Algorithm for tracking peaks amongst numerous datasets in comprehensive two-dimensional chromatography to enhance data analysis and interpretation

Molenaar, S.R.A.; Mommers, J.H.M.; Stoll, D.R.; Ngxangxa, S.; de Villiers, A.J.; Schoenmakers, P.J.; Pirok, B.W.J.

DOI
10.1016/j.chroma.2023.464223

Publication date
2023

Document Version
Final published version

Published in
Journal of Chromatography A

Journal of Chromatography A 1705 (2023) 464223

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Algorithm for tracking peaks amongst numerous datasets in comprehensive two-dimensional chromatography to enhance data analysis and interpretation

Stef R.A. Molenaar [a,b,*], John H.M. Mommers [c], Dwight R. Stoll [d], Sithandile Ngxangxa [e], André J. de Villiers [e], Peter J. Schoenmakers [a,b], Bob W.J. Pirok [a,b]

[a] Analytical Chemistry Group, van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
[b] Centre for Analytical Sciences Amsterdam (CASA), The Netherlands
[c] DSM Engineering Materials, Geleen, The Netherlands
[d] Department of Chemistry, Gustavus Adolphus College, Saint Peter, MN 56082, United States
[e] Department of Chemistry and Polymer Science, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa

ARTICLE INFO

Keywords:
Peak tracking, LC×LC-MS
Method optimization
Data interpretation

ABSTRACT

Analytical data processing often requires the comparison of data, *i.e.* finding similarities and differences within separations. In this context, a peak-tracking algorithm was developed to compare multiple datasets in one-dimensional (1D) and two-dimensional (2D) chromatography. Two application strategies were investigated: *i*) data processing where all chromatograms are produced in one sequence and processed simultaneously, and *ii*) method optimization where chromatograms are produced and processed cumulatively. The first strategy was tested on data from comprehensive 2D liquid chromatography and comprehensive 2D gas chromatography separations of academic and industrial samples of varying compound classes (monoclonal-antibody digest, wine volatiles, polymer granulate headspace, and mayonnaise). Peaks were tracked in up to 29 chromatograms at once, but this could be upscaled when necessary. However, the peak-tracking algorithm performed less accurate for trace analytes, since, peaks that are difficult to detect are also difficult to track. The second strategy was tested with 1D liquid chromatography separations, that were optimized using automated method-development. The strategy for method optimization was quicker to detect peaks that were still poorly separated in earlier chromatograms compared to assigning a target chromatogram, to which all other chromatograms are compared. Rendering it a useful tool for automated method optimization.

1. Introduction

Data processing in chromatography and especially comprehensive two-dimensional (2D) chromatography is a complex and growing field [1]. Before any useful information can be extracted from a chromatogram, the raw data needs to be pre-processed. This typically starts with noise-removal and background correction. Several pre-processing methods have been published over the years and, depending on the needs of the data analyst (*e.g.* speed or quantification) and peak coverage, performances vary greatly [2]. After pre-processing, peaks can be extracted from the chromatogram. Any errors during this extraction translate into a decrease in quality of the information that can impact routine analysis [3,4], investigating unknown samples in various fields (*e.g.* forensics [5,6] or protein analysis [7]) and even optimization [8,9]. Therefore, peak-detection algorithms have been developed for both one-dimensional (1D) [10–12] and 2D [13–15] chromatography.

Ultimately, when multiple chromatograms are compared (*e.g.* sample comparison), knowing which peaks actually belong to the same analyte, across the chromatograms, is required. This is often referred to as peak tracking or peak alignment [16] and becomes more challenging as the sample complexity and the degree of peak coelution increases. Moreover, peaks can be shifted due to chromatographic effects, such as column aging, mobile-phase impurities, or an inconsistent mobile phase flow. The group of Synovec is particularly active in this field, and have developed alignment strategies for both gas chromatography (GC) and comprehensive 2DGC (GC×GC) [17–19].

* Corresponding author: Science Park 904, 1098 XH, Amsterdam, The Netherlands.
E-mail address: s.r.a.molenaar@uva.nl (S.R.A. Molenaar).

https://doi.org/10.1016/j.chroma.2023.464223
Received 14 March 2023; Received in revised form 6 July 2023; Accepted 18 July 2023
Available online 20 July 2023

Peak tracking can also be useful for cases where the method is altered, but the sample is the same. A typical example is retention modeling, where analytes are studied under various conditions to use the retention data for the construction of retention models. Here, the mobile phase composition or temperature program can be intentionally changed [20–23]. Small retention time deviations due to repeatability issues can often be corrected using peak-alignment strategies [24–29]. However, when gradient programs are intentionally altered, peak alignment might be unsuccessful due to shifts that are too large to be properly corrected or due to shifts in retention order. For these cases, peak-tracking algorithms have been developed for both 1D [30–33] and 2D [34,35] chromatography. What these peak-tracking algorithms have in common is that they only work for a limited number (typically 2 or 3) chromatograms, or work with a target chromatogram to which all other chromatograms are compared [36]. In the context of a routine application of a method scientists may face large numbers of chromatograms for comparison and selecting a target chromatogram might introduce unintentional bias. Furthermore, when performing routine work in batch analysis, chromatograms which deviate from the others might be of interest. This can be challenging when each chromatogram can only be compared to a few others and consumes a lot of computational power if all peaks need to be tracked across all other chromatograms.

In this paper, we describe a workflow to track peaks across numerous datasets that minimizes computational costs. The workflow leverages our earlier peak-tracking work for 1D liquid chromatography (LC) and gas chromatography (GC) [33], and the peak-tracking algorithm for comprehensive 2D-LC (LC×LC) and GC×GC [35]. These peak-tracking algorithms first create a list of likely candidates, then compare spectral information and peak moments in both the total-ion current and the extracted-ion current to track peaks between two chromatograms. Subsequently, remaining unpaired peaks are compared to determine their positions on both chromatograms. Even though this work makes use of our own peak-tracking algorithms, the principles that are discussed here can be applied to other peak-tracking algorithms in both 1D and 2D separations.

Performing peak tracking between all chromatograms requires a number of peak-tracking steps that increases exponentially with the number of samples. Computational problems that increase exponentially can rapidly become too expensive and time consuming. A linear increase in the number of tracing steps with the number of chromatograms is preferred, but a high accuracy within the data analysis step needs to be maintained. Here we propose a strategy to follow analytes over multiple chromatograms by linking all chromatograms to each other with a linear-increasing number of connections. Each analyte can be "followed" between the chromatograms, however not all chromatograms have to be linked to each other. Two strategies are proposed, each with their own advantages and disadvantages. Lastly, it will be shown that the strategy for gradient optimization and the strategy for batch analysis will differ in making the most efficient use of computational resources on multiple industrial and academic samples differing in compound class, such as an antibody digest, wine volatiles, headspace of polymer granulates and mayonnaise.

2. Experimental

2.1. Experimental conditions

2.1.1. Tryptic digest of a monoclonal antibody (mAb)

Experimental conditions for the LC×LC separations of the mAb digest can be found in previous work [37].

2.1.1.1. Chemicals. A tryptic digest of an in-house produced IgG1 mAb was used. The mobile phase consisted of an aqueous solution (Milli-Q water, Billerica, MA, USA) of 10 mM ammonium bicarbonate (Sigma-Aldrich, St. Louis, MO, USA), pH 9.5, an aqueous solution of 10 mM phosphoric acid and acetonitrile (ACN, Chromasolv LC/MS grade).

2.1.1.2. Instrumentation. The used LC modules were from the 1290 series from Agilent technologies (Waldbronn, Germany) coupled to a quadrupole-time-of-flight (Q-TOF) instrument (Agilent, model G6549A) equipped with the Agilent JetStream electrospray ionization source. In the ^{1}D a Poroshell HPH C18 column (Agilent Technologies, 200 mm × 2.1 mm, 2.7 μm d_p) was used, in the ^{2}D a custom column was packed with commercially available Zorbax Eclipse Plus C18 particles (20 mm × 2.1 mm, 1.8 μm d_p).

2.1.2. Wine volatiles

2.1.2.1. Sample preparation. The divinybenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) solid-phase micro extraction (SPME) fiber (Supelco, St. Louis, MO, USA) used was first conditioned for 30 min at 270 °C. A 20 mL vial containing 1 g of NaCl, 0.5 mL wine and 4.5 mL of deionized water spiked with 4-methyl-3-penten-2-one as an internal standard was pre-incubated for 5 min at 50 °C at an agitation speed of 250 rpm. The fiber was exposed in the headspace for 30 min at 50 °C for the extraction of wine volatiles. Subsequently, the SPME fiber with extracted wine volatiles was immediately desorbed in the GC injection port at 250 °C for 10 min.

2.1.2.2. Instrumentation. The analysis of wine volatiles was performed on a Pegasus 4D GC×GC-TOFMS system (LECO Corporation, St. Joseph, MI, USA) equipped with a 7890A Agilent GC (Agilent Technologies, Palo Alto, GA USA) and Gerstel MPS2 autosampler (Gerstel GmbH & co. KG, Mülheim Van der Ruhr, Germany). A split/splitless injector was used at 250 °C. Split injections were performed in triplicate analysis, with an additional splitless injection performed using a splitless time of 2 min. Helium was used as a carrier gas with split ratio 1:10 at a constant column flow rate of 1 mL·min^{-1}. The GC column configuration consisted of an Rxi-5Sil MS column (30 m × 0.25 mm internal diameter (i.d.) 0.25 μm film thickness (d_f); Restek, Bellefonte, PA, USA) in the ^{1}D, and aStabilwax (0.6 m × 0.25 mm i.d., 0.25 μm d_f; Restek) in the ^{2}D.

2.1.2.3. Procedures. A dual stage cryogenic modulator (LECO) was used with a modulation period of 4 s, and hot and cold pulse times of 0.90 s and 1.10 s, respectively. A secondary oven temperature offset of +50 °C relative to the primary oven was used, while the offset temperature of the modulator relative to the secondary oven was +15 °C. High resolution MS data using electron ionization at an ionization energy of 70 eV. The acquisition rate was 120 Hz for a mass range of 40–520 *m/z* at an extraction frequency of 2 kHz. Tuning and mass calibration was performed daily using perfluorotributylamine (PFTBA) with an acquisition delay of 180 s.

2.1.3. Headspace of polymer granulates

2.1.3.1. Chemicals. About 0.3 gram of polymer granulates (eight different batches), in a capped 20 mL headspace vial, were heated to 50 °C for 30 min, while sampled by solid phase micro extraction (SPME) using a Supelco 65 μm Polydimethylsiloxane/Divinylbenzene (PDMS/DVB) fiber, 23 Ga, autosampler, fiber.

2.1.3.2. Instrumentation. All SPME-GC×GC–TOFMS analyses were carried out on a Leco (St. Joseph, MI, USA) GC×GC 4D system equipped with an Agilent 7683 autosampler, a hot split/splitless injector and a Pegasus TOFMS. The ^{1}D column was a 30 m × 0.25 mm i.d., 0.25 μm d_f Agilent VF1ms and the ^{2}D column was a 2 m × 0.1 mm i.d., 0.2 μm d_f Agilent VF17ms.

2.1.3.3. Procedures. After sampling the fiber was thermally desorbed in the GC inlet, which was set at 300 °C. A constant column flow of 1.2

mL·min^{-1} of helium was used. The oven temperature was programmed from 50 °C which was held for 2 min to 280 °C which was held for 1 min, at a rate of 3 °C min^{-1}. The modulation time was 4 s, the offset +10 °C, and the hot pulse time was set to 1 s. The main TOFMS settings were set as follows: mass scan range was set to 15–600, acquisition rate 100 spectra·s^{-1}, detector voltage 1700, ionization energy −70 eV and ion source 200 °C.

2.1.4. Mayonnaise

2.1.4.1. Sample preparation. Different mayonnaise samples of the same brand with varying 'best-before dates' were purchased at local supermarkets. The sample were equilibrated for 15 min at 40 °C, after which the headspace was injected with a volume of 1 ml at a split ratio of 1:10.

2.1.4.2. Instrumentation. The analysis was performed on a LECO GC×GC-TOFMS system (LECO Corporation, St. Joseph, MI, USA) equipped with an Agilent 6890 GC instrument (Agilent Technologies, Palo Alto, GA USA). Helium was used as a carrier gas at a constant column flow rate of 1.50 mL·min^{-1}. The GC column configuration consisted of StabilWax-DA (30 m × 0.25 mm i.d., 0.50 µm d_f; Restek, Bellefonte, PA, USA) in ^{1}D, and BPX35 (1.0 m × 0.1 mm i.d., 0.1 µm d_f; Trajan, Ringwood, AU) in ^{2}D.

2.1.4.3. Procedures. The primary oven temperature program was: 25 °C for 2 min and then ramped up with 5 °C·min^{-1} to a final temperature of 235 °C for 6 min. Cryomodulation was used with modulation time of 2.5 s with a hot pulse of 0.60 s and 0.65 s cool time between stages. A secondary oven temperature offset of +5 °C relative to the primary oven was used, while the offset temperature of the modulator relative to the secondary oven was +15 °C.

2.1.5. mAb tryptic digest for 1D-LC

Data from previous work [38] was used to make a comparison between the used strategies. Below follows a summary of experimental conditions. Sample preparation can be found in prior work [37].

2.1.5.1. Chemicals. The sample consisted of a monoclonal-antibody digest by trypsin. The mobile phase consisted of an aqueous solution (Milli-Q water, 18.2 MΩ·cm, from a Arium 6111UV, Sartorius, Germany) of 0.1% formic acid (reagent grade ≥ 95%, Sigma-Aldrich, Darmstadt, Germany) and the modifier consisted of ACN (LC-MS grade, Bisolve, Valkenswaard, The Netherlands).

2.1.5.2. Instrumentation. The used instrument was an Agilent Infinity II 2D-LC System, with a binary pump (G7120), a Jet Weaver V35 mixer (G7120–68,135), an autosampler (G4226A), a column oven (G7116B) and a Q-TOF mass spectrometer (G6549A, MS). A Poroshell HPH—C18 (693,675–702, 150 mm × 2.1 mm, 1.9 µm d_p) was column used for all experiments.

2.1.5.3. Procedures. Three scouting gradients were programmed to start with $\varphi = 0.02$ to $\varphi = 1.00$ in 30, 20 and 10 min respectively. The nine other gradient profiles were calculated by the automated method-development protocol using 5 gradient steps to provide improved separations.

2.2. Data processing

The entire algorithm was written using MATLAB 2020b (Mathworks, Natick, MA, USA) for the in-house open access "Multivariate optimization and refinement program for efficient analysis of key separations (MOREPEAKS)" [39]. Raw MS data were converted into .mz5 format by ProteoWizard 3.0.19202 64-bit [40], peak detection was performed with the two-step algorithm by Peters *et al.* [13]. The algorithm was adjusted to loop through the highest abundant *m/z* values until 80% of the total-ion-current was explained. Subsequently, peaks at the same position but with different *m/z* values were merged.

3. Results & discussion

3.1. Connection strategies

Tracking all peaks across all chromatograms can induce an exponential computational problem as the number of chromatograms increases. Each comparison can be envisaged as a link between chromatograms which corresponds to a computation. When all chromatograms in a dataset are compared with each other, then all chromatograms are linked. This can be described as $C = \frac{1}{2}n * (n-1)$, with C the needed number of connections and n the number of chromatograms. Indeed, the number of connections, and thus computations, exponentially increases with increasing n. The exponential increase in demand of computational resources rapidly limits the number of included chromatograms in an assessment. Transforming this exponential increase into a linear demand on computational resources thus allows data to be processed in a more efficient manner. In the present work we thus propose two linking strategies that still allow peaks to be tracked across all chromatograms, yet offer a linear increase in computational resource demand proportional to n. These linking strategies focus on limiting the number of required connections or links for each chromatogram.

The linking strategies could be compared to a game of whispers. Players 1 and 2, or chromatograms 1 and 2, tell each other the locations of their peaks, and player 2 also exchanges their peak locations with player 3. Now player 2 knows how to relate the peaks from player 1 to the peaks of player 3, even though players 1 and 3 never exchanged their peak locations. However, just as in a game of whispers, some chromatograms might provide little or incorrect information to the others. Therefore, we propose that each chromatogram doesn't communicate with just two other chromatograms but with at least three other chromatograms. The number of links however, can be provided as an input to the algorithm. This minimizes the risk that one divergent (*i.e.* missing or co-eluting peaks) chromatogram interrupts the loop. Although if it is certain no chromatograms are divergent, then just two connections per chromatogram could be sufficient. These strategies would provide a linear computational need with $C = \lceil \frac{1}{2}L \,.\, n \rceil$ where L is the number of links for each chromatogram $\{L \mid L \in \mathbb{N},\ L \geq 2\}$. In the case of an uneven number of links and chromatograms, the number needed is rounded up since half connections cannot be made. This linear formula works for any number of chromatograms higher than $\{n \mid n \in \mathbb{N},\ n > L\}$. When n is equal to or lower than L, fewer connections are needed as each chromatogram is already compared with all other chromatograms. To illustrate, if L equals 3, five connections cannot be created without making double connections if only three chromatograms are present (#1 to #2, #2 to #3 and #1 to #3). Fig. 1 shows the number of connections needed for the linear cases with three, four and five links per chromatogram, the exponential case and a cumulative case, which will be discussed in Section 3.1.2.

When all connections are made, the algorithm will combine the information provided by each individual pairing connection into one peak table. To achieve this, the algorithm adds the first results to a peak table. Then it will loop though all remaining results and for each peak in each results set it checks if the peak is already present in the peak table (comparing retention times and chromatogram number). If the peak is already present, the peak and its match on the other chromatogram are combined with the peak table. In the case the match is also found at a different location in the peak table, both groups are combined to one. If both peaks are not present in the peak table, they are added as a new entry in the peak table (for graphical representations please refer to Supporting Information Section S-1).

The algorithm stores all entries and their relationships. In case

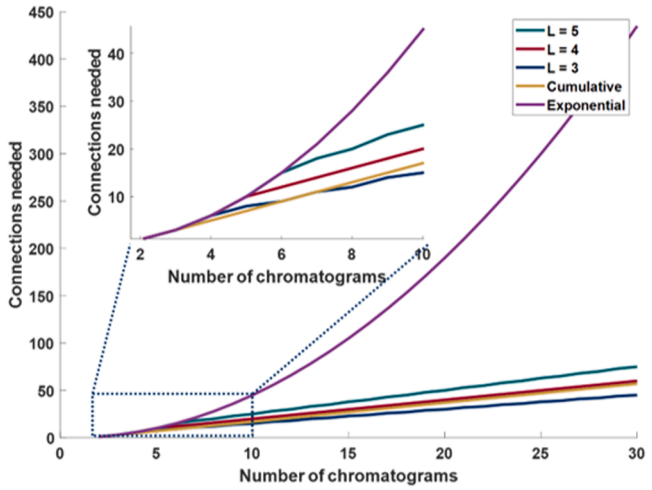

Fig. 1. The required number of connections to compare chromatograms when each set is compared to all available other chromatograms (purple), the cumulative strategy (yellow), with three links (blue), four links (red), or five links (green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contradictory information is available (*e.g.* a peak has been paired to different peaks in another chromatogram through different connections), the algorithm checks all entries to find the most common peak and as a final result will provide the most common (and thus the most likely) match. When both possibilities are just as likely (*i.e.* same number of matches), it is assumed there are co-eluting peaks, and the peak will appear as different entries in the peak table with equal retention times, except for the divergent peaks.

3.1.1. *Semi-randomization strategy for batch analysis*

When creating the network of connections, it is crucial to not create multiple clusters of datasets that are not connected to each other. Therefore, a semi-randomization strategy was selected. In principle, each chromatogram is connected to its neighbor creating a loop (Fig. 2A, black dotted line). This adds two connections to each chromatogram. Thereafter, depending on the number of links required, all other connections are made as random crosslinks within the loop. Double connections are not allowed. If L and n are odd, one chromatogram is randomly chosen that will have an extra connection. When all connections are determined, peak tracking can be performed by parallel computing. Parallel computing can then significantly decrease computational time.

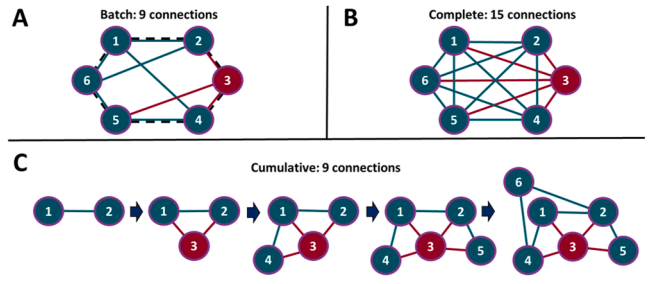

Fig. 2. Different clustering strategies. Chromatogram #3 (red) represents a dataset where something went wrong, and thus connections with chromatogram #3 don't provide information about at least one peak. However, following other connections still creates an intact network (green). **A**) Batch strategy: Each chromatogram is connected to the next, creating a loop indicated with the black dotted line, and then random cross connections are made within the loop. **B**) Complete strategy: All chromatograms are connected to all others. **C**) Cumulative strategy: Every new chromatogram is connected to two randomly selected chromatograms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 2A shows an example with $L = 3$ and $n = 6$ while Fig. 2B shows the case where all chromatograms will be compared to all others. The cumulative strategy in Fig. 2C will be explained in Section 3.1.2. Fig. 2 shows that chromatogram #3 represents a dataset in which no information, about at least one peak, can be extracted (*e.g.* at least one peak is not detected, and thus not tracked in this chromatogram). However, all three strategies still provide paths of connections to maintain an intact network. While the network is intact, information from *e.g.* chromatogram #2 to chromatogram #4 can still be passed along through chromatogram #1, even though there is no direct connection in Fig. 2A and C between these chromatograms #2 and #4.

3.1.2. *Cumulative strategy for automated method-development*

Recently, we introduced a workflow for the automated method-development in 1D-LC [38]. This strategy relied heavily on the ability to perform accurate retention modeling by iterative feedback (*i.e.* predicting and measuring a better separation based on formerly obtained results). Consequently, there is a need to accurately track peaks across all chromatograms and not just across two or three chromatograms. In that work, we solved this by appointing a target chromatogram of which the highest resolution was expected and tracking peaks across all other chromatograms to this target chromatogram. This required to track peaks across all chromatograms yet again, when a new target chromatogram was chosen. This method required a lot of computational power due to the re-track step. However, it also introduced the risk of choosing a chromatogram as target even though the separation performed poorly, possibly leading to poor peak-tracking.

The batch strategy described in Section 3.1.1 cannot address this problem, because the chromatograms are not produced simultaneously. Each iteration of the workflow provides one more chromatogram that needs to be processed. Then, its peaks are added to the retention tables that contain data from the other chromatograms that have already been analyzed. Thus, a peak-tracking strategy where each chromatogram is linked to two randomly selected previous chromatogram is more suitable. This strategy has several disadvantages, namely it (*i*) has an approximately equal cost efficiency compared to the batch strategy with four links ($C = 2n - 3$ vs $C = 2n$, see Fig. 1), (*ii*) cannot be performed in parallel for more than two calculations at the same time and (*iii*) creates the possibility that some chromatograms have a large number of connections. However, it does add the strong advantage of continuing data analysis even as subsequent chromatograms are being recorded. Fig. 2C shows an example of adding connections for up to 6 chromatograms.

3.2. *Peak-tracking results for samples from industrial and academic laboratories*

The peak-tracking workflow has been tested on four different samples containing compounds from different classes. First the batch strategy is applied with three and five links to 29 replicates of a LC×LC separation of peptides from a monoclonal antibody to determine if more links provide a more accurate assessment. Afterwards, GC×GC separations of samples from industrial and academic laboratories are evaluated. Lastly, the cumulative strategy is applied to LC-MS data from previous work [38] to determine if the strategy is a viable alternative to the target-chromatogram strategy.

3.2.1. *Monoclonal-antibody digest: 29 LC×LC-HRMS separations*

Initially, the proposed algorithm was tested on 29 LC×LC—HRMS separations of a mAb digest. If all chromatograms are compared to each other, this would result in 406 connections (*i.e.* computations) that need to be processed. These measurements were approximately 120 min long, resulting in datafiles that required about 10GB of memory. A high-end commercially available data processing computer was used, which required roughly 8 min to load in a chromatogram. Peak tracking would

require two chromatograms to be loaded in and then 8 to 10 min to be able to track peaks between the two chromatograms. This would result in a processing time in the order of 160–175 h of computations. Evidently, parallelization will reduce the needed time, but parallelization is limited by internal memory (≈20 GB per set). Utilizing the batch strategy, the number of connections needed would result in 44 connections for three links (≈9 fold gain in efficiency), 58 connections for four links (≈7 fold gain in efficiency), or 73 connections for five links (≈6 fold gain in efficiency),. First, we ran the algorithm with three links per chromatogram and with five links per chromatogram to assess differences in performance (*i.e.* percentage of chromatograms the peaks were tracked in). Figs. 3 and 4 and Table 1 show the peak-tracking results for three links and five links, respectively. The figures show all the peaks that were found in at least 25% of the chromatograms. A large fraction of the peaks (40% and 41% for three and five links, respectively) were found in less than 25% of the chromatograms. A large number of this latter group of peaks (66% and 59%, respectively) featured a unique dominant *m/z* value exclusive to just that chromatogram and were thus not detected in the other chromatograms (See Figure S-5 and S-7 in Supporting Information Section S-2). Since they were not detected in any other chromatogram, they could not be tracked to any other chromatograms, and therefore could not to be back-tracked through the tracking network. Investigation into all peaks that were tracked in less than 25% of the cases, showed that they were trace analytes that often were not abundant enough to have a dominant *m/z* value above the background signal and were discarded as noise in most chromatograms.

Fig. 5 shows the performance versus the maximum signal-to-noise ratio (S/N) when five links were used. Note that the maximum S/N is used, so in many chromatograms the S/N was even lower than the value provided in Fig. 5. Of the 103 peaks that were tracked in less than 25% of the chromatograms, 93 (90%) had a maximum S/N of less than 20 and 74 of the peaks (72%) even had a S/N ratio below 4. Peak detection is the bottleneck in the performance of the proposed algorithm, as expected. If peaks were not abundant enough to be detected robustly, the peak-tracking algorithm was unable to compensate for this. For these calculations the noise level and average baseline were estimated in the first 5 min of the chromatograms. All chromatograms with peak-tracking results, including those that were tracked in less than 25% of the chromatograms, and the tracking network are shown in Supporting Information Section S-2.

Table 1 shows the results obtained when tracking with three or five links. It is most notable that more trace analytes were tracked when more links were utilized. Whereas five links required more connections, the number of different peaks that were tracked in at least 25% of the chromatograms remained equal (134). However 15 more trace analytes were tracked in less than 25% of the chromatograms. This is due the nature of the peak-tracking algorithm. When comparing two chromatograms, the algorithm will actively search for analytes that are found in at least one of the chromatograms. However, when two chromatograms are not directly linked to each other, the algorithm has no way of knowing to look for newly found peaks in the other chromatograms. Thus, more connections result in more active searches for some trace analytes. This is also reflected in the total coverage of the peaks that are tracked in more than 25% of the chromatograms. There is a 3% higher coverage (*i.e.* percentage of possible peak relations across all chromatograms) compared to the results with three links. In total 3412 peak relations were made across all 29 chromatograms when utilizing three links, and 3913 peak relations were made utilizing five links. This translates to 9% increase in the total number of peak relations (299, of which 117 in the below 25% region), while the computational cost increased by 66%. A user can determine if the small increase in peak coverage is worth the extra computational cost. Since the other datasets discussed in this paper were relatively small, we chose to proceed using just three links in each case.

3.2.2. *Wine volatiles: four GC×GC-HRMS separations*

The algorithm was further tested on four GC×GC—HRMS separations of wine volatiles. In the case of four chromatograms, the maximum links per chromatogram is three, with six connections in total. The separations were performed in triplicate in split mode, with one additional analysis performed in splitless mode. As a consequence, the fourth chromatogram showed higher intensities for most analytes, but overloading may occur for this analysis, affecting peak shape. Statistical moments of the peaks in chromatogram #4 can therefore not be reliably compared to the other chromatograms, and the peak-tracking algorithm relies more on the *m/z* information. However, the higher intensities in chromatogram #4 mean that more peaks were clearly visible. Peaks that had a too-low intensity on chromatograms #1–3 could therefore still be tracked, as they were found in chromatogram #4 and thus the peak-tracking algorithm actively searched for those low intensity peaks in the other chromatograms. Fig. 6 shows the peak-tracking results of the four wine volatile samples. In total 183 peaks were tracked in the four chromatograms. However, 27 of these peaks had an exact *m/z* value of 68.9947 Da, corresponding to the base peak ion of the PFTBA mass calibrant, which was infused throughout the analysis to ensure mass accuracy. Therefore, peaks with this *m/z* were removed from the candidate lists, leaving 144 different peaks that were tracked in the four chromatograms, resulting in 464 peak relations with a coverage of 80%.

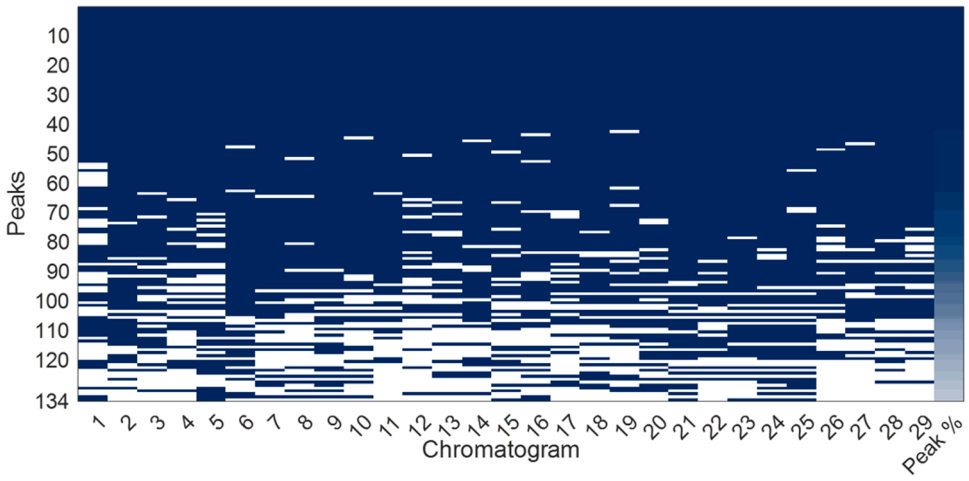

Fig. 3. Tracking results of 29 LC×LC—HRMS repeats of the antibody digest with $L = 3$ ($C = 44$). If a coordinate is filled in, the indicated peak was tracked in the corresponding chromatogram. If a coordinate is left blank, the peak was not tracked in the chromatogram. The color scale in the right-most column indicates the percentage of the chromatograms the peaks were found in.

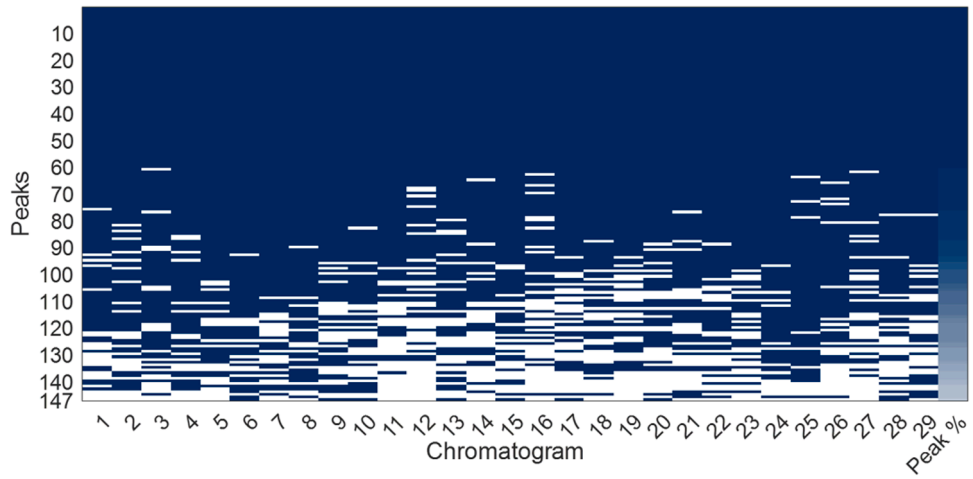

Fig. 4. Tracking results of 29 LC×LC—HRMS repeats of the antibody digest with $L = 5$ ($C = 73$). If a coordinate is filled in, the indicated peak was tracked in the corresponding chromatogram. If a coordinate is left blank, the peak was not tracked in the chromatogram. The color scale in the right-most column indicates the percentage of the chromatograms the peaks were found in.

Table 1
Tracking performances of three links and five links.

L	C	Coverage (%)	Number of peaks	Unique m/z	Peaks with performances below 25%	Unique m/z below 25% performance	Coverage above 25% performance
3	44	53	222	65	88	58	80
5	73	54	250	79	103	62	83

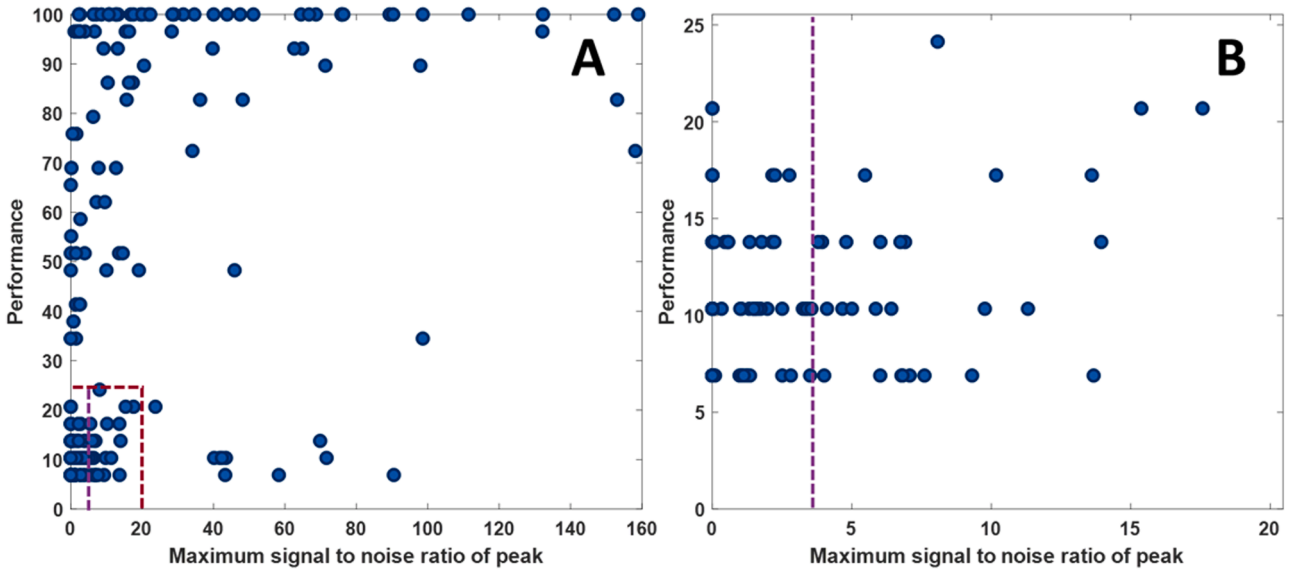

Fig. 5. **A**) Percentage of chromatograms a peak was tracked in, *i.e.* performance, versus the maximum signal to noise ratio over all 29 chromatograms. Of all datapoints with a performance less than 25%, 90% have a signal-to-noise ratio of less than 20 as indicated by the red dashed square. 72% of them have a maximum signal-to-noise ratio of less than 4, as indicated by the purple line. **B**) Zoom in of the red square. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As expected due to the higher intensities, most peaks (137) were tracked in chromatogram #4. All chromatograms with peak-tracking results are shown in Supporting Information Section S-3.

3.2.3. Headspace of polypropylene granulates: eight GC×GC–MS separations

The headspace of polypropylene granulates samples was measured by GC×GC–MS in centroid mode. Thus, less specific information about each peak is available. Many peaks (> 60%) were detected with a m/z of 28 Da. This is mostly the background signal from the carrier gas (N_2). Looking at the chromatograms (Supporting Information Section S-4) there is a large section at the column dead time with a high intensity. Therefore, all peaks with an abundant m/z of 28 Da were deleted from the candidate lists. This could mean that analytes with the same m/z (*e.g.* ethylene, C_2H_4) or present at only trace amounts (*i.e.* the background m/z is most abundant) could be removed unintentionally. Furthermore, centroid data results in more computation need during peak tracking, as more options are available. Especially in separations where many analytes containing C, H, O and N are present, the algorithm has to rely more on peak moments and needs to compare them with extra

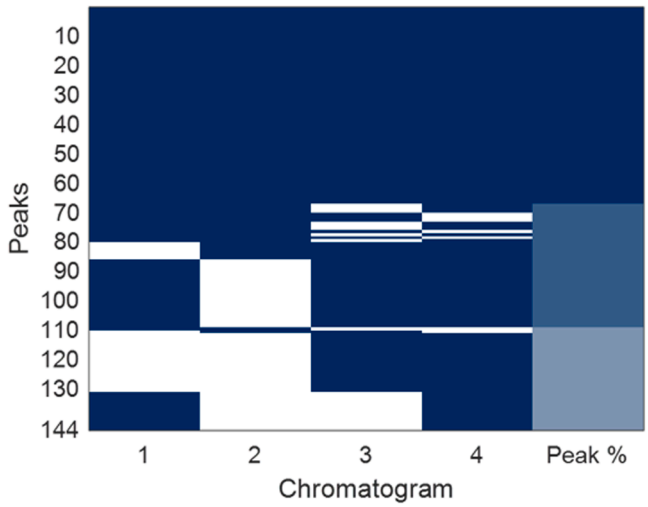

Fig. 6. Tracking results of 4 GC×GC—HRMS separations of the wine volatiles ($C = 6$). If a coordinate is filled in, the indicated peak was tracked in the corresponding chromatogram. If a coordinate is left blank, the peak was not tracked in the chromatogram. The color scale in the right-most column indicates the percentage of the chromatograms the peaks were found in.

candidates. For example, carbon dioxide (CO_2), acetaldehyde (C_2H_4O) and nitrous oxide (N_2O) all have a centroid m/z of 44 Da (*i.e.* isobars with m/z 44 Da). When comparing chromatograms where these compounds are present, the statistical moments of all these peaks need to be compared. When using HRMS, these isobars are represented with different exact m/z values (43.9898 Da, 44.0262 Da and 44.0011 Da respectively), and are therefore more easily distinguished from each other. With larger compounds, the number of isomers is already large, but the number of possible candidates becomes even larger if isobars are taken into account. Thus, the computational cost is higher when performing peak tracking on centroid data.

The results of the batch tracking strategy are shown in Fig. 7. Differences can clearly been seen, most clearly when comparing chromatogram #2 and #8 to the others. Differences were expected, as eight different batches were compared to each other. In total 218 peaks were tracked, resulting in 1256 peak relations (72% of the possible relations). All chromatograms with peak-tracking results and the tracking network are shown in Supporting Information Section S-4.

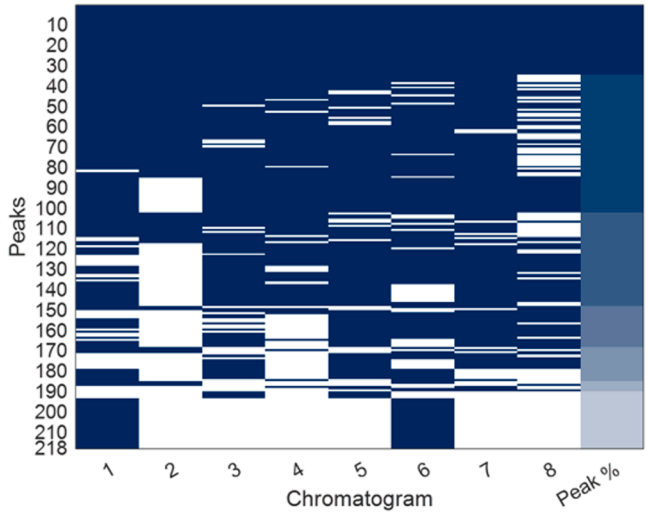

Fig. 7. Tracking results of eight GC×GC–MS separations of the headspace of polypropylene granulates with $L = 3$ ($C = 12$). If a coordinate is filled in, the indicated peak was tracked in the corresponding chromatogram. If a coordinate is left blank, the peak was not tracked in the chromatogram. The color scale in the right-most column indicates the percentage of the chromatograms the peaks were found in.

3.2.4. *Mayonnaise: 12 GC×GC–MS separations*

The headspace of two different brands of mayonnaise at varying "best-before dates" were analyzed with GC×GC–MS. Fresh samples, samples at the middle of their respective end-dates and samples that were at the end of their best-before date were measured in duplicate for both brands. Fig. 8 shows the peak-tracking results from all samples. The samples are shown in order: Twice fresh brand A (#1, 2), twice mid-point brand A (#3, 4), twice end-date brand A (#5, 6), then twice fresh brand B (#7, 8), twice mid-point brand B (#9, 10) and finally twice end-date brand B (#11, 12).

As with the polypropylene headspace analyses, centroid mass data were provided in this case as well. Even though analytes signals were more intense compared to the background than in the polypropylene headspace analysis, many trace analytes were not abundant enough to have a dominant m/z above the background signal. This again requires better peak detection for 2D chromatography that utilizes the mass domain to distinguish these trace analytes from the background. Due to the fragmentation patterns however, the peak-tracking is still able to track peaks correctly, as the mass spectra are distinctive from each other. In total 106 peaks were tracked over the 12 chromatograms, resulting in 3995 peak relations (82% of the area filled). Fewer peaks were tracked with increasing sample age (most notably in chromatograms #10 and #11). This could be because these particular analytes decreased in concentration with age of the sample, and thus "disappeared" into the detector baseline. All chromatograms with peak-tracking results and the tracking network are shown in Supporting Information Section S-5.

3.2.5. *Cumulative strategy for automated method-development for 1D-LC—HRMS of a monoclonal-antibody digest: 12 LC measurements*

As an alternative to comparing each chromatogram to a target chromatogram during automated method-development, the cumulative strategy was tested. A set of 12 LC—HRMS chromatograms from previous work [38] were added to the tracking cluster one by one. This is in contrast to the results in Sections 3.2.1 to 3.2.4, where all chromatograms were added to the cluster at once. In our prior work with this dataset, all chromatograms were first compared to scanning

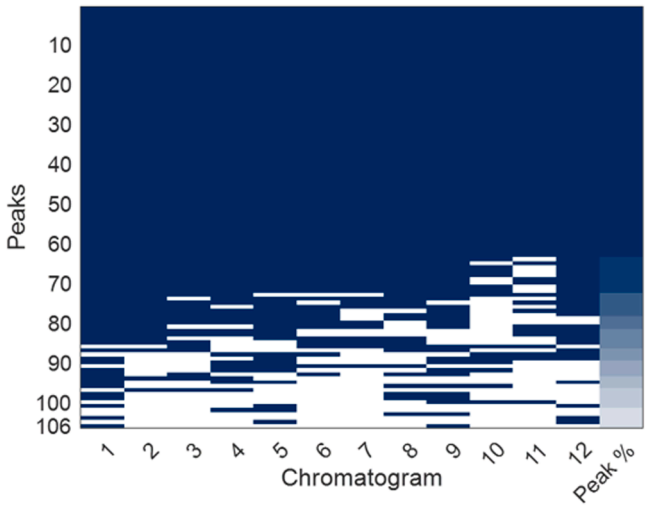

Fig. 8. Tracking results of 12 GC×GC–MS separations of mayonnaise with $L = 3$ ($C = 18$). If a coordinate is filled in, the indicated peak was tracked in the corresponding chromatogram. If a coordinate is left blank, the peak was not tracked in the chromatogram. The color scale in the right-most column indicates the percentage of the chromatograms the peaks were found in.

chromatogram #1. Then, after the first calculation of an ideal separation, all subsequent chromatograms were compared to target chromatogram #4. After five additional iterations, it was expected to have found conditions yielding a better separation and all subsequent chromatograms were compared to target chromatogram #9.

Fig. 9 shows the peak-tracking results of the cumulative strategy compared to the obtained tracking results from the previous work using target chromatograms. Note that there are a different number of total peaks tracked per plot as indicated in the y-label of the plots. When the separation improves, it is expected to detect more peaks in a chromatogram and thus more peaks can be tracked. Furthermore, chromatograms #1, 4 and 9 show that all tracked peaks are present within those chromatograms in Fig. 9B to 9D, respectively. This is of course expected as all other chromatograms were compared to these target chromatograms in the target-chromatogram approach. All chromatograms with peak-tracking results from the cumulative strategy and the tracking network are shown in Supporting Information Section S-6.

Table 2 shows the performance of the cumulative strategy compared to the previous target-chromatogram results. The cumulative strategy resulted in 39 more total peaks tracked compared to the target-chromatogram results of all 12 chromatograms and 160 more peak relations (1328 and 1168). Furthermore, on average 13.33 more peaks were tracked in each chromatogram. Compared to the target-chromatogram strategy, the cumulative strategy is less reliant on the peaks detected in one chromatogram. Thus, it is not surprising more peaks were detected and thus tracked using the cumulative strategy. However, peaks that were tracked are on average found in 1.30 less chromatograms compared to the target-chromatogram strategy. Thus, average performance (*i.e.* percentage of chromatograms it was tracked in) per peak decreased. In the previous work, a target chromatogram was selected where the separation would be expected to be optimal, while the cumulative strategy selects a random chromatogram. This selected chromatogram can be one with a worse separation and thus peaks can be tracked less reliably. In the context of automated optimization however, peaks only have to be found in a few chromatograms for retention modeling to be possible. More tracked peaks reflect a better understanding of the sample and thus a possibility to achieve a better optimal separation. It can therefore be debated if it is preferred to track more peaks, than to track peaks with a higher accuracy. Fig. 10 shows the total number of tracked peaks per tracking step. Here it can be seen that after four chromatograms, the cumulative strategy starts to outperform the target-chromatogram strategy.

4. Conclusion

A more computationally efficient strategy for tracking peaks across numerous chromatograms was developed. In contrast to tracking all peaks across all chromatograms, the strategy requires a linear computational cost to the number of datasets. The algorithm was tested with different samples from industrial and academic laboratories that represent different compound classes.

In the largest dataset of 29 LC×LC—HRMS replicates of a mAb digest, a substantial number of peaks (40%) were tracked in less than 25% of the chromatograms. This was due trace analytes that were rarely detected above the background signal. Given the fact that this entire algorithm hinges on the success of peak detection, it is important to investigate whether other peak-detection strategies would improve the peak tracking. In this light, the recent development of a region of interest-multivariate curve resolution based strategy for peak detection is of high interest [41]. More links between chromatograms did not

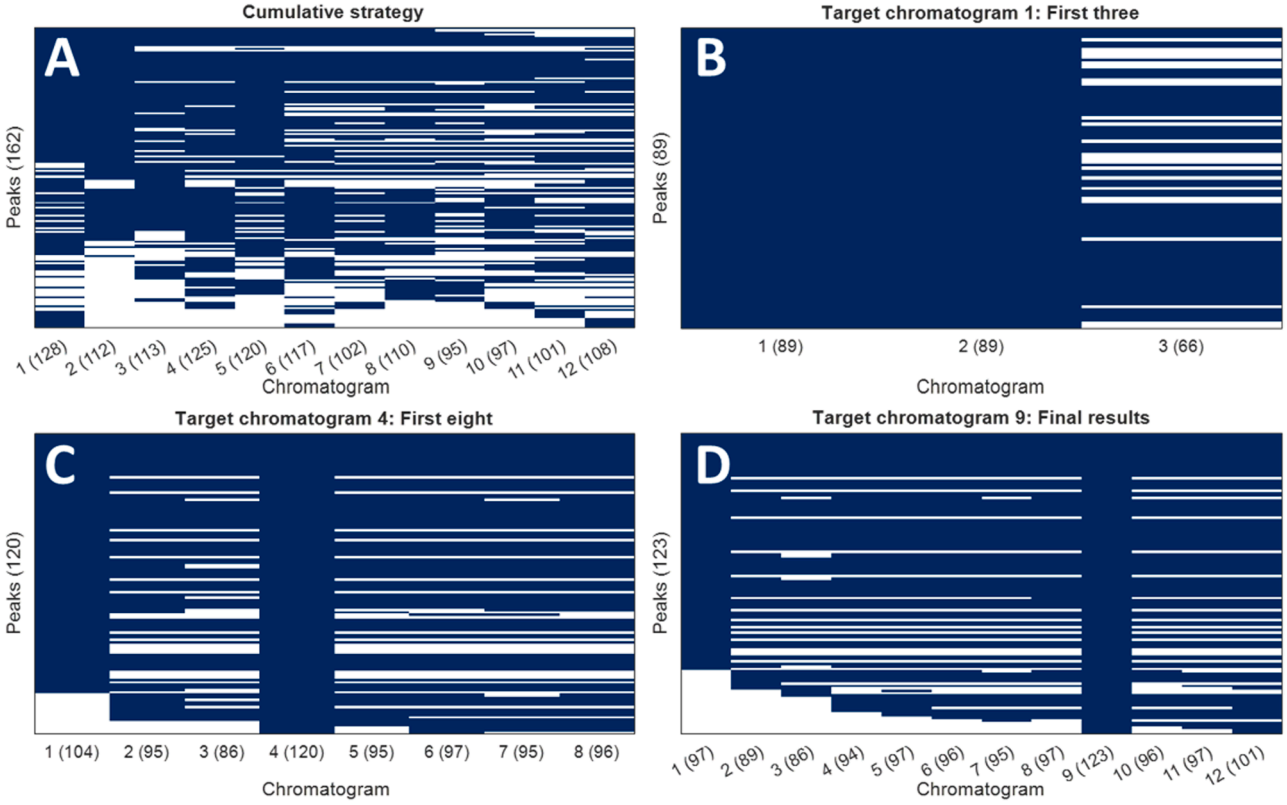

Fig. 9. Comparison of tracking results from 12 LC—HRMS separations of the mAb digest obtained with the cumulative strategy ($C = 21$) and comparing all chromatograms to one target chromatogram. Total numbers of tracked peaks are indicated in the y-labels, the numbers of tracked peaks per chromatogram are indicated between brackets in the x-labels. **A**) Tracking results of all 12 chromatograms using the cumulative strategy. **B**) Tracking results of the first 3 chromatograms with target chromatogram #1. **C**) Tracking results of the first eight chromatograms with target chromatogram #4. **D**) Tracking results of all 12 chromatograms with target chromatogram #9.

Table 2
Comparison of the cumulative strategy to the target-chromatogram strategy in previous work [38].

	Total number of peaks	Average number of peaks per chromatogram	Average number of chromatograms per peak	Total peak relations
Cumulative strategy	162	110.67	8.20	1328
Target-chromatogram strategy	123	97.33	9.50	1168

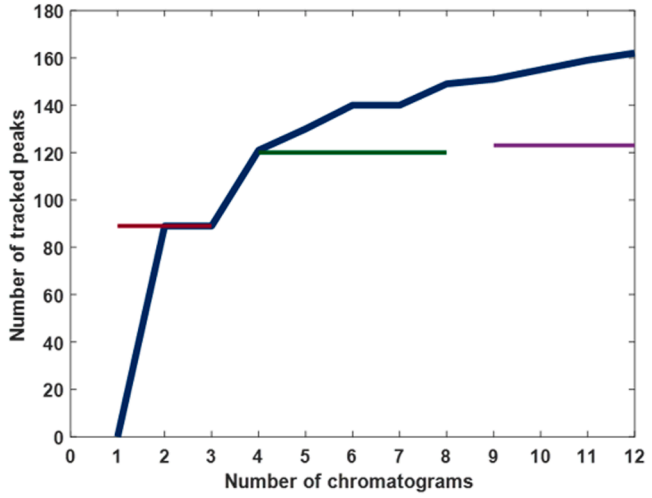

Fig. 10. Number of tracked peaks after each tracking step. The blue line indicated the cumulative strategy, the red line indicates the first three chromatograms with target chromatogram #1, green line indicated the first eight chromatograms with target chromatogram #4 and the purple line indicates the final tracking results with target chromatogram #9. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reduce the number of peaks that were found in less than 25% of the chromatograms, but did slightly improve the performance (*i.e.* number of chromatogram in which the peak was tracked) for those above this threshold. The computational need, however, grew more substantially than the performance with an increasing number of links, and thus a user can decide if the improvement in accuracy is worth the additional computational need.

Centroid MS data increased the algorithm's computational need considerably. As a result of isobars within the sample, the algorithm had to compare more possible candidates compared to data where the exact mass of the analytes is measured (*i.e.* high resolution MS). The centroid mass data depended on fragmentation patterns to be accurately tracked, which provides more (albeit less accurate) information to the peak-tracking algorithm. Due to the fragmentation however, the most abundant *m/z* is often not representative of the molecular weight, but for the most stable fragment. If this fragment is an isobar of the most abundant background *m/z*, there is a possibility that the peak is unintentionally removed from candidate lists by the algorithm.

Finally, the cumulative strategy was tested for the use in automated method-development. Automated method-developed required a different strategy, since chromatograms are acquired sequentially during the development process, and thus are not all available initially for simultaneously processing. The cumulative strategy resulted in a quicker assessment of the number of peaks present within the samples, but was slightly less accurate in tracking those peaks over all chromatograms compared to the target-chromatogram strategy.

CRediT authorship contribution statement

Stef R.A. Molenaar: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization. **John H.M. Mommers:** Investigation, Resources, Writing – review & editing. **Dwight R. Stoll:** Investigation, Resources, Writing – review & editing. **Sithandile Ngxangxa:** Investigation, Writing – review & editing. **André J. de Villiers:** Resources, Writing – review & editing. **Peter J. Schoenmakers:** Supervision, Project administration, Funding acquisition. **Bob W.J. Pirok:** Writing – review & editing, Supervision, Project administration.

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.

Acknowledgements

The authors would like to thank prof. dr. ir. Hans-Gerd Janssen for the provided data. We also thank Hayley Lhotka for her effort in collecting the mAb digest data. A special thanks to Tijmen S. Bos for the use of his scripts to significantly decrease the loading time of the .mz5 data format. SM acknowledges the UNMATCHED project, which is supported by BASF, DSM and Nouryon, and receives funding from the Dutch Research Council (NWO) in the framework of the Innovation Fund for Chemistry (CHIPP Project 731.017.303) and from the Ministry of Economic Affairs in the framework of the "PPS-toeslagregeling". DS acknowledges support from an Agilent Technologies Thought Leader Award. All experiments in his laboratory were carried out using instrumentation provided by Agilent. BP acknowledges the TTW VENI research program (Project 19173, "Unleashing the Potential of Separation Technology to Achieve Innovation in Research and Society (UPSTAIRS)"), which is financed by the Dutch Research Council (NWO). This work was performed in the context of the Chemometrics and Advanced Separations Team (CAST) within the centre for Analytical Sciences Amsterdam (CASA). The valuable contributions of the CAST members are gratefully acknowledged.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2023.464223.

References

[1] 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.
[2] L.E. Niezen, P.J. Schoenmakers, B.W.J. Pirok, Critical comparison of background correction algorithms used in chromatography, Anal. Chim. Acta. 1201 (2022), 339605, https://doi.org/10.1016/j.aca.2022.339605.
[3] J. Kuligowski, D. Pérez-Guaita, I. Lliso, J. Escobar, Z. León, L. Gombau, R. Solberg, O.D. Saugstad, M. Vento, G. Quintás, Detection of batch effects in liquid chromatography-mass spectrometry metabolomic data using guided principal component analysis, Talanta 130 (2014) 442–448, https://doi.org/10.1016/j.talanta.2014.07.031.
[4] N. Kawase, H. Tsugawa, T. Bamba, E. Fukusaki, Different-batch metabolome analysis of Saccharomyces cerevisiae based on gas chromatography/mass spectrometry, J. Biosci. Bioeng. 117 (2014) 248–255, https://doi.org/10.1016/J.JBIOSC.2013.07.008.

[5] M. Wood, M. Laloup, N. Samyn, M. del Mar Ramirez Fernandez, E.A. de Bruijn, R. A.A. Maes, G. De Boeck, Recent applications of liquid chromatography-mass spectrometry in forensic science, J. Chromatogr. A. 1130 (2006) 3–15, https://doi.org/10.1016/J.CHROMA.2006.04.084.
[6] B. Gruber, B.A. Weggler, R. Jaramillo, K.A. Murrell, P.K. Piotrowski, F.L. Dorman, Comprehensive two-dimensional gas chromatography in forensic science: a critical review of recent trends, TrAC - Trends Anal. Chem. 105 (2018) 292–301, https://doi.org/10.1016/J.TRAC.2018.05.017.
[7] M. Qu, B. An, S. Shen, M. Zhang, X. Shen, X. Duan, J.P. Balthasar, J. Qu, Qualitative and quantitative characterization of protein biotherapeutics with liquid chromatography mass spectrometry, Mass Spectrom. Rev. 36 (2017) 734–754, https://doi.org/10.1002/MAS.21500.
[8] 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.
[9] 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.
[10] M. Woldegebriel, G. Vivó-Truyols, Probabilistic model for untargeted peak detection in LC-MS using Bayesian statistics, Anal. Chem. 87 (2015) 7345–7355, https://doi.org/10.1021/acs.analchem.5b01521.
[11] M. Lopatka, G. Vivó-Truyols, M.J. Sjerps, Probabilistic peak detection for first-order chromatographic data, Anal. Chim. Acta. 817 (2014) 9–16, https://doi.org/10.1016/j.aca.2014.02.015.
[12] G. Vivó-Truyols, J.R. Torres-Lapasió, A.M. Van Nederkassel, Y. Vander Heyden, D. L. Massart, Automatic program for peak detection and deconvolution of multi-overlapped chromatographic signals: part I: peak detection, J. Chromatogr. A. 1096 (2005) 133–145, https://doi.org/10.1016/j.chroma.2005.03.092.
[13] S. Peters, G. Vivó-Truyols, P.J. Marriott, P.J. Schoenmakers, Development of an algorithm for peak detection in comprehensive two-dimensional chromatography, J. Chromatogr. A. 1156 (2007) 14–24, https://doi.org/10.1016/j.chroma.2006.10.066.
[14] B. Li, S.E. Reichenbach, Q. Tao, R. Zhu, A streak detection approach for comprehensive two-dimensional gas chromatography based on image analysis, Neural Comput. Appl. (2018), https://doi.org/10.1007/s00521-018-3917-z.
[15] J. Xu, L. Zheng, G. Su, B. Sun, M. Zhao, An improved peak clustering algorithm for comprehensive two-dimensional liquid chromatography data analysis, J. Chromatogr. A. 1602 (2019) 273–283, https://doi.org/10.1016/j.chroma.2019.05.046.
[16] J.K. Strasters, H.A.H. Billiet, L. de Galan, B.G.M. Vandeginste, Strategy for peak tracking in liquid chromatography on the basis of a multivariate analysis of spectral data, J. Chromatogr. A. 499 (1990) 499–522, https://doi.org/10.1016/S0021-9673(00)96996-6.
[17] 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.
[18] C.N. Cain, T.J. Trinklein, G.S. Ochoa, R.E. Synovec, Tile-based pairwise analysis of GC × GC-TOFMS data to facilitate analyte discovery and mass spectrum purification, Anal. Chem. 94 (2022) 5658–5666, https://doi.org/10.1021/acs.analchem.2c00223.
[19] T.J. Trinklein, R.E. Synovec, Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis, J. Chromatogr. A. 1677 (2022), 463321, https://doi.org/10.1016/j.chroma.2022.463321.
[20] 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.
[21] M.J. den Uijl, P.J. Schoenmakers, B.W.J. Pirok, M.R. van Bommel, Recent applications of retention modelling in liquid chromatography, J. Sep. Sci. 44 (2021) 88–114, https://doi.org/10.1002/jssc.202000905.
[22] 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.
[23] A. Barcaru, A. Anroedh-Sampat, H.G. Janssen, G. Vivó-Truyols, Retention time prediction in temperature-programmed, comprehensive two-dimensional gas chromatography: modeling and error assessment, J. Chromatogr. A. 1368 (2014) 190–198, https://doi.org/10.1016/j.chroma.2014.09.055.
[24] 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.
[25] S.E. Reichenbach, P.W. Carr, D.R. Stoll, Q. Tao, Smart Templates for peak pattern matching with comprehensive two-dimensional liquid chromatography, J. Chromatogr. A. 1216 (2009) 3458–3466, https://doi.org/10.1016/j.chroma.2008.09.058.
[26] Y. Liu, L. Chen, H. Chang, H. Wu, P. Liao, V.S. Tseng, A novel peak alignment method for LC-MS data analysis using cluster-based techniques. 2010 IEEE Int. Conf. Bioinforma. Biomed. Work, IEEE, 2010, pp. 525–530, https://doi.org/10.1109/BIBMW.2010.5703856.
[27] B.Q. Li, J. Chen, X. Wang, M.L. Xu, H.L. Zhai, Longest distance shifting: a simple and efficient approach for the alignment of shifted chromatographic peaks, J. Sep. Sci. 39 (2016) 4549–4556, https://doi.org/10.1002/jssc.201600811.
[28] J. Mommers, J. Knooren, Y. Mengerink, A. Wilbers, R. Vreuls, S. van der Wal, Retention time locking procedure for comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1218 (2011) 3159–3165, https://doi.org/10.1016/j.chroma.2010.08.065.
[29] J.J.A.M. Weusten, E.P.P.A. Derks, J.H.M. Mommers, S. van der Wal, Alignment and clustering strategies for GC×GC-MS features using a cylindrical mapping, Anal. Chim. Acta. 726 (2012) 9–21, https://doi.org/10.1016/j.aca.2012.03.009.
[30] 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.
[31] 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.
[32] 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.
[33] 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.
[34] 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.
[35] 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.
[36] 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.
[37] 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 Anal. Technol. Biomed. Life Sci. 1134–1135 (2019), 121832, https://doi.org/10.1016/j.jchromb.2019.121832.
[38] 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.
[39] S.R.A. Molenaar, P.J. Schoenmakers, B.W.J. Pirok, Multivariate Optimization and Refinement Program for Efficient Analysis of Key Separations (MOREPEAKS), (2021). doi:10.5281/zenodo.5710442.
[40] 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.
[41] M. Pérez-Cova, S. Platikanov, R. Tauler, J. Jaumot, Quantification strategies for two-dimensional liquid chromatography datasets using regions of interest and multivariate curve resolution approaches, Talanta 247 (2022), 123586, https://doi.org/10.1016/j.talanta.2022.123586.