, X. Xue, Y. Jin, X. Liang, Study on the retention equation in hydrophilic interaction liquid chromatography, Talanta 76 (2008) 522–527, https://doi.org/10.1016/j.talanta.2008.03.042.

[38] E. Tyteca, V. Desfontaine, G. Desmet, D. Guillarme, Possibilities of retention modeling and computer assisted method development in supercritical fluid chromatography, J Chromatogr A 1381 (2015) 219–228, https://doi.org/10.1016/j.chroma.2014.12.077.

[39] C.R. Yonker, R.D. Smith, Effect of the partial molar volume of the solute in the stationary phase on retention in supercritical fluid chromatography, J Chromatogr A 459 (1988) 183–191, https://doi.org/10.1016/S0021-9673(01)82026-4.

[40] D.R. Luffer, W. Ecknig, M. Novotny, Physicochemical model of retention for capillary supercritical fluid chromatography, J Chromatogr 505 (1990) 79–97, https://doi.org/10.1016/S0021-9673(01)93069-9.

[41] S.R.A. Molenaar, M.V. Savova, R. Cross, P.D. Ferguson, P.J. Schoenmakers, B.W.J. Pirok, Improving retention-time prediction in supercritical-fluid chromatography by multivariate modelling, J Chromatogr A 1668 (2022), 462909, https://doi.org/10.1016/j.chroma.2022.462909.

[42] I. Groeneveld, B.W.J. Pirok, S.R.A. Molenaar, P.J. Schoenmakers, M.R. van Bommel, The development of a generic analysis method for natural and synthetic dyes by ultra-high-pressure liquid chromatography with photo-diode-array detection and triethylamine as an ion-pairing agent, J Chromatogr A 1673 (2022), 463038, https://doi.org/10.1016/j.chroma.2022.463038.

[43] G. van Schaick, B.W.J. Pirok, R. Haselberg, G.W. Somsen, A.F.G. Gargano, Computer-aided gradient optimization of hydrophilic interaction liquid chromatographic separations of intact proteins and protein glycoforms, J Chromatogr A 1598 (2019) 67–76, https://doi.org/10.1016/j.chroma.2019.03.038.

[44] J.W. Dolan, D.C. Lommen, L.R. Snyder, Drylab® computer simulation for high-performance liquid chromatographic method development. II. Gradient Elution, J Chromatogr 485 (1989) 91–112, https://doi.org/10.1016/S0021-9673(01)89134-2.

[45] L.R. Snyder, J.W. Dolan, D.C. Lommen, Drylab® computer simulation for high-performance liquid chromatographic method development. I. Isocratic elution, J Chromatogr 485 (1989) 65–89, https://doi.org/10.1016/S0021-9673(01)89133-0.

[46] B.W.J. Pirok, S. Pous-Torres, C. Ortiz-Bolsico, G. Vivó-Truyols, P.J. Schoenmakers, Program for the interpretive optimization of two-dimensional resolution, J Chromatogr A 1450 (2016) 29–37, https://doi.org/10.1016/j.chroma.2016.04.061.

[47] S.R.A. Molenaar, P.J. Schoenmakers, B.W.J. Pirok, Multivariate Optimization and Refinement Program for Efficient Analysis of Key Separations (MOREPEAKS), (2021). https://doi.org/10.5281/zenodo.5710442.

[48] T.S. Bos, W.C. Knol, S.R.A. Molenaar, L.E. Niezen, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Recent applications of chemometrics in one- and two-dimensional chromatography, J Sep Sci 43 (2020) 1678–1727, https://doi.org/10.1002/jssc.202000011.

[49] K.M. Åberg, R.J.O. Torgrip, J. Kolmert, I. Schuppe-Koistinen, J. Lindberg, Feature detection and alignment of hyphenated chromatographic-mass spectrometric data. Extraction of pure ion chromatograms using Kalman tracking, J Chromatogr A 1192 (2008) 139–146, https://doi.org/10.1016/j.chroma.2008.03.033.

[50] A.J. Round, M.I. Aguilar, M.T.W. Hearn, High-performance liquid chromatography of amino acids, peptides and proteins. CXXXIII. Peak tracking of peptides in reversed-phase high-performance liquid chromatography, J Chromatogr A 661 (1994) 61–75, https://doi.org/10.1016/0021-9673(93)E0874-T.

[51] A. Bogomolov, M. McBrien, Mutual peak matching in a series of HPLC-DAD mixture analyses, Anal Chim Acta 490 (2003) 41–58, https://doi.org/10.1016/S0003-2670(03)00667-6.

[52] M.J. Fredriksson, P. Petersson, B.O. Axelsson, D. Bylund, Combined use of algorithms for peak picking, peak tracking and retention modelling to optimize the chromatographic conditions for liquid chromatography-mass spectrometry analysis of fluocinolone acetonide and its degradation products, Anal Chim Acta 704 (2011) 180–188, https://doi.org/10.1016/j.aca.2011.07.047.

[53] B.W.J. Pirok, S.R.A. Molenaar, L.S. Roca, P.J. Schoenmakers, Peak-Tracking Algorithm for Use in Automated Interpretive Method-Development Tools in Liquid Chromatography, Anal Chem 90 (2018) 14011–14019, https://doi.org/10.1021/acs.analchem.8b03929.

[54] K.J. Johnson, B.W. Wright, K.H. Jarman, R.E. Synovec, High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis, J Chromatogr A 996 (2003) 141–155, https://doi.org/10.1016/S0021-9673(03)00616-2.

[55] A. Barcaru, E. Derks, G. Vivó-Truyols, Bayesian peak tracking: A novel probabilistic approach to match GCxGC chromatograms, Anal Chim Acta 940 (2016) 46–55, https://doi.org/10.1016/j.aca.2016.09.001.

[56] S.R.A. Molenaar, T.A. Dahlseid, G.M. Leme, D.R. Stoll, P.J. Schoenmakers, B.W.J. Pirok, Peak-tracking algorithm for use in comprehensive two-dimensional liquid chromatography – Application to monoclonal-antibody peptides, J Chromatogr A 1639 (2021), 461922, https://doi.org/10.1016/j.chroma.2021.461922.

[57] K.M. Pierce, L.F. Wood, B.W. Wright, R.E. Synovec, A comprehensive two-dimensional retention time alignment algorithm to enhance chemometric analysis of comprehensive two-dimensional separation data, Anal Chem 77 (2005) 7735–7743, https://doi.org/10.1021/ac0511142.

[58] S.R.A. Molenaar, J.H.M. Mommers, D.R. Stoll, S. Ngxangxa, A.J. De Villiers, P.J. Schoenmakers, B.W.J. Pirok, Algorithm for tracking peaks amongst numerous datasets in comprehensive two-dimensional chromatography to enhance data analysis and interpretation, J Chromatogr A. Accepted (2023).

[59] T.S. Bos, J. Boelrijk, S.R.A. Molenaar, B. van 't Veer, L.E. Niezen, D. van Herwerden, S. Samanipour, D.R. Stoll, P. Forré, B. Ensing, G.W. Somsen, B.W.J. Pirok, Chemometric Strategies for Fully Automated Interpretive Method Development in Liquid Chromatography, Anal Chem 94 (2022) 16060–16068, https://doi.org/10.1021/acs.analchem.2c03160.

[60] D. Li, C. Jakob, O. Schmitz, Practical considerations in comprehensive two-dimensional liquid chromatography systems (LCxLC) with reversed-phases in both dimensions, Anal Bioanal Chem 407 (2015) 153–167, https://doi.org/10.1007/s00216-014-8179-8.

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

[62] L.E. Niezen, T.S. Bos, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Capacitively coupled contactless conductivity detection to account for system-induced gradient deformation in liquid chromatography, Anal Chim Acta (2023) 1271, https://doi.org/10.1016/j.aca.2023.341466.

[63] D.R. Stoll, H.R. Lhotka, D.C. Harmes, B. Madigan, J.J. Hsiao, G.O. Staples, High resolution two-dimensional liquid chromatography coupled with mass spectrometry for robust and sensitive characterization of therapeutic antibodies at the peptide level, J Chromatogr B Analyt Technol Biomed Life Sci 1134–1135 (2019), 121832, https://doi.org/10.1016/j.jchromb.2019.121832.

[64] D.R. Stoll, K. Shoykhet, P. Petersson, S. Buckenmaier, Active Solvent Modulation: A Valve-Based Approach to Improve Separation Compatibility in Two-Dimensional Liquid Chromatography, Anal Chem 89 (2017) 9260–9267, https://doi.org/10.1021/acs.analchem.7b02046.

[65] M.C. Chambers, B. MacLean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman,

N. Shulman, B. Frewen, T.A. Baker, M.Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R. L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, Nat Biotechnol 30 (2012) 918–920, https://doi.org/10.1038/nbt.2377.

[66] M.A. Quarry, R.L. Grob, L.R. Snyder, Prediction of Precise Isocratic Retention Data from Two or More Gradient Elution Runs. Analysis of Some Associated Errors, Anal Chem 58 (1986) 907–917, https://doi.org/10.1021/ac00295a056.

[67] G. Vivó-Truyols, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Error analysis and performance of different retention models in the transference of data from/to isocratic/gradient elution, J Chromatogr A 1018 (2003) 169–181, https://doi.org/10.1016/j.chroma.2003.08.044.

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

[69] P. Nikitas, A. Pappa-Louisi, A. Papageorgiou, Simple algorithms for fitting and optimisation for multilinear gradient elution in reversed-phase liquid chromatography, J Chromatogr A 1157 (2007) 178–186, https://doi.org/10.1016/j.chroma.2007.04.059.

[70] T. Brau, B. Pirok, S. Rutan, D. Stoll, Accuracy of retention model parameters obtained from retention data in liquid chromatography, J Sep Sci 45 (2022) 3241–3255, https://doi.org/10.1002/jssc.202100911.

[71] Elimelech. Grushka, M.N. Myers, P.D. Schettler, J.Calvin. Giddings, Computer characterization of chromatographic peaks by plate height and higher central moments, Anal Chem 41 (1969) 889–892, https://doi.org/10.1021/ac60276a014.

[72] L.R. Snyder, J.W. Dolan, J.R. Gant, Gradient elution in high-performance liquid chromatography: I. Theoretical basis for reversed-phase systems, J Chromatogr 165 (1979) 3–30, https://doi.org/10.1016/S0021-9673(00)85726-X.

[73] M.R. Schure, Quantification of Resolution for Two-Dimensional Separations, Journal of Microcolumn Separations 9 (1997) 169–176, https://doi.org/10.1002/(SICI)1520-667X(1997)9:3<169::AID-MCS5>3.0.CO;2-%23.

[74] D.R. Stoll, G.M. Leme, Instrumentation for Two-Dimensional Liquid Chromatography, in: D.R. Stoll, P.W. Carr (Eds.), Multi-Dimensional Liquid Chromatography, 1st ed., CRC Press, Boca Raton, 2023, pp. 115–165, https://doi.org/10.1201/9781003090557.

[75] F. Bedani, W.T. Kok, H.G. Janssen, Optimal gradient operation in comprehensive liquid chromatography × liquid chromatography systems with limited orthogonality, Anal Chim Acta 654 (2009) 77–84, https://doi.org/10.1016/j.aca.2009.06.042.

[76] S. Chapel, F. Rouvière, S. Heinisch, Sense and nonsense of shifting gradients in on-line comprehensive reversed-phase LC × reversed-phase LC, Journal of Chromatography B 1212 (2022), 123512, https://doi.org/10.1016/j.jchromb.2022.123512.

[77] M. Przybyciel, R.E. Majors, Phase Collapse in Reversed-Phase LC, LC GC Eur 15 (2002) 652–657.

[78] U.D. Neue, D.H. Marchand, L.R. Snyder, Peak compression in reversed-phase gradient elution, J Chromatogr A 1111 (2006) 32–39, https://doi.org/10.1016/j.chroma.2006.01.104.

[79] G.M. Homsy, Viscous Fingering in Porous Media, Annu Rev Fluid Mech 19 (1987) 271–311, https://doi.org/10.1146/annurev.fl.19.010187.001415.

[80] R.E. Bellman, Adaptive Control Processes: A Guided Tour, Princeton University Press, Princeton, 1961.

[81] C.M. Bishop, Pattern Recognition and Machine Learning, Springer New York, New York, 2006.

[82] A. Vaast, E. Tyteca, G. Desmet, P.J. Schoenmakers, S. Eeltink, Gradient-elution parameters in capillary liquid chromatography for high-speed separations of peptides and intact proteins, J Chromatogr A 1355 (2014) 149–157, https://doi.org/10.1016/j.chroma.2014.06.010.

[83] J.D. Stuart, D.D. Lisi, L.R. Snyder, Separation of mixtures of o-phthalaldehyde-derivatized amino acids by reversed-phase gradient elution, J Chromatogr A 485 (1989) 657–672, https://doi.org/10.1016/S0021-9673(01)89171-8.

[84] R.W. Stout, J.J. DeStefano, L.R. Snyder, High-performance liquid chromatographic column efficiency as a function of particle composition and geometry and capacity factor, J Chromatogr A 282 (1983) 263–286, https://doi.org/10.1016/S0021-9673(00)91607-8.

[85] J.T.V. Matos, R.M.B.O. Duarte, A.C. Duarte, Chromatographic response functions in 1D and 2D chromatography as tools for assessing chemical complexity, TrAC Trends in Analytical Chemistry 45 (2013) 14–23, https://doi.org/10.1016/j.trac.2012.12.013.

[86] R.M.B.O. Duarte, J.T.V Matos, A.C. Duarte, A new chromatographic response function for assessing the separation quality in comprehensive two-dimensional liquid chromatography, J Chromatogr A 1225 (2012) 121–131, https://doi.org/10.1016/j.chroma.2011.12.082.

[87] J.T.V. Matos, R.M.B.O. Duarte, A.C. Duarte, A generalization of a chromatographic response function for application in non-target one- and two-dimensional chromatography of complex samples, J Chromatogr A 1263 (2012) 141–150, https://doi.org/10.1016/j.chroma.2012.09.037.

[88] W. Nowik, S. Héron, M. Bonose, M. Nowik, A. Tchapla, Assessment of Two-Dimensional Separative Systems Using Nearest-Neighbor Distances Approach. Part 1: Orthogonality Aspects, Anal Chem 85 (2013) 9449–9458, https://doi.org/10.1021/ac4012705.

[89] W. Nowik, M. Bonose, S. Héron, M. Nowik, A. Tchapla, Assessment of two-dimensional separative systems using the nearest neighbor distances approach. Part 2: Separation quality aspects, Anal Chem 85 (2013) 9459–9468, https://doi.org/10.1021/ac4012717.

[90] J. Boelrijk, B. Ensing, P. Forré, Multi-objective optimization via equivariant deep hypervolume approximation, in: The Eleventh International Conference on Learning Representations, 2023.