

Parameter selection for peak alignment in chromatographic sample profiling: objective quality indicators and use of control samples

Sonja Peters · Ewoud van Velzen · Hans-Gerd Janssen

Received: 27 October 2008 / Revised: 19 January 2009 / Accepted: 28 January 2009 / Published online: 21 February 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract In chromatographic profiling applications, peak alignment is often essential as most chromatographic systems exhibit small peak shifts over time. When using currently available alignment algorithms, there are several parameters that determine the outcome of the alignment process. Selecting the optimum set of parameters, however, is not straightforward, and the quality of an alignment result is at least partly determined by subjective decisions. Here, we demonstrate a new strategy to objectively determine the quality of an alignment result. This strategy makes use of a set of control samples that are analysed both spiked and non-spiked. With this set, not only the system and the method can be checked but also the quality of the peak alignment can be evaluated. The developed strategy was tested on a representative metabolomics data set using three software packages, namely Markerlynx™, MZmine and MetAlign. The results indicate that the method was able to assess and define the

quality of an alignment process without any subjective interference of the analyst, making the method a valuable contribution to the data handling process of chromatography-based metabolomics data.

Keywords Chromatographic profiling · Metabolomics · Peak alignment · Quality control · Control samples

Introduction

Chromatography in combination with mass spectrometry is widely used for the analysis of complex samples. Recently, a new way of looking at the obtained chromatograms has evolved. Rather than focussing on a limited set of target compounds, the whole chromatogram, either as the total ion chromatogram or the set of spectra, is considered. The chromatogram is thus treated as a fingerprint of the sample. This approach especially gained popularity in metabolic profiling where it is often applied to detect new (bio) markers in large data sets by means of multivariate analytical techniques. These multivariate techniques require peaks to be aligned, meaning that any given compound must be present at exactly the same time point in all chromatograms. While the potential of chromatographic fingerprinting is undisputed, the choice of a suitable alignment strategy out of the algorithms available is not an easy task and represents a major obstruction for further acceptance and use of chromatographic profiling.

The development of alignment strategies has received a great deal of attention in recent literature. The most commonly applied approaches make use of warping of the (raw) signal [e.g. 1–3] or use algorithms based on matching detected peaks [e.g. 4, 5]. In the current study, only the latter strategy is investigated. An overview on alignment software packages freely or commercially available can be

S. Peters (✉) · E. van Velzen · H.-G. Janssen
Advanced Measurement and Imaging,
Unilever Research and Development,
Unilever Food and Health Research Institute,
P. O. Box 114, 3130 AC, Vlaardingen, The Netherlands
e-mail: sonja.peters@unilever.com

S. Peters · H.-G. Janssen
Polymer Analysis Group,
van't Hoff Institute for Molecular Sciences,
University of Amsterdam,
Nieuwe Achtergracht 166,
1018 WV, Amsterdam, The Netherlands

E. van Velzen
Biosystems Data Analysis,
Swammerdam Institute for Life Sciences,
University of Amsterdam,
Nieuwe Achtergracht 166,
1018 WV, Amsterdam, The Netherlands

found in a recent review by Katajamaa et al. [6]. The problem in using and improving published alignment strategies is that not all algorithms are well described. Moreover, input parameters usually need to be selected for peak alignment that are not always straightforward to determine.

Parameters to be set in most software packages can be divided into two groups. The first group comprises parameter settings that are instrument dependent or a property of the acquired chromatograms (e.g. maximum output level of the mass spectrometer or average chromatographic peak width). The optimum values for these settings are easily determined and (usually) constant for all chromatograms of the data set. In contrast, other settings are less objective and must be optimised by trial-and-error. These include for example parameters that define real chromatographic peaks versus noise or window sizes in which peaks in two chromatograms are considered the same. Selecting these parameters correctly is not only a time-consuming process but it also involves a decision making by gut feeling or experience. So far, little has been published describing options for eliminating this subjective nature of parameter selection in data alignment.

In our opinion, the use of chromatograms in chromatographic profiling and metabolomics also requires a new way of assuring a good quality of the data, both during the recording and in the alignment step. The quality of data acquisition can be assessed by means of control samples. This part of the quality control is similar to that in classical target compound analysis, albeit that the test is more critical. Especially, the condition of retention time stability will be stricter. For the second part of quality control testing, i.e. assessing the quality of peak alignment, new criteria need to be defined. Whereas one control sample very often suffices for the first (chemical) part of the quality control procedure, more samples are needed to be able to judge the quality of an alignment operation. Only by using more samples from different individuals, inter-individual variation that might affect the alignment can be taken into account. A good alignment result is obtained if all peaks of the same origin are aligned, while peaks of different origin are not misaligned. The easiest way to check the quality of an alignment process is to use spiked samples of different individuals from the trial and monitor the alignment of the spiked compounds. Inclusion of the non-spiked analogues in the sample set and comparing the chromatograms pair-wise adds additional information on incorrect alignment in the other areas of the chromatogram. These regions should be identical in the aligned spiked and non-spiked chromatograms. All deviations are false positives, which might result from not properly aligned peaks or other sources of variation due to the chemical analysis. In this way, the settings resulting in the best alignment throughout the entire chromatogram of the set of control samples can be determined. Since the control samples are actually samples from the trial, the parameter settings

resulting in the best alignment of this representative control sample set will then also result in the best alignment of the entire sample set.

In this article, we will discuss a new strategy for the selection of optimum parameter settings for data alignment procedures in chromatographic sample profiling. We will define two quality indicators that allow an objective judgment on what constitutes a good alignment of a given data set. The basis for the evaluation is a balanced control sample set containing a given number of real samples from the trial analysed both spiked and non-spiked. Several sets of input parameters will be evaluated, and by means of the developed quality indicators, the optimum alignment settings are derived.

Theory

The quality of data alignment is determined by two factors: (1) Are spiked compounds aligned to the same retention time and (2) are there false positives? Both factors can be investigated by comparing the pair of the aligned spiked and non-spiked chromatograms of one control sample. Assuming optimal alignment, a residual chromatogram, obtained by the subtraction of the aligned non-spiked chromatogram from the aligned spiked chromatogram, should only contain peaks resulting from spiked compounds. All other peaks are false positives. To establish the values for the two quality indicators defined above, a procedure is needed to determine whether a peak in the residual chromatogram is a spiked compound or a false positive. A schematic representation of the approach is shown in Fig. 1. The starting point of the procedure is the peak list obtained from the given alignment software. Inputs for the alignment software were chromatograms of ten different individuals analysed both spiked and non-spiked (see “Control samples” for details) and the set of parameters for peak detection and alignment. The output peak list contains intensity values for each detected retention time/mass pair in all chromatograms. Next, for one sample pair (i.e. spiked and non-spiked) at a time, the residual chromatogram is calculated as described above. Peaks are evaluated in the residual chromatogram according to their intensity, starting with the most-intense peak. For each peak, a retention time range is defined as the retention time of the maximum \pm half a peak width. Since the retention times of the spiked compounds are known, it can be determined whether the detected peak originates from a spiked compound: Only if the previously established retention time of one spiked compound falls within this time window will the detected peak be considered a potential spike. In a next step, the masses of this spike found in the retention time window are compared to the

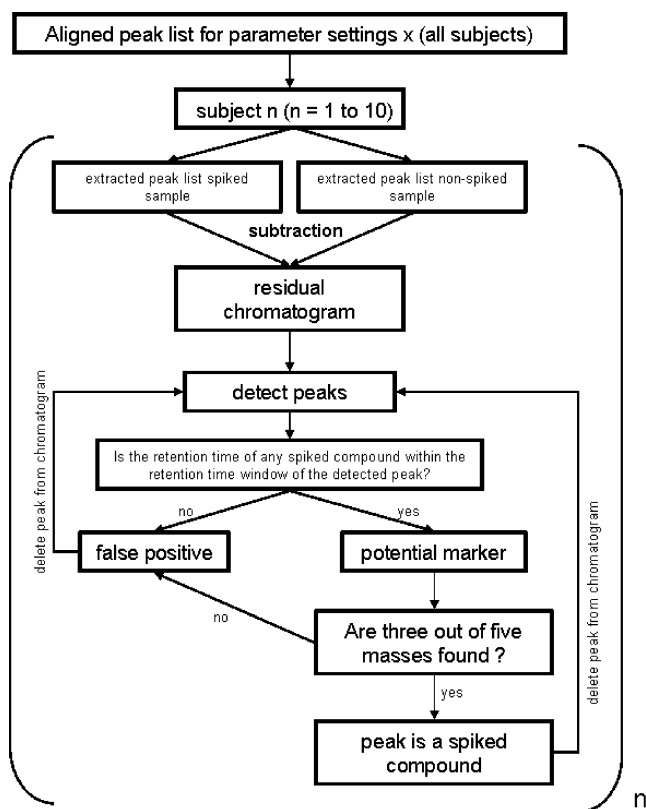


Fig. 1 Schematic overview of the strategy developed to determine the optimum parameter settings for peak alignment

mass spectrum (consisting only of the top five masses round to nominal values) of the spiked compound: Only if three out of five masses of the mass spectrum can be found in the mass list of the potential marker will it be considered to originate from this (spiked) compound. Otherwise, it is registered as a false positive. If no spiked compound can be found in the retention time window of the detected peak, it is registered as a false positive as well. The intensities of all detected peaks in the retention time window are now set to zero in order for them to not be evaluated again. This process is repeated until all 19 spiked compounds are found once in the chromatogram or with a maximum of 100 repetitions (selected as five times the number of spiked compounds). The number of times this process should be repeated also depends on the response level of the spikes. If the intensities are very low, the number of repetitions might have to be increased.

The whole process is repeated for all subjects, and the mean values of spiked compounds and false positives retrieved of all subjects are taken and collected for each set of parameter settings.

Plotting the residual chromatogram is a rapid qualitative method to visually assess the quality of a set of parameters. In contrast, calculating the above-mentioned two quality indicators is a more quantitative, objective measure for the

determination of the quality of alignment of a given data set and the suitability of the set of input parameters for alignment.

Instrumentation and methods

Control samples

The control set was prepared by a random selection of ten out of 150 urine samples from a nutritional intervention study where the bioavailability of polyphenols was studied (see Ref. [7] for details). The urines of these ten volunteers were split, and one part was spiked with a solution of 19 reference compounds. This resulted in two groups of samples for the control set: ten spiked urine samples and ten “blanks” (i.e. non-spiked urine samples). The identities of the 19 reference compounds used as spikes are given in Table 1, together with their retention times and the top five mass fragments used as a simplified mass spectrum. The reference compounds selected belong to the same compound class as the target compounds investigated in the trial.

Sample preparation and chromatographic analysis

Details on the preparation and chromatographic analysis of the samples can be found in a previous publication [7]. In short, the acidified urine samples were extracted three times by liquid–liquid extraction using ethyl acetate. The combined organic layers were evaporated to dryness and derivatised using *N,O*-bis[trimethylsilyl]trifluoroacetamide. The gas chromatographic analysis included a 1:20 split injection (1 μ L injection volume) and a temperature programmed separation from 45 to 300 $^{\circ}$ C at 3 $^{\circ}$ C/min. The column used was a VF-17 ms (30 m \times 0.25 mm, d_f =0.1 μ m) from Varian (Varian, Middelburg, The Netherlands). The gas chromatograph used was an Agilent 6890 (Agilent, Amstelveen, The Netherlands) with a Waters MicroMass GCT accurate-mass mass spectrometer (Waters, Etten-Leur, The Netherlands).

The internal standard used was *trans*-cinnamic acid-d6 (Sigma-Aldrich, Zwijndrecht, The Netherlands).

Software packages

Three software packages for data alignment were used in this study: MarkerlynxTM (an add-in to MasslynxTM, Waters (MA, USA)) [8], MZmine [9] and MetAlign [10], the two latter being freely available online. It is not within the scope of this article to explain in detail how the software packages work nor is it fully possible to give detailed explanations as the real algorithms are not always disclosed by the authors. However, one remark is appropriate. While all three

Table 1 Retention times, molecular weights and the five most dominant mass peaks of the 19 spiked compounds used in this study

Retention time [min]	Compound	Molecular weight after TMS derivatisation	Main mass fragments				
10.18	<i>m</i> -Toluic acid	208.1	65	91	119	193	208
11.23	3-Phenylpropionic acid	222.1	75	91	104	207	222
12.12	Mandelic acid	296.1	73	147	163	179	253
12.61	Salicylic acid	282.1	73	91	135	147	267
13.25	<i>Trans</i> -cinnamic acid	220.1	75	103	131	161	205
13.52	3-Hydroxybenzoic acid	282.1	73	193	223	267	282
14.14	3-Hydroxyphenylacetic acid	296.1	73	147	164	281	296
14.42	4-Hydroxybenzoic acid	282.1	73	193	223	267	282
15.97	2,3-Dihydroxybenzoic acid	370.1	73	75	147	193	355
16.27	2,6-Dihydroxybenzoic acid	370.1	73	75	147	267	355
16.97	<i>Trans</i> -2-hydroxycinnamic acid	308.1	73	147	161	293	308
17.05	2,4-Dihydroxybenzoic acid	370.1	73	147	281	355	370
17.13	3,4-Dihydroxybenzoic acid	370.1	73	193	281	355	370
17.2	3,5-Dihydroxybenzoic acid	370.1	73	75	147	355	370
17.90	<i>Trans</i> -3-hydroxycinnamic acid	308.1	73	203	219	293	308
18.72	<i>Trans</i> -4-hydroxycinnamic acid	308.1	73	219	249	293	308
18.96	3,4,5-Trihydroxybenzoic acid	458.1	73	75	281	443	458
19.25	2,4,6-Trihydroxybenzoic acid	458.1	73	75	147	355	443
21.17	Caffeic acid	397.1	73	75	219	381	396

packages are originally designed for LC-MS data sets, their applicability to GC-MS data should be feasible. Nevertheless, one aspect must be taken into consideration. In LC-MS, eluting compounds usually only give rise to a molecular ion peak with little fragmentation. In GC-MS, however, stronger fragmentation occurs, i.e. one compound is described by several fragments. In the peak lists obtained from the software packages, several detected retention time/mass pairs thus describe the same compound, which makes data interpretation slightly more complicated.

From their main principles, the three data alignment packages are based upon a similar alignment principle. However, whereas Markerlynx™ and MZmine make use of the accurate-mass dimension, at least if acquired by the instrument, MetAlign rounds to nominal masses. If wanted, the data can then be normalised, and/or further classification of the samples can be performed. Since all three packages include normalisation on a user-defined internal standard, it was opted to include the normalisation step in the processing in the software packages. All aligned data sets were then exported in the ASCII-format for further processing in Matlab 7.1 (The Mathworks, Natick, MA, USA).

Markerlynx™

The parameters to be set in Markerlynx™ pertain to the peak detection algorithm, alignment of detected peaks and the selection of the internal standard. Parameters related to

peak properties such as peak width or general settings such as the *m/z* range scanned by the mass spectrometer or the retention time window acquired by the gas chromatograph were set to fixed values. All other parameters were varied one at a time resulting in 11 sets of parameters that were evaluated using the procedure presented here.

To allow Markerlynx™ to detect the internal standard, its retention time and mass must be specified. In our case, the internal standard eluted at 13.21 ± 0.05 min and its main ion had a mass of 211.1 ± 0.1 Da.

MZmine

For peak detection, MZmine requires parameters that define peaks and noise. As in Markerlynx™, parameter settings concerning the chromatographic behaviour of peaks such as peak width were not varied during the optimisation procedure. For peak alignment, only the (recommended) option of the “fast aligner” was tested as the other alignment procedure resulted in frequent crashes of the computer. The main parameters to be set here define the window, both in the chromatographic and the mass dimension, in which two peaks in two chromatograms are considered to be the same.

The selection of the internal standard must be performed manually in the aligned peak list. The requirement is that the retention time/mass pair used for normalisation must be present in all samples in order to be used. The retention time of the pair used for normalisation was 792.762 s and

its m/z value was 211.107 Da. In MZmine, 13 sets of parameters were evaluated.

MetAlign

In MetAlign, instrument-dependent parameters concern the retention time region in the chromatogram to consider and the maximum amplitude of the mass spectrometer. The values for these parameters were taken from the data acquisition software. These parameters were not varied throughout the study. The internal standard is defined by its (nominal) mass and its scan number (211 Da and scan 1082) that can be readily obtained from the chromatograms. For peak detection, three parameters are required of which one, the average peak width, was not varied as it is rather constant throughout the chromatograms. Two alignment strategies can be selected, “rough” or “iterative”, with the latter requiring the user to select input parameters for the calculation of the chromatographic shift profiles. In our case, both options were tested, and altogether, nine sets of parameters were tried in MetAlign.

Data processing of the aligned data sets

As described in the “[Instrumentation and methods](#)” section, the control sample set used for the evaluation of the parameter sets comprised 20 samples, i.e. ten pairs of a spiked urine sample and its non-spiked equivalent (i.e. the ten sample pairs each differ only by the 19 added reference compounds).

For one set of parameter settings, the data set contains intensity values for all detected retention time/mass pairs of the 20 samples. The number of detected retention time/mass pairs depends on the peak detection parameters applied and varies with varying settings. Depending on the size of the obtained data sets and the power of the computer used for the calculations, the data may need to be reduced. In our case, it was opted to only allow retention times between 10 and 22 min. This was possible since it was known that all compounds of interest elute within this time window. By reducing the data this way, the number of detected retention time/mass pairs was between 1,000 and 30,000, depending on the software package used and the selected parameter settings.

Results and discussion

Proper selection of the input parameters for the alignment software is essential to obtain optimum alignment results in chromatographic profiling experiments. Only identical peaks should be aligned to the same retention time. Using spiked samples, only the correct alignment of the spiked compounds can be investigated. While this is important information, all information on the quality of alignment in

other areas of the chromatogram is missed, and the argument on the overall quality of the alignment is less strong. When including the non-spiked samples as well, information is obtained on the alignment of the spiked compounds as well as on incorrect alignment in other areas.

The nature of the spiked compounds should be related to the analytical question of the sample set. In most cases, one or more classes of compounds are of interest (e.g. amino acids or fatty acids). The spikes should be of the same class (es), and one should ensure that the intensities vary. This can be done by spiking various concentrations or by testing the response factors of the spiked compounds. When only high-intense spikes are used, the algorithm is not so robust as the quality of alignment may decrease for less intense peaks. Another factor that is related to the intensity levels is the number of repetitions of the procedure shown in Fig. 1. For our sample set, the two quality indicators were determined, repeating the procedure until all spiked compounds are found or after 100 repetitions. This number was selected as five times the number of spikes. Since the algorithm detects peaks according to intensities (starting with the most-intense peak), this number might have to be increased in order to ensure that the non-detection of low-intense spikes is due to poor alignment and not due to the low intensities.

The residual chromatogram as defined in “[Theory](#)” can be a useful tool in determining the quality of alignment. If no variation (e.g. from the instrument or the sample preparation) is present in the data, a residual chromatogram of one subject should only contain retention time/mass pairs originating from the spiked components. The optimum parameter settings thus result in a set of residual chromatograms containing the maximum number of pairs that originate from the spiked compounds, while detecting the lowest number of false positives. Figure 2 gives an example of chromatograms reconstructed from the aligned peak list of one subject and the resulting residual chromatograms for two sets of parameter settings: one resulting in good peak alignment (Fig. 2a and b) and one set resulting in poor peak alignment (Fig. 2c and d). The reconstructed chromatograms in Fig. 2a and c show an overlay of the spiked (purple) and non-spiked (black) samples by plotting the intensity values versus the respective retention time. Figure 2b and d represent the respective residual chromatograms. Note that only the chromatograms of one sample, spiked and non-spiked, are shown, whereas for the determination of the number of spikes and false positives found, an average of all samples is taken. In case of good alignment (Fig. 2a and b), the residual chromatogram is dominated by peaks originating from spikes. The spikes named A to G in Fig. 2a (purple chromatogram) can all be found back in the residual chromatogram (Fig. 2b). If the parameter set for peak detection and alignment is poor (Figs. 2c and d), not all spikes can be identified, and the number of false positives in

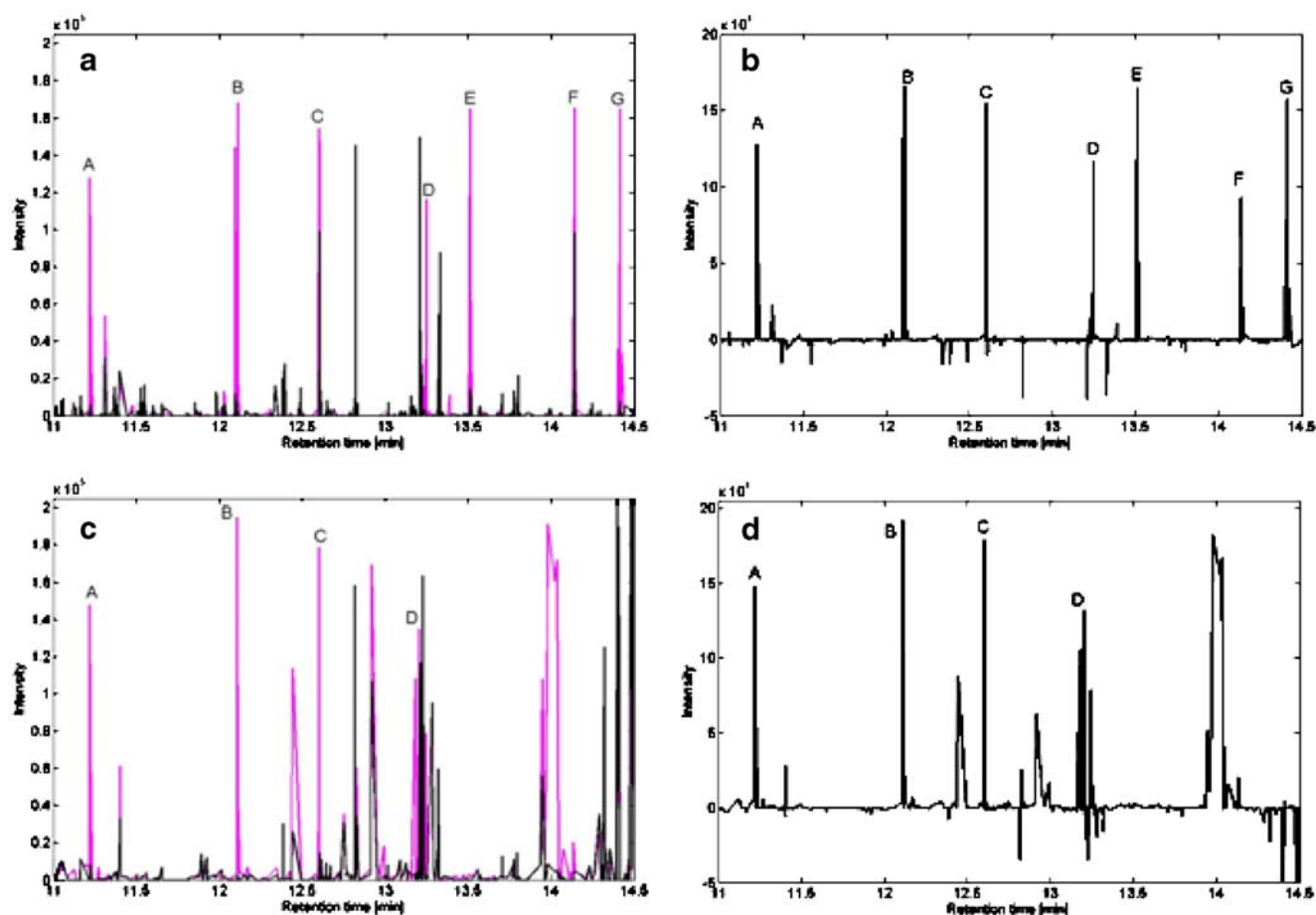


Fig. 2 Reconstructed chromatograms obtained with two sets of parameter using Markerlynx: one set resulting in good peak alignment (a) and one resulting in poor alignment (c). The spiked (purple) and non-spiked (black) chromatograms are overlaid in a and c, whereas

their respective residual chromatograms are shown in b and d. These representative settings are taken from sets 5 and 11 using Markerlynx for peak alignment (see Table 2 for details)

the residual chromatogram is higher (see Table 2). Negative peaks in the residual chromatograms mean that the intensity at this retention time was lower in the spiked sample than in the blank. These peaks can also be considered as false positives, meaning that the input data for the algorithm used to establish the two quality indicators must be investigated as absolute values.

A point of discussion of the proposed strategy remains the window in which retention times are included for finding potential markers as well as the criterion of having an agreement of three out of five masses between potential marker and spiked compound. For our data set, however, these criteria have shown to be reasonable.

Selection of optimum parameter settings in the alignment software packages

For a given software package, the starting values of the parameter settings were the ones given by the authors. Parameters that were not instrument or method dependent

were then systematically varied, resulting in as many data sets as the number of times the parameter settings were changed. This resulted in 11 sets for Markerlynx™, 13 sets for MZmine and nine sets for MetAlign.

Table 2 gives all sets of parameters used to detect and align peaks in Markerlynx™. For all 11 settings, a different number of peaks are detected, and a varying quality of the alignment is obtained. Evidently, the number of detected peaks is largely dependent on the noise threshold. Nevertheless, the number of detected peaks does not significantly influence the alignment result as a comparable number of spikes are found back, and little variation is found for the number of false positives as defined here.

Varying the parameter “number of retained masses per retention time” (see sets 2, 4 and 5) has little influence on the number of detected peaks, but does strongly influence the alignment result: If only very few masses are retained (e.g. five), only 3.7 out of 19 spiked compounds are found back on average, compared with 18.4 when 100 masses are collected (set 5). When the minimum intensity of a mass

Table 2 Sets of parameters applied in Markerlynx, MZmine and MetAlign and the number of detected peaks as well as number of markers and false positives found

Settings	1	2	3	4	5	6	7	8	9	10	11	12	13
Markerlynx													
Mass tolerance (abs)	0.01	0.01	0.01	0.01	0.01	0.01	0.1	1	0.1	0.1	0.1		
Noise elimination level	1	6	20	6	6	6	6	6	6	6	6		
Masses per retention time	50	50	50	5	100	50	50	50	50	50	50		
Minimum intensity (%)	1	1	1	1	1	10	1	1	1	1	1		
Mass window (Da)	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.5	1	0.5		
Retention time window	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1		
Number of peaks	13,296	4,920	2,781	2,434	2,882	1,925	4,415	4,350	2,596	2,256	1,398		
Number of spikes found	18.4	18	18.3	3.7	18.4	6.9	18.8	18.4	17.7	17.8	8.1		
Number of false positives	58.6	70.9	55.7	91	58	93.1	47.1	58.3	82.3	82.2	91.9		
MZmine													
<i>m/z</i> bin size (Da)	0.25	0.25	0.25	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Noise level	4	16	0.5	4	4	4	4	50	4	4	4	4	4
Tolerance for <i>m/z</i> variation (Da)	0.05	0.05	0.05	0.05	0.5	1	0.05	0.5	0.5	0.5	0.5	0.5	0.5
Tolerance for intensity variation (%)	20	20	20	20	20	20	50	20	10	20	20	20	20
Balance between <i>m/z</i> and RT	10	10	10	10	10	10	10	10	10	40	5	10	10
<i>m/z</i> tolerance size	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1	0.2
RT tolerance size (%)	1	1	1	1	1	1	1	1	1	1	1	1	10
Number of variables	7,450	4,171	8,251	7,450	7,830	8,535	7,450	1,551	6,089	6,083	6,091	5,868	5,127
Number of spikes found	19	19	19	19	19	19	19	18.3	19	19	19	19	18.7
Number of false positives	6	13.7	7.1	6	5	5.3	6	55.8	3.8	3.5	4	3.9	51.1
MetAlign													
Peak slope factor	1	5	1	1	1	1	1	1	1				
Peak threshold factor	2	2	10	2	2	2	2	2	2				
Regions	100/200	100/200	100/200	10/20	10/20	10/20	10/20	10/20	10/20				
Alignment	Rough	Rough	Rough	Rough	It.	It.	It.	It.	It.				
Maximum shift per 100 scans	–	–	–	–	40	60	40	40	40				
Minimum factor	–	–	–	–	10/5	10/5	50/5	50/5	10/5				
Minimum number of masses	–	–	–	–	4/2	4/2	4/2	8/2	8/2				
Number of variables	19,268	5,879	6,172	27,913	33,336	33,343	33,315	33,308	33,291				
Number of spikes found	19	19	19	19	19	19	19	19	19				
Number of false positives	7.8	9	7.8	3.2	3.7	3.7	10.4	10.4	3.7				

peak to be included is changed from the recommended 1% of the height of the base peak to 10% (sets 2 and 6), less peaks are detected as expected, but unfortunately, also the quality of alignment is tremendously decreased. Using the parameter settings of set 6, only 6.9 markers are found on average (in comparison to 18 in set 2). Also, more false positives are included.

For comparison of the peak list obtained for a sample with the detected peaks of another (reference) sample, two windows are used: the mass window and the retention time window. When the mass window is increased from 0.05 to 1 Da (sets 7, 9, 10), the number of detected markers is decreased, but more importantly, many more false positives are included in sets 9 and 10. Increasing the retention time

window from 0.2 to 1 results in only 8.1 markers being found back (see sets 9 and 11).

Varying the mass tolerance parameter (see sets 2, 7 and 8) neither has a great influence on peak detection nor on the quality of the alignment, with the intermediate value of 0.1 resulting in a slightly better alignment result (18.8 out of 19 markers found on average versus 18).

The (peak detection and alignment) algorithms in MZmine are more robust, and only extreme changes from given standard values have a strong influence on the alignment result (see Table 2). With all markers being found in all settings, the number of false positives becomes more important. With most settings, it averages between 3 and 7, and only for some sets did the number increase

substantially. This was particularly the case when the noise level was set to extreme values or the tolerance size for the retention times in the alignment settings was set extensively large (e.g. 10% in set 13). An important problem with MZmine is the possible misalignment of the internal standard. Since it is a prerequisite that the internal standard must be present in all samples with the same retention time/mass pair, other less selective ions than the most specific ion may need to be used for normalisation or no normalisation on an internal standard may be possible. In our case, when the noise level was set too high, the characteristic ion of 211.1 Da could not be found in all samples, and the second, less-characteristic ion of 73.1 Da was used for normalisation for this set. For all other sets, however, the ion of 211.1 Da could be used.

In MetAlign, most parameter settings are defined by the nature of the chromatogram or the instrumental method. Varying the peak slope factors or peak threshold factors around reasonable values did not change the peak detection and alignment outcome. Most influence on the alignment result had the maximum shift allowed of the window for peak search. If chosen too high, the number of false positives is increasing. Nevertheless, the actual number still is rather low, being decreased from around eight to around three when allowing a maximum of ten to 20 scans. Varying the other alignment parameters did not change the alignment results much with all markers being found, and only small changes in the number of detected false positives. This is of course only true if varied around reasonable values from a chromatographic point of view. The different sets of settings tried in MetAlign and the alignment results obtained can be found in Table 2.

For our data set, we found that the selection of the parameter settings is rather robust for all packages and that slight variations around reasonable values did not markedly influence the quality of the alignment. However, it cannot always be predicted from theory which set of parameters will result in optimum alignment of the data set and a small set of combinations needs to be evaluated. The assessment of around ten sets of settings should be sufficient to select the optimum parameter settings in an objective way. The starting values as proposed by the authors for LC-MS data sets were adequate for our data set, even though it was obtained by GC-MS. Using the strategy developed here, the optimum settings for alignment of chromatographic data sets can be obtained in a systematic and objective manner. A similar approach might also be applicable to spectroscopic data, though the investigation of this was not within the scope of this article.

When comparing the output of the three packages, MetAlign results in by far the largest number of detected peaks, suggesting a lot of noise still being present in the data set. Using our defined criteria for analysing the quality

of alignment, the large number of detected peaks did not negatively influence the end result. On the contrary, all markers were detected in all sets of parameters tested, and on average, the number of false positives was very low. In Markerlynx, however, this number was very large even though the numbers of detected peaks in the various sets are limited. In addition, for some sets, only very few markers were found back. The number of detected variables could, however, compose a problem for a common personal computer, and data reduction may need to be applied in order to obtain data of a workable size.

Theoretically, an infinite number of combinations for sets of parameters are possible per package. In practice, the computational time to process one set is around 4 to 6 h, and therefore, the number of sets of parameters that can be evaluated is limited. The requirement for user input during the computations further restricts the number of sets of parameters that can be assessed. In Markerlynx™, the user is only required to enter the values in the beginning, and no further user input is necessary. For MZmine, first, the user needs to load and read the data, then define the peak detection parameter and in a third step, align the data. If required, the marker for the internal standard must be found and selected manually from the aligned peak list and normalisation is then performed. MetAlign first performs baseline correction using a user-defined parameter, which is the most time-consuming process. The second step, in which the user has to define the internal standard and the alignment parameter, is rather fast.

Conclusion

Chromatographic fingerprinting as e.g. used in metabolomics experiments usually requires the use of multivariate techniques in order to obtain useful information. An important prerequisite in the analysis is that the chromatographic peaks are aligned.

An objective assessment of the quality of an alignment operation can be derived from a set of appropriately selected control samples. These control samples are routinely analysed within a chromatographic experiment. In contrast to the multiple analysis of one control sample as performed in targeted approaches, control samples from a representative number of individuals in a trial need to be included for this type of quality control. In addition, they must be analysed both spiked and non-spiked. The obtained new (control) data set is then used for the determination of the optimum values for peak alignment in the respective software packages.

The strategy of investigating the number of retrieved spiked compounds together with the number of false positives is a suitable tool to select the settings which result in the best alignment of all chromatograms.

Acknowledgement Arjen Lommen (RIKILT, Institute for Food Safety, Wageningen, The Netherlands) is gratefully acknowledged for the detailed discussions on the principles of MetAlign.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Pravdova V, Walczak B, Massart DL (2002) *Anal Chimica Acta* 456:77–92
2. Tomasi G, van den Berg F, Andersson C (2004) *J Chemometr* 18:231–241
3. van Nederkassel AM, Daszykowski M, Eilers PHC, Vander Heyden Y (2006) *J Chromatogr A* 1118:199–210
4. Johnson KJ, Wright BW, Jarman KH, Synovec RE (2003) *J Chromatogr A* 996:141–155
5. Frenzel T, Miller A, Engel K-H (2003) *Eur Food Technol* 216:335–342
6. Katajamaa M, Orešič M (2007) *J Chromatogr A* 1158:318–328
7. Grün CH, van Dorsten FA, Jacobs DM, Le Belleguic M, van Velzen EJJ, Bingham MO, Janssen H-G, van Duynhoven JPM (2008) *J Chromatogr B* 871:212–219
8. Waters (2008) Company website. Waters, Mass., USA. <http://www.waters.com>, last accessed 22 Sept 2008
9. Orešič M, Katajamaa M (2007) <http://mzmine.sourceforge.net>, last accessed 23 May 2008
10. Lommen A (2008) www.rikilt.wur.nl/UK, last accessed 23 May 2008