SUPPLEMENTARY MATERIAL

Peak-Tracking Algorithm for Use in Automated Interpretive Method-Development Tools in Liquid Chromatography

Bob W.J. Pirok^{a,b,*}, Stef R.A. Molenaar^a, Liana S. Roca^a, Peter J. Schoenmakers^a

a University of Amsterdam, van 't Hoff Institute for Molecular Sciences, Analytical-Chemistry Group, Science Park 904, 1098 XH Amsterdam, The Netherlands. **b TI-COAST, Science Park 904, 1098 XH Amsterdam, The Netherlands.**

(*) Corresponding author E-mail: B.W.J.Pirok@uva.nl

Contents

S-1 Example of relative-retention matrix

Table S1 – Example of matrix with relative retention times of all possible combinations of peaks across chromatograms and . Blue squares indicate combinations which met the relative-retention criterion (in this case between 0.33 and 0.55) and are considered to be "logical combinations".

S-2 Narrowing evaluation limits through the relation between retention times

In the preparation block of the algorithm, relative retention times are calculated and compared to the mean relative retention time to define a search window for the most likely combinations. Although this rather crude selection method significantly reduces the number of possible pairs, it is still prudent to verify whether or not peak pairings are sensible.

Figure S1 displays the relative retention times for the data obtained from the comparison block of the algorithm after considering only the initially determined logical combinations. By plotting the retention time of all pairs from chromatogram m against the ratio of the retention times for all pairs, a clear relation between the retention time becomes apparent.

For the evaluation, a second-degree polynomial is fitted to the data. For further evaluation of the pairs this fitted relation is used. Pairs that significantly deviate from this relation (>5%) are considered suspects and may be excluded. Moreover, in the event that more than one candidate peaks in chromatogram n fall within the window for any given peak of chromatogram m , the algorithm will initiate more-exhaustive *post-hoc* comparison methods. An example of the latter case can be found in section S-7 of this Supporting Information document.

Figure S1 – Retention time of likely combinations (post-comparison of the logical combination pool) from chromatogram *m* plotted against A) the retention times from chromatogram *n*, B) the ratio $t_{R,m}/t_{R,n}$.

S-3 Example of unfiltered chromatograms for alizarin

Figure S2 – Examples of typical unprocessed chromatograms after initial peak detection. Pink parkers represent detected regions of interest.

S-4 Alizarin analysis by ion-pair reversed-phase LC

This section displays more results from the alizarin analysis detailed in section 3.1.

Figure S3 - Gradient-scanning LC-MS chromatograms of light-degraded eosin using a gradient time of A) 6, and B) 18 min. Detected, filtered peaks are depicted with the pink marker, the green numbers reflect identified peak pairs. Note, pair number 16 was not manually assigned in the user interface and thus is missing.

S-5 Minimum prominence

To detect chromatographic bands of isomers in the extracted-ion chromatogram (XIC), the algorithm applies a simple local-maxima search function. A relatively strict (*i.e.* high) minimal prominence of the peak is used. The advantage of this is that the algorithm will ignore relatively prominent noise bands at the particular *m*/*z* value. However, if the threshold is set incorrectly, the algorithm may find an unequal number of isomer peaks in the two chromatograms. The latter will result in rejection of some peaks. As can be seen in Figure S4, three isomer bands are present in the chromatogram. However, in chromatogram A the peak marked with 6 is barely resolved. While the algorithm was capable of pairing the first two of the three peaks initially without an issue, the algorithm only counted two peaks in total for chromatogram A. As a result, the block described in section 2.3.3 merely searched for two isomers, assuming that two peaks had co-eluted in chromatogram A. An adjustment of the prominence threshold would be an obvious solution, but the optimal threshold to find the "true" number of pairs differs significantly from experiment to experiment.

Figure S4 – Extracted-ion chromatograms for the analysis of alizarin at m/z 522.22, recorded using a gradient time of A) 6, and B) 18 min.

S-6 Results of retention-parameter fit and retention-time prediction for alizarin

The following table lists the results of the fitting of retention parameters for the paired peaks from the alizarin sample.

Figure S5 – Retention plot based on all the fitted retention parameters (see Table S3). In such a plot, the natural logarithm of the retention factor, $\ln k$, is plotted against the organic modifier fraction (φ) . Note how two **retention curves are almost vertical, indicating that they are strongly influenced by the mobile phase. However, a closer inspection of the retention parameters (see also Table S3) reveals that the parameters are unrealistic. These compounds were also found to elute at the identical retention times during blank runs and thus are system peaks. The algorithm can recognize such spurious vertical lines and reject the pairs.**

Next, the retention times were predicted for a chromatogram recorded at a gradient time of 12 minutes (scanning gradients utilized 6 and 18 min). The retention times of the predicted and experimentally obtained chromatogram are compared in the table below. The prediction error is in the range of 1%, which is in agreement with different modelling studies based on manual peak pairing. Consequently, it can be concluded that the error arises from the models and not as a result of peak pairing.

Also note the automatic rejection of pairs 14 and 15, which were found to have unrealistic retention curves in the $\ln k$ vs φ plot (Figure S5).

The following table lists the retention parameters determined through automatic tracking of the peaks. These parameters were used for prediction. One interesting point is that two peaks were found to have unrealistically high model parameters. The automatic fit function of MATLAB **fminsearch** estimates values for $\ln k_0$ and S until it reaches a point where the predicted retention times are as close as possible to the experimental retention times. In the event that the analytes do not follow the retention model (such as invariable system peaks), no parameter values will reproduce the experimental retention times. MATLAB thus pointlessly increases the parameters until a maximum number of iterations is reached. This results in extremely high retention parameters, which – upon plotting in the retention plot – appear as "polymeric species" which are dramatically influenced by the mobile-phase composition within a specific composition domain. The contrary is true, however. These species do not move under any circumstances and are simply not correctly modelled.

Table S3 – Overview of retention parameters and prediction accuracies for the paired analytes from the alizarin sample. The average prediction error was 0.9%. Two analytes were rejected due to their abnormal and unrealistically high retention parameters.

Peak ID	$ln k_0$	S	AIC	t_{pred}	$t_{\rm exp}$	$ t_{\text{pred}} - t_{\text{exp}} $	Error
$\mathbf{1}$	4.65	10.14	-44.99	4.42	4.57	0.15	3%
$\overline{2}$	4.92	10.34	-45.85	4.70	4.72	0.02	1%
3	4.97	9.41	-42.99	5.08	5.01	0.07	1%
4	5.07	9.42	-41.51	5.21	5.15	0.06	1%
5	5.36	8.02	-44.35	6.30	6.23	0.07	1%
6	5.69	7.77	-41.03	7.00	6.92	0.07	1%
$\overline{7}$	6.13	8.36	-42.89	7.32	7.24	0.08	1%
8	7.54	8.85	-43.79	9.14	8.92	0.22	2%
9	7.87	7.32	-41.00	11.25	11.19	0.07	1%
10	8.36	7.63	-43.68	11.72	11.67	0.06	1%
11	9.35	8.37	-41.57	12.42	12.41	0.01	0%
12	11.95	10.72	-42.38	13.37	13.39	0.02	0%
13	14.61	13.04	-44.33	14.59	14.57	0.02	0%
14	$1.37 \cdot 10^{9}$	$1.44 \!\cdot\! 10^{9}$	-15.68	Rejected	15.84	Rejected	Rejected
15	$2.73 \cdot 10^{2}$	$2.85 \cdot 10^{2}$	-69.91	Rejected	16.05	Rejected	Rejected
16	5.21	9.02	-45.51	5.57	5.50	0.07	1%

S-7 Occlusion of isomer chromatographic bands

The algorithm was found to successfully resolve the case depicted in Figure S6 below for two isomer species. As can be seen in chromatogram A, the chromatographic bands of analytes of different *m*/*z*, but with multiple isomers each, co-elute. The three isomers of *m*/*z* 488.88 and the two isomers for *m*/*z* 568.79 are all correctly paired despite the occurrence of significant co-elution of the individual analyte bands. Note that the green number indicators correspond to the blue (*m*/*z* 488.88) trace. For the orange trace, the number indicators 9 and 11 are plotted at the bottom of the left and right peak respectively to indicate the pair. The chromatographic bands for the two isomers with *m*/*z* 408.97 were, however, rejected by the algorithm. Section S-8 details this case further.

Figure S6 – Extracted-ion chromatograms (XIC) and total-ion chromatograms (TIC) of eosin recorded using a gradient time of A) 6 and B) 18 min. In the XICs, the blue curve represents *m***/***z* **488.88, orange** *m***/***z* **568.97 and yellow** *m***/***z* **408.97.**

S-8 Tracking a peak with a different most-abundant *m***/***z* **across chromatograms**

In Figure S7 below, the algorithm evaluates the earlier-accepted likely pair a_m, a_n (blue arrows and dots).

Figure S7 – Total-ion-current chromatograms (TIC) of experiment and , with the peak-of-interest (POI) marked.

To assess the likelihood of the pair, the algorithm determines the most abundant mass in the mass spectrum at the apex of peaks a_m and a_n . The algorithm verifies that the location of the peak apex in the total-ion-current chromatogram (TIC) precisely matches that of the peak apex in the extracted-ion chromatogram (XIC) of the corresponding abundant mass. In this case, the test is passed, as is the test of relative peak areas in both XICs.

Next, the algorithm checks the values of the most abundant m/z and finds that for a_m this is 408.97, whereas it is 410.97 for a_n . While this suspicious inequality may be reason to reject the matched pair, it is good to realize that the earlier comparison of the mass spectra in the Comparison block of the algorithm (see Section 2.2) determined this match to be likely (>75% similarity based on the 30 most-abundant peaks for each spectrum).

Instead of rejecting the pair, the algorithm was programmed to now search for neighbours in sufficiently close proximity. The most abundant mass of a_m (408.97) is investigated in chromatogram *n* and the most abundant mass of a_n (410.97) is searched in chromatogram m . Figure S8 displays the XIC chromatograms for both masses in both chromatograms.

Figure S8 – Extracted-ion-current chromatograms (XIC) of m/z values 408.97 (blue) and 410.97 (red) in chromatograms and .

From Figure S8 it becomes immediately clear that a nearby neighbour represents a very similar molecule. In fact, they are isomers of a degradation product of eosin, in which three of the four bromine atoms have been removed as a result of light-induced degradation. One bromine atom is left, which explains the two *m*/*z*-values with similar intensity featured in the mass spectra. The algorithm cannot rely on such background information and must verify whether the earlier likely match is true or false.

The algorithm takes the m/z value of 408.97 from the most abundant signal in the of a_m spectrum as reference. The retention time of a_m of $t_{R,m}$ = 4.26 minutes is compared with the retention time of peaks found in the XIC chromatogram of n at the same m/z value. For $a_{n,1}$ and $a_{n,2}$ (Figure S8n, blue) the retention times are $t_{R,a_{n,1}}$ 8.20 and $t_{R,a_{n,2}}$ 8.75 minutes, respectively. According to the relation in Figure S2B, these retention times can be related to the times of chromatogram m by

$$
t_{R,m}/t_{R,n} = 0.005 \, t_{R,m}^2 - 0.0757 t_{R,m} + 0.7194 \tag{S1}
$$

From the result of the calculation, the algorithm can determine which pair is most likely from the perspective of retention. In this case, the values for $a_{n,2}$ and $a_{m,2}$ are more similar than $a_{n,1}$ and $a_{m,2}$. The algorithm thus concludes that $a_{n,1}$ is to be paired with $a_{m,1}$, and $a_{n,2}$ with $a_{m,2}$. The fact that the most-abundant masses do not match is not of an issue. While indeed $a_{m,1}$ has the most abundant mass of 408.97 and $a_{n,2}$ 410.97, the XIC of chromatogram n at 408.97 did show that there also is an apex at 408.97. This piece of evidence, the MS-30 similarity, and the relative retention all suggest that the peaks are in fact correctly paired.

S-9 Eosin analysis by ion-pair reversed-phase LC

This section displays more data from the eosin analysis described in section 3.2.

Figure S9 – Total-ion-current chromatograms of degraded eosin using a gradient duration of (top) 6 and (bottom) 18 min. Green numbers refer to paired peaks that were tracked across the chromatograms.

Figure S10 – (A) Relative retention of chromatograms m and n ($t_{R,m}$ = 100) plotted for each peak pair. Any **significant changes in elution order can easily be spotted using this plot. (B) Tracking information for all identified pairs with several scores contributing to the total score. Using this information, the user can easily identify suspect pairs and reject them if necessary.**

Table S4 – Retention times and most abundant m/z-values for tracked peak pairs of the eosin sample.

Figure S11 - Retention plot based on all the fitted retention parameters for the eosin sample (see Table S5). Here, the natural logarithm of the retention factor, $\ln k$, is plotted against the organic modifier fraction (φ).

Peak ID	$\ln k_0$	\boldsymbol{S}	AIC	t_{pred}	$t_{\rm exp}$	$ t_{\text{pred}} - t_{\text{exp}} $	Error
$\mathbf{1}$	6.87	18.30	-45.93	4.81	4.80	0.01	0%
$\overline{2}$	6.36	13.69	-43.41	5.50	5.48	0.01	0%
3	6.17	12.29	-40.24	5.74	5.70	0.04	1%
4	6.48	12.82	-47.34	5.89	5.85	0.04	1%
5	6.25	11.28	-44.71	6.21	6.24	0.03	0%
6	6.63	12.06	-45.61	6.33	6.30	0.03	0%
$\overline{7}$	7.13	12.39	-44.18	6.73	6.71	0.02	0%
8	7.29	12.17	-43.88	7.01	6.97	0.03	0%
9	7.40	11.74	-40.80	7.33	7.30	0.03	0%
10	7.50	11.85	-42.17	7.39	7.34	0.04	1%
11	7.53	11.48	-43.02	7.62	7.59	0.03	0%
12	7.75	11.59	-41.54	7.81	7.78	0.03	0%
13	7.67	10.91	-45.47	8.09	8.07	0.03	0%
14	7.97	10.44	-43.86	8.76	8.71	0.05	1%
15	7.67	8.81	-41.36	9.59	9.49	0.10	1%
16	7.79	8.78	-42.01	9.79	9.76	0.02	0%
17	8.72	8.77	-39.80	11.22	11.19	0.03	0%
18	9.35	9.21	-39.41	11.69	11.67	0.01	0%
19	10.10	9.51	-40.09	12.43	12.41	0.02	0%
20	12.07	11.01	-40.74	13.38	13.39	0.01	0%

Table S5 – Overview of retention parameters and prediction accuracies for the different paired analytes from the eosin sample. The average prediction error was 0.4%.