Tutorial

Correction of shifts in single-stage LC-MS(/MS) data

Mitra, V.; Smilde, A.K; Bischoff, R.; Horvatovich, P.

DOI
10.1016/j.aca.2017.09.039

Publication date
2018

Document Version
Final published version

Published in
Analytica Chimica Acta

Analytica Chimica Acta 999 (2018) 37–53

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Review

Tutorial: Correction of shifts in single-stage LC-MS(/MS) data

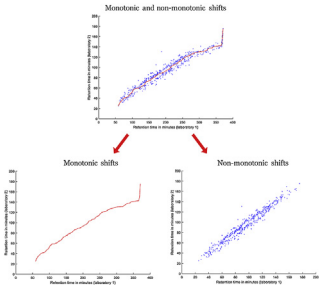
Vikram Mitra [a], Age K. Smilde [b], Rainer Bischoff [a], Péter Horvatovich [a, *]

[a] Analytical Biochemistry, Department of Pharmacy, University of Groningen, A. Deusinglaan 1, 9713 AV Groningen, The Netherlands
[b] Swammerdam Institute for Life Science, University of Amsterdam, the Netherlands, Science Park 904, 1098 XH Amsterdam, The Netherlands

HIGHLIGHTS

- Single-stage LC-MS data (MS1 map) should be comparable for accurate quantification.
- Comparable MS1 maps can be accurately corrected with single monotonic function.
- Monotonic and non-monotonic shifts exist jointly between MS1 maps.
- Monotonic shift can be corrected, non-monotonic shift cannot be corrected.
- Non-monotonic shift affects the quality of quantitative LC-MS(/MS) pre-processing.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 10 December 2016
Received in revised form
26 September 2017
Accepted 27 September 2017
Available online 2 November 2017

Keywords:
Shift correction
Retention time alignment
Label-free quantification
Orthogonality

ABSTRACT

Label-free LC-MS(/MS) provides accurate quantitative profiling of proteins and metabolites in complex biological samples such as cell lines, tissues and body fluids. A label-free experiment consists of several LC-MS(/MS) chromatograms that might be acquired over several days, across multiple laboratories using different instruments. Single-stage part (MS1 map) of the LC-MS(/MS) contains quantitative information on all compounds that can be detected by LC-MS(/MS) and is the data of choice used by quantitative LC-MS(/MS) data pre-processing workflows. Differences in experimental conditions and fluctuation of analytical parameters influence the overall quality of the MS1 maps and are factors hampering comparative statistical analyses and data interpretation. The quality of the obtained MS1 maps can be assessed based on changes in the two separation dimensions (retention time, mass-to-charge ratio) and the readout (ion intensity) of MS1 maps. In this tutorial we discuss two types of changes, monotonic and non-monotonic shifts, which may occur in the two separation dimensions and the readout of MS1 map. Monotonic shifts of MS1 maps can be corrected, while non-monotonic ones can only be assessed but not corrected, since correction would require precise modelling of the underlying physicochemical effects, which would require additional parameters and analysis. We discuss reasons for monotonic and non-monotonic shifts in the two separation dimensions and readout of MS1 maps, as well as algorithms that can be used to correct monotonic or to assess the extent non-monotonic shifts. Relation of non-monotonic shift with peak elution order inversion and orthogonality as defined in analytical chemistry is discussed. We aim this tutorial for data generator and evaluators scientists who aim to known the condition and approaches to produce and pre-processed comparable MS1 maps.

* Corresponding author.
E-mail address: p.l.horvatovich@rug.nl (P. Horvatovich).

https://doi.org/10.1016/j.aca.2017.09.039

Contents

Abbreviations

AMT	Accurate Mass Tag
CE	Capillary electrophoresis
CODA	Component Detection Algorithm
COW	Correlation Optimized Warping
DTW	Dynamic Time Warping
GC×GC	2 dimensional gas chromatography
GC-MS	Gas chromatography coupled to mass spectrometry
ICAT	isotope-coded affinity tags
ICPL	Isotope-Coded Protein Labelling
MS1	single stage part of LC-MS(/MS) data
(s)PTW	(semi)Parametric Time Warping
SILAC	Stable Isotope Labelling by Amino acids in Cell culture
SIMA	SImultaneous Multiple Alignment
SPC	Spectral Counting
TIC	Total Ion Chromatograms

1. Introduction

Over the past decade LC-MS(/MS) technology has been routinely used in proteomics and metabolomics laboratories to analyse complex biological samples [1,2]. However, to understand system level perturbations and molecular mechanisms of biological events and diseases, quantitative values of biomolecules are required to determine which compounds show differential levels between sample groups [3–6].

The non-fragmented single-stage part (MS1) of LC-MS(/MS) data is described with two separation dimensions such as mass to charge ratio (*m/z*) and retention time (*rt*) and one readout (ion intensity *iin*), which data is considered as second order tensor obtained from second order analytical instrument (Fig. 1) [7]. MS1 data contains signals from all compounds that can be detected by an LC-MS(/MS) system and is the signal of choice used by label-free quantification approach[8,10,11]. MS1 signal is used for quantification for stable-isotope chemical labelling methods, which provide sample specific signal in MS1 domain such as stable isotope labelling by amino acids in cell culture (SILAC) [12], isotope-coded affinity tags (ICAT) [13] and isotope-coded protein labelling (ICPL) [14]. In ideal case, *rt* and *m/z* coordinates of one compound would not differ in MS1 maps facilitating the identification of MS1 signal of identical compounds using these coordinates independently to their identification status and intensity of compounds in multiple MS1 maps. However, these coordinates in the MS1 maps are not constant and are subject to variation. Examples for these variations include those correctable by a single monotonic function and those that are not correctable by such a function. These variabilities in conjunction with local compound density and other signal processing parameters such as the presence of chemical and electronic noise should be taken into account by the quantitative LC-MS(/MS) data pre-processing workflow to provide accurate quantitative tables with columns (or rows) corresponding to samples and rows (or columns) to compounds, which data are used subsequently for statistical evaluation. The *iin* readout may include variations for example as a result of differences in the injected sample amount due to variation in the quantity of all or of a subset of compounds varying in intensity in one batch. Also, ion suppression effects may reduce or increase the *iin* readout value of the compounds affected by it. These variations affect the quality of the quantitative table obtained upon LC-MS(/MS) pre-processing and ultimately the statistical outcome of biomarker discovery or differential expression analysis.

The minimal MS1 data pre-processing workflow includes only modules for peak detection and matching and assumes no shift in the *rt* and *m/z* dimensions and in the *iin* readout of MS1 data (Fig. 2a). Typical quantitative MS1 LC-MS(/MS) data pre-processing (Fig. 2b) consists of modules for data format conversion, raw data resampling in retention time and *m/z* dimensions, denoising, correction for background ion intensity, peak detection and quantification followed by correction of shifts occurring in each of the *rt* and *m/z* dimensions and the *iin* readout of the MS1 data. Algorithms, which corrects for shifts in the *rt* domain are named retention time alignment methods, algorithms that correct shifts in the *m/z* domain are called mass (re)calibration and algorithms that corrects “shifts” in the *iin* readout are classified as normalisation approaches. The term “shift” in the *iin* readout cannot be interpreted similarly to the separation dimensions, but similar phenomena can be observed e.g. when total amount of injected sample differs. Correction of “shifts” can be treated mathematically similarly to those of separation dimensions. The final step after correction of shifts in the two separation dimensions of MS1 is the peak matching step, which identifies the MS1 information of identical compounds in multiple LC-MS(/MS) chromatograms based on *m/z* and *rt* coordinates, by matching based on an identified peptide sequence or based on the similarity between MS/MS spectra in data-dependent acquired LC-MS/MS data. All these steps

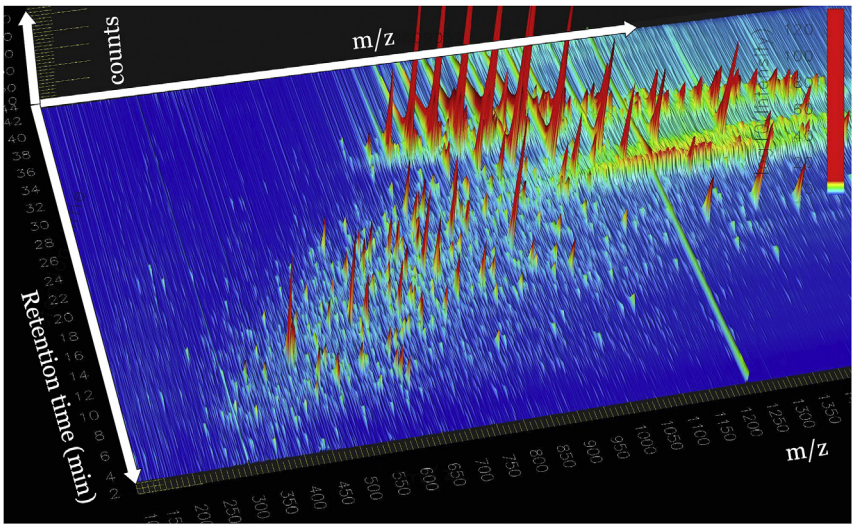


Fig. 1. The *rt* and *m/z* dimensions and *iin* readout of a single-stage LC-MS data (MS1 map). The dimensions are mass-to-charge ratio (*m/z*), retention time (*rt*) and ion intensity (*iin*) readout. Chromatographic pairs can show monotonic shift and non-monotonic shift with orthogonality component, where monotonic shift can be corrected, while the remaining non-monotonic shifts including orthogonality determines the uncertainty to find corresponding peaks in the chromatograms using *rt* and *m/z* dimensions. Orthogonality in *iin* readout leads to statistical bias and increase false discovery in statistical differential analysis.

are required prior to statistical evaluation and are implemented in automated data pre-processing pipelines [15–20]. One of the most critical steps is the accurate correction of shifts in the *rt* and *m/z* dimensions and in the *iin* readout of MS1 maps. Improper correction of shifts may lead to inaccurately matched peaks and to quantification bias which may ultimately lead to inappropriate conclusions after the statistical analysis. Presence of such error is often only recognized much later during experimental validation of the original biomarker discovery results contributing to irreproducibility of biological and preclinical studies and leading to loss of analysis time, research effort and resources [21].

In this tutorial, we focus on the LC-MS(/MS) analysis conditions, which results in comparable MS1 LC-MS(/MS) maps and on algorithms which are able to accurately pre-process the obtained MS1 maps. This paper restricts the discussion of LC-MS(/MS) pre-processing with respect to sources of variability, variability assessment and their correction approaches used between MS1 maps of LC-MS(/MS) data. Special attention is devoted to discuss the physico-chemical origins and algorithmic treatment of correctable (monotonic) and non-correctable non-monotonic shifts in *m/z* and *rt* separation dimensions, and in the *iin* quantitative readout. This tutorial is aimed for experimental scientists planning molecular profiling experiments, aiming to generate MS1 data that can be pre-processed accurately, as well for bioinformaticians, who are developing new algorithms for LC-MS(/MS) data pre-processing and quality control.

2. Accurate alignment of single-stage LC-MS(/MS) data

2.1. Definitions and statements

In order to avoid confusion and facilitate the reading of the article we define here terms that will be used throughout the manuscript. **Single-stage LC-MS(/MS) or MS1 dimensions**: dimension definition is used both for the separation (*rt* or *m/z*) and for the readout (*iin*) variables of MS1 map. **Monotonic shifts**: monotonic shifts are differences (fluctuations) of values in one of the *rt* and *m/z* dimensions or in the *iin* readout of MS1 map pair of the same compounds (for *rt* and *m/z* dimensions) or the same compounds with the same quantity (*iin* readout) that can be corrected using a monotonic function. **Non-monotonic shifts**: is the differences (fluctuation) of values in one of the *rt* and *m/z* dimensions or in *iin* readout of MS1 map pair of the same compounds (for *rt* and *m/z* dimensions) or the same compounds with the same quantity (*iin* readout), which remains after correction with monotonic shift. Monotonic and non-monotonic shifts are always defined in the same dimension (or readout) of MS1 map pairs i.e. between *m/z*, *rt* or *iin*. **Orthogonality**: Orthogonality has many definitions in different science disciplines. In mathematics, algebra defines orthogonality of two vectors, which have dot product of zero. More general definition of orthogonality relates to synonyms such as independence, non-correlated or non-overlapping properties. Analytical chemistry uses the term orthogonality to measure the similarities and differences of two separation systems e.g. in liquid chromatography. Camenzuli et al. defines the orthogonality measure of two chromatographic separations as characteristics, which describes the degree of independence of two separation systems [22]. Gilar et al. provided similar but more practical definition of orthogonality as characteristics, which defines orthogonality as the joint peak capacity of two chromatographic system evaluated by occupancy percentage of bins with the same compound in the complete peak capacity space [23]. The analytical chemistry definition of orthogonality allow to interpret smaller and larger orthogonality differently from the algebraic binary definition, where two vector are either orthogonal or not. There are different metrics for orthogonality reported in the literature of analytical chemistry [22–25] and each of them refer to the fraction of area occupied by common compounds in the separation space of two chromatographic systems. These metrics can take values between 0 and 1, where 0 means two equivalent, and 1 reflects two fully independent separation systems. Since orthogonality is assessed using the common compounds therefore its value is dependent not only from the separation dimensions, but also from chemical space of the analysed compounds. We interpret orthogonality following the analytical chemistry's definition.

2.2. Conditions for correcting shifts

MS1 data has two separation dimensions (*m/z* and *rt*) and one readout (*iin*) as described in the introduction. Quantitative

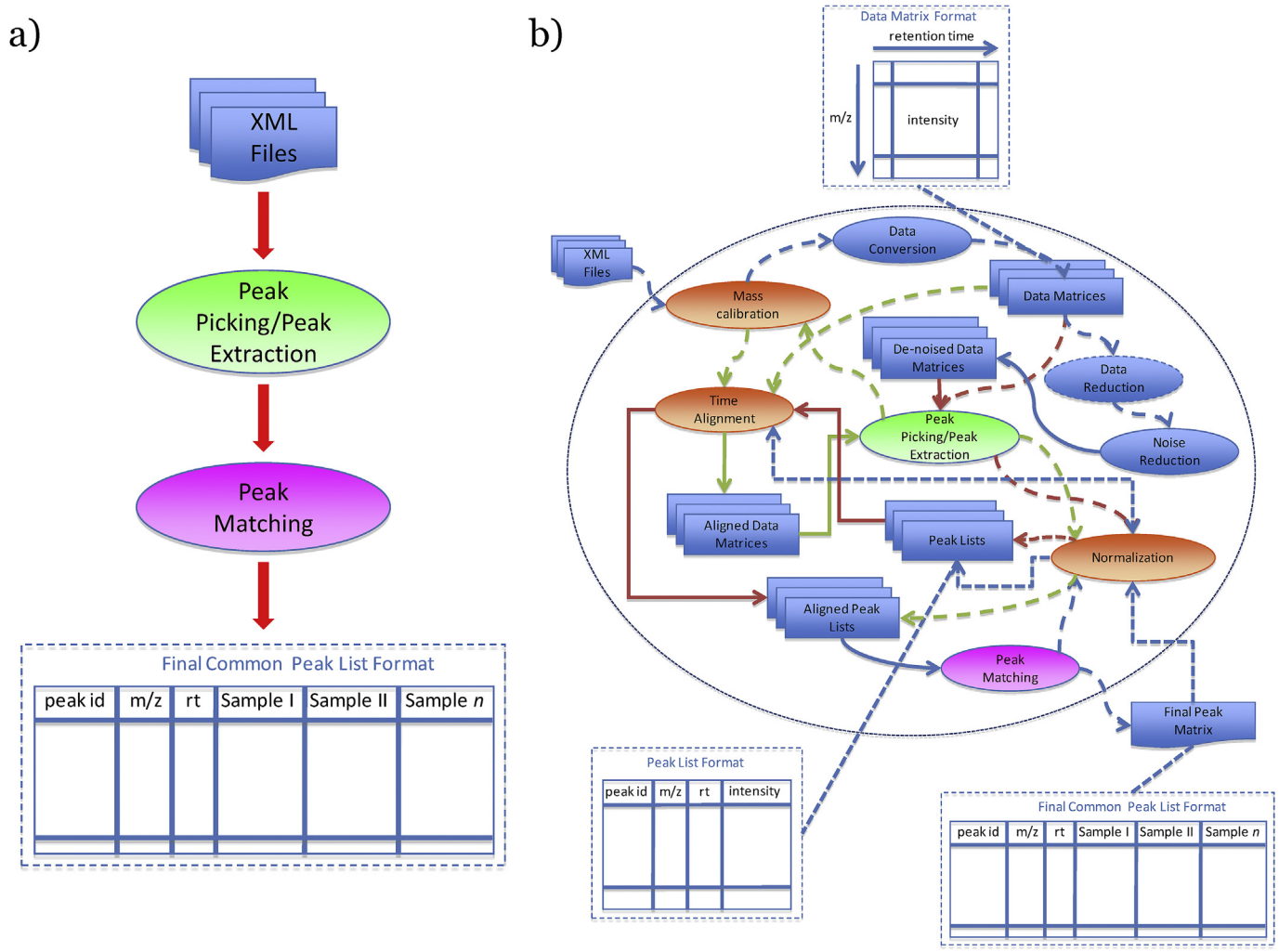


Fig. 2. Scheme of a) minimal and b) optimal label-free MS1 data pre-processing workflows. Two modules are required for minimal workflow, which includes peak detection/quantification modules (green) and module that matches the corresponding peaks across multiple chromatograms (purple). The minimal module assumes no monotonic shift and orthogonality in *rt*, *m/z* dimensions and *iin* readout. The optimal workflow implements modules for correction to monotonic shifts in the *rt* and *m/z* separation dimensions and in the *iin* readout of MS1 map corresponding to time alignment (correction in *rt*), to mass (re)calibration (correction in *m/z*) and to normalization (correction to *iin*). Other modules such as noise, data reduction, and resampling are additional modules of the workflow. Although not present in current pipelines, orthogonality assessment and modelling module e.g. by use of retention time prediction or feature decharging algorithms may add additional precision for LC-MS(/MS) data pre-processing workflow. The result of LC-MS(/MS) pre-processing is a quantitative table of compounds detected in multiple chromatograms serving as input for differential statistical analysis. Scheme b) was adopted from Christin et al. [39]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

information of compounds in MS1 data is represented as 3-dimensional Gaussian (or Lorentzian) peaks, where *iin* is the extent of the peak while *rt* and *m/z* represent the location of the peak maxima. The distinction between *iin* readout and the *m/z* and *rt* dimensions is reflected by the role of these variables. *m/z* and *rt* characterise the peak capacity of the analytical system and are related to the physicochemical properties of a compound, while the quantity of a compound is expressed in the *iin* readout, which is the main interest of the subsequent quantitative statistical analysis. Algorithms correcting for shifts are generally applied to LC-MS(/MS) chromatographic pairs, but some approaches perform alignment of the complete dataset in one step such as the Continuous Profile Model [26,27]. This method assumes one common underlying molecular profile, to which all chromatograms are aligned using a hidden Markov model [26]. In pairwise alignment, generally the MS1 coordinate of the raw data or feature list in one chromatogram (often called sample chromatogram) is corrected to the other non-altered chromatogram considered to be the reference. In this tutorial we discuss pairwise alignment of MS1 maps approaches but similar conditions apply for methods that align the complete data set in one step. Shifts in two separation dimensions and readout of MS1 map may occur, and these shifts have a physicochemical and/or instrumental cause or originate as error of LC-MS(/MS) data pre-processing. In *rt* and *m/z* dimensions and in the *iin* readout of MS1 map, monotonic shifts can be corrected when the following conditions are met:

1. Sample chromatograms should contain common compounds for alignment in the *m/z* and *rt* dimension, while for normalization (correction in the *iin* readout) the samples should contain common compounds with the same quantity in the chromatographic pairs.
2. The alignment algorithm should identify an adequate number of common peaks accurately for alignment in *rt* and *m/z* dimensions, while the *iin* readout (normalisation) should identify common compounds that are present in the same quantity in sufficient numbers and in sufficient distribution in the range of interest, which allows accurate alignment.

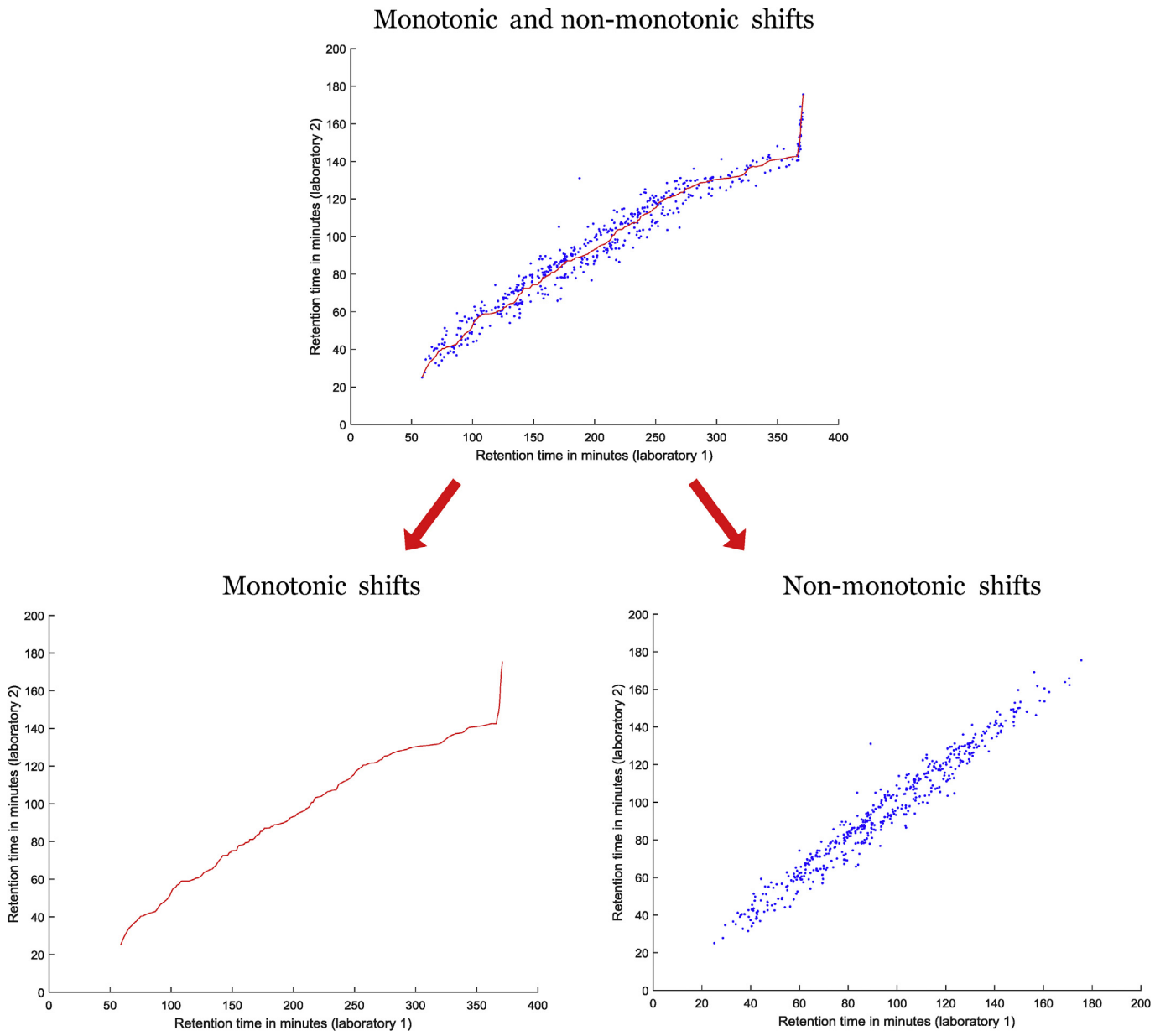


Fig. 3. Monotonic and non-monotonic shifts in MS1 data. Mixing of monotonic (red line) and non-monotonic shifts in the scatter plot of the retention time of identified peptides (blue dots) matched based on agreement of the identified primary amino acid sequence. The data originate from same trypsin digested porcine cerebrospinal fluid sample analysed in two different laboratories using different eluent programs and LC-MS/MS platforms (QTOF and Orbitrap). The upper plot shows the original retention time of peptides, which includes perturbations that are due to monotonic and non-monotonic shift in the liquid chromatography separation. The lower left plot shows the monotonic retention time correction function, which can be used to remove correctable monotonic shift from the raw data. The lower right plot shows the scatter plot of the retention time of identified peptides after correction with monotonic retention time correction function. The remaining fluctuation of peptides reflect the non-monotonic shift, which includes orthogonality of the liquid chromatography separation and shows the uncertainty to found corresponding compounds based on *rt* and *m/z* coordinates in other chromatograms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Common compounds should follow the same order in both chromatograms for *m/z* and *rt* dimensions. In *iin* readout, the order of ion intensity of the common compounds present in the same quantity should be the same in the two chromatograms.

It is important to note that accurate single monotonic correction function applied to all compound cannot be derived if one or more of these conditions are not met. It is the common compounds (in *rt* and *m/z* dimension) and the common compounds that are present in the same quantity (in *iin* readout) in the two chromatograms that convey the information, that should be used to derive the single monotonic correction function. After obtaining the correction function, all *rt*, *m/z* and *iin* values of the other compounds will be corrected with the derived correction function. The requirement that common compounds should have the same quantity in the two chromatograms for alignment in *iin* is due to the fact that detector response and ion suppression/competition effects may be different at different concentration ranges. In fact the condition of having the same compounds in the same quantity can be seen to be too restrictive compared to requirement of known quantity. However, in *iin* readout the signal of compounds may be affected by the other compounds present e.g. due to ion-supression, while this coupling

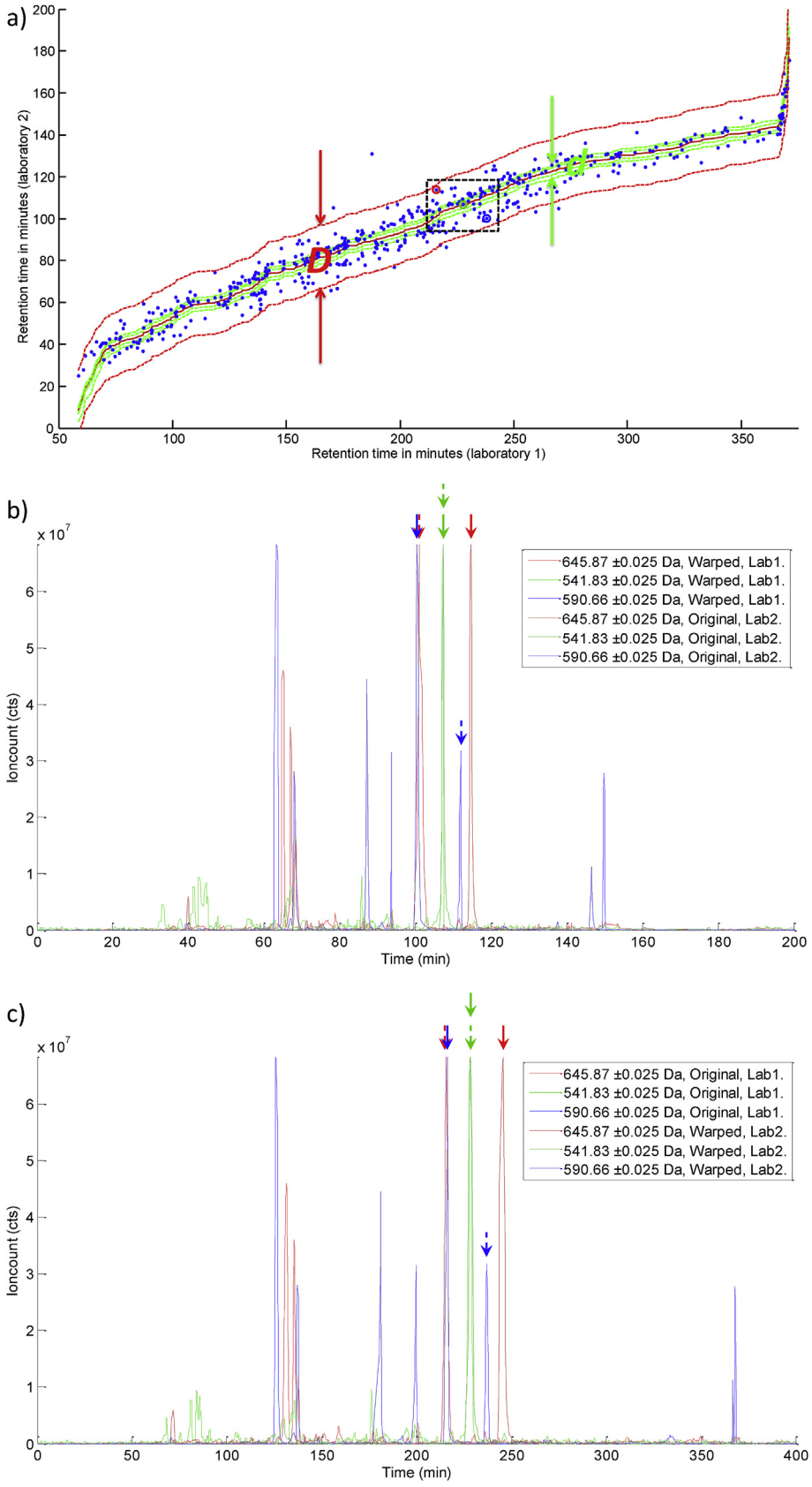


Fig. 4. Orthogonality results in considerable mismatching of LC-MS(/MS) peaks. a) shows a scatter plot of retention times of peptides matched based on agreement of peptide sequence (blue dots) in two chromatograms acquired with two different LC-MS/MS platforms, in the different laboratories under different gradient programs (same data is presented in Fig. 3). The monotonic retention time correction function is shown as a red solid line. The maximal deviation of peptides from the monotonic correction function obtained with robust kernel density approach and between laboratories is shown with red dashed line (red D). Green dashed line and greed "d" label shows the maximal deviation of

is negligible in the *rt* and *m/z* dimensions, i.e. the influence of other compounds on the *rt* and *m/z* of one particular compound in the sample is limited. Using compounds with known but different quantities in the two chromatograms would result in compounds that are in different concentration ranges and their values could be affected by different detector response and/or ion suppression.

When the second condition is not met, common compounds or compounds with the same quantity are present in the two chromatograms, but the correction algorithm is unable to find them in sufficient number, density and accuracy to perform accurate correction. Beside the numbers of common compounds and common compounds with the same quantity, the distribution of them along the full measured range is important as well. If there are domains with no or low number of common compounds or compounds with the same quantity present, then information for monotonic shift correction is lacking at these locations and local misalignment may occur. In highly complex proteomics samples, common compounds and compounds with the same quantity are present in sufficient number and density across the full measured range. This may be challenging however for lower complexity metabolomics data. Typical examples of lack of information is at the beginning or end of the chromatogram where no compounds elute. Other important aspect is the accuracy of the alignment algorithm to select the common compound or the compound present with same quantity. If mismatched compounds or noise is present with large extent, then correction algorithm may be inaccurate. When the third point is not met, the common compounds or compounds with the same quantity are mixed-up and the exact location or quantity of a compound cannot be exactly determined in the other chromatogram by deriving a single monotonic correction function.

2.3. Distinction between monotonic and non-monotonic shifts and orthogonality

Correctable shift should be monotonic since any deviation from monotonicity would lead to a break the one-to-one correspondence of coordinate transformation. Monotonicity of shifts also ensures the mathematical inversion of the shift correcting function, which in fact inverses the role of sample and reference in the aligned chromatographic pairs. Monotonic and non-monotonic shifts have a different physicochemical origins and should be algorithmically treated differently. Monotonic shifts can be corrected, but non-monotonic one not unless the physicochemical process that leads to non-monotonic shift can be fully modelled. It is important to note that monotonic shift should be corrected with single monotonic function generally applied to all compounds in MS1 maps. Correction for non-monotonic shift requires compound specific monotonic correction function obtained from precise modelling of retention mechanisms or intensity changes of compounds. The application of a monotonic function to a group of signals is rare, but one example is provided later when individual monotonic function is applied for each *m/z* channel of MS1 map to correct small fluctuation in ion trap data caused by charge repulsion. Assessment of monotonic and non-monotonic shifts are performed using only compounds that are present in both chromatograms (common compounds) and using common compounds that are present with the same quantity in the two chromatograms in *iin* readout.

Non-monotonic shift may have two components. One component is related to data pre-processing errors such as to determine compound signal location in MS1 map (*m/z* and *rt* dimensions) or compound quantification (*iin* readout). The second is related to elution order inversion of common compounds and therefore can be interpreted as the analytical chemistry definition of orthogonality. The metric to calculate orthogonality should be calculated after correction for monotonic shift and will inevitably contain the data pre-processing error. Comparable MS1 maps without the need for complex modelling of orthogonality can be therefore obtained for MS1 map pairs, which includes only monotonic shift and non-monotonic shifts with data pre-processing error component.

Publications so far discuss separately alignment (correctable monotonic shift) and assessment of orthogonality in LC-MS(/MS) (and GC-MS or CE-MS) data. For example orthogonality is considered absent when it comes to design of retention time alignment algorithm even the existence of elution order i.e. presence of small orthogonality was recognised in multiple articles [28,29]. However, it is obvious that the two phenomena may be present to a different extent in various datasets, and may influence the performance of monotonic shift correction and orthogonality assessment algorithms. Orthogonality in the literature was related solely to the retention time domain and was not mentioned for the *m/z* dimension or in the *iin* readout of MS1 map [22–25]. With correction of single monotonic function, we separate monotonic shift from non-monotonic ones, which may have orthogonality component. Fig. 3 shows a pair of chromatograms of the same complex proteomics sample that shows non-linear monotonic shifts mixed with orthogonality and non-monotonic shift due to data pre-processing error. The figure also shows the monotonic retention time correction function and the non-monotonic shift after correcting for monotonic shift with a single monotonic function applied to all compounds.

Since orthogonality cannot be corrected without accurate modelling and without knowing the identity of the peak in the MS1 data it has as consequence that either *rt* or *m/z* coordinates of a compound cannot be predicted precisely in other LC-MS data, while in the *iin* readout the normalisation will have limited precision. Fig. 4a shows a scatterplot of retention time of identical peptides in two chromatograms that were obtained with analysis of the same sample using two different LC-MS/MS platforms and gradient LC programs. Non-linear monotonic shift and orthogonality is obviously visible on the plot. Alignment of the two chromatograms using monotonic best fitted retention time correction function on the scatterplot using LOWESS regression constrained for monotonicity results in accurate alignment of peaks that are located on the correction function, while peaks far from this function are misaligned (Fig. 4b and c). Orthogonality in this tutorial is assumed to have a symmetric form around a main monotonic trend, which is generally the case when the goal is to align datasets corrected for non-monotonic shift with small orthogonality component (i.e. strong correlation of *rt* of the same compounds in

peptides from the main monotonic retention time correction function in data that was acquired in the same laboratory using the same LC-MS/MS platform and the same eluent program. The difference between red "D" and green "d" is related to the non-monotonic shift of the liquid chromatographic separation and shows the uncertainty to determine corresponding peak locations in two different chromatograms. Peak pairs with red, blue and green circles in the black dashed box area are corresponding to the three peak pairs that are used to illustrate the effect of peak elution order inversion in extracted ion chromatograms (EICs) in plots b and c after aligning one of the chromatograms to the other one. In plot b), the chromatogram of laboratory 1 was aligned to the chromatogram of laboratory 2, while in plot c) the chromatogram of laboratory 2 was aligned to the chromatogram of laboratory 1. Peptide LTLPQLEIR (green arrows) is located on the monotonic retention-time correction function, while the peptides DIAPTLTLYVGK (red arrows) and VHQFFNVGLIQPGSVK (blue arrows) are located far from this function. Retention time alignment using a single monotonic retention time correction function provides well aligned peaks for the first peptide (green traces). The two other peptides (red and blue arrows) suffer from considerable misalignment with retention time error close to the distance D due to considerable orthogonality. The EICs are normalized to the highest peaks and the Y axis represent ion counts relative to the most abundant signal intensity. Figures adapted from Mitra et al. [33]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the two chromatograms). This situation may be different when orthogonality is large e.g. in case of optimisation of peak capacity in multidimensional chromatography [22,30]. Another assumption that we include in the discussion of monotonic and non-monotonic shifts is that these shifts are independent between the two separation dimensions of *m/z*, and *rt*, except for *iin* which lead to the requirement of having the same compounds with the same quantity present in chromatographic pairs. Interaction between *rt* and *m/z* dimensions exist but their effect is generally small [31,32].

Orthogonality can be also considered between the *rt* and *m/z* dimensions, and the *iin* readout of single MS1 map, however this orthogonality is not related to the assessment of comparable MS1 maps and is therefore outside of the scope of this paper.

3. Shifts and orthogonality in single-stage LC-MS data

In this section the physicochemical origins of monotonic and non-monotonic shifts in the *rt*, *m/z* dimensions and in the *iin* readout along with algorithms that are used to correct for monotonic shift or assess the degree of non-monotonic shifts is discussed in detail. One pertinent problem relates to the definition of the term "same compound" in multiple samples. A chemical compound can be modified in different ways ranging from chemical modifications, adduct formation, charge state differences, or can be present at different degrees of dissimilarity when it comes to chemical and 3D structures such as diastereomerisation, *cis/trans* isomerization, structural (constitutional) isomers, chiral isomerisation and conformation changes. Table 1 lists molecular variants and modifications that describe how compounds in the same chemical structure family can be discriminated in the *rt* and *m/z* dimensions and in the *iin* readout of the MS1 map.

3.1. Retention time dimension

Physicochemical background. The dimension most prone for shift and orthogonality is the chromatographic dimension. Multiple factors may influence the elution time of a compound which may result in non-linear retention time shifts between chromatograms, such as slight changes in column/eluent temperature, slight changes in eluent's pH, modification of the stationary phase surface e.g. due to accumulation of the non-eluted components from previously analysed samples, degradation of the surface chemistry or mechanical changes of the stationary phase due to high pressure and slight changes in the solvent delivery and/or mixing system of the liquid chromatography apparatus [9].

Within a quantitative profiling study, orthogonality of separation is a property that is attempted to be minimized since orthogonality lowers the precision to predict the retention time of a compound in different MS1 maps [33]. Orthogonality may have different origins compared to monotonic shifts, such as those listed as cause of non-linear monotonic shifts. For example, simple change of the gradient program leads to slight orthogonality. The reason of this orthogonality has been already described in the linear solvent strength theory introduced by Snyder and his coworkers in the 60's [34] and this effect was considered by other researchers as well [35,36]. As a consequence, chromatograms acquired with different gradient programs will show different degrees of orthogonality, which in turn determines the maximal accuracy that can be achieved by retention time alignment using single non-linear monotonic correcting function. It is therefore important to consider for data generator and data evaluator scientists, that the same LC column the same gradient program and eluent composition should be used to obtain comparable MS1 maps. However these conditions are not sufficient in obtaining comparable MS1 maps, since it does not account of e.g. degradation of the LC column nor in change of gradient delivery systems.

Monotonic shift correction algorithms. In the last two decades multiple retention time correction algorithms were developed as part of label-free LC-MS(/MS) data pre-processing workflows [19,33,37–50]. A comprehensive review by Smith et al. [9] includes discussion of 50 open source retention time alignment algorithms. Although several retention time alignment algorithms exist, the general objective of every time alignment algorithm is to first identify peaks (or signal) of the same compound in two (or more) chromatograms and provide a retention time transformation function, that corrects for monotonic retention time shifts and aligns LC-MS(/MS) datasets. Retention time correction algorithms can be classified in many ways such as: i) type of data and MS1 map dimensions used for the alignment, such as using the complete MS1 map, total ion or base peak chromatograms, peak lists [39]; ii) if alignment is performed pairwise or in one step and iii) type of benefit or objective function used to measure similarity of the chromatographic pair, which is used subsequently to derive retention time correction function (e.g. sum of the squared ion intensity distance of raw data, correlation of raw ion intensity or sum of overlapping peak volume).

One of the most widely used algorithmic approach to derive the correction function is dynamic time warping (DTW) [51] that identifies the optimal retention time correspondence path. This path can be obtained by minimizing the cumulative differences between the LC-MS signal at different sampling points either using peak lists [52], TIC [47] or the regions of MS1 maps [53]. Correlation-Optimized time Warping (COW) [54] performs segment-wise stretching or shrinking of the retention time segments and uses a cumulative benefit function that maximizes segment profile similarity such as correlation [54] or sum of overlapping peak volumes [55]. The combination of segments positions that fit best the reference chromatogram is obtained using dynamic programming. Christin et al. [45] combined COmponent Detection Algorithm (CODA) with COW, which algorithm includes only information from LC-MS mass traces that contain low noise and background and large number of high abundant peaks from the sample and reference chromatograms. CODA implements a moving window, to detect *m/z* traces in different retention time domains with high quality peak content. Another algorithm called parametric and semi-parametric time warping ((s)PTW) uses fitted polynomial as a warping function that minimize the profile abundance differences between LC-MS chromatograms using TIC [56–58] or combined CODA selected mass traces [53]. OpenMS [59] applies an affine transformation to the retention time coordinates of sample feature list using linear regression on features obtained with robust matching (pose clustering) of the *rt* and *m/z* coordinates.

Commonly used time alignment methods either use centroid peak lists or charge-state- and isotope-deconvoluted feature lists. These lists are then used to model a retention time alignment function based on retention time values of correspondences. Correspondences could be defined as matched peak pairs within certain *rt* and *m/z* coordinates or bins or matched landmark isotopic features between datasets. However algorithms such as PEPPeR [60], SuperHirn [18], IDEAL-Q [42] and LCMSWARP [61] use a combination of isotopic feature detection and MS/MS identification to enhance the "Landmark Matching" process prior to retention time alignment. Many time alignment algorithms perform alignment pairwise, which poses the problem of reference selection. Star type of alignment using one reference to which all other

Table 1
Summary of molecular variants, which effects the definition of same compound (molecular entity). The table contains molecular variants at various levels and presents how molecular variants can be distinguished in the *rt* and *m/z* dimensions and the *iin* readout of the MS1 LC-MS(/MS) data.

Type of modification/molecular variant	Retention time (*rt*) dimension	Mass-to-charge ratio (*m/z*) dimension	Ion intensity (*iin*) readout
Chemical modifications (covalent bond changes)	Difference can be expected, which extent is depending form the type and size of the modification	Difference is expected if there is a change in molecular mass of the target compound.	Chemical modification leads to differences in ionisation properties, therefore same ion intensity may express different amount of compounds.
Same chemical but different isotopic constitution	No difference in retention time, only slight difference is expected when deuterium/hydrogen replacement occurs.	Difference should be observed when mass of the intact ion changes.	No difference between members of this type of compounds is to be expected.
Different charge state	Certain eluent composition (e.g. pH) may influence charge of the peak and therefore the retention time. The effect is depending form the time scale of hydrogen exchange and the pH.	In principle the charge states during liquid chromatography influence the charge distribution of the analytes in the MS. The same holds in changing electrospray conditions such as voltage, application of shearing gas (ionspray), different eluent or uses of eluent modifiers etc).	Charge state differences in chromatography or at the MS interface may influence the number of formed ions and may provide different detected response.
Adduct formation (Na^+, K^+, NH_4^+, Mg^{2+}, Ca^{2+} etc.)	May result in distinct peaks in the LC dimension.	Results in distinct peaks if mass of the compound changes.	Adduction formation may influence the competition for charges and this could lead to different detector response.
Diastereomers, *cis/trans* isomers	Physicochemical property changes of the analyte may result in different retention time.	Undistinguishable in this dimensions without fragmentation.	Very small (mass defect) or no difference is to be expected.
Constitutional isomers	May be resolved in chromatographic domain, but retention time are expected to be close, except when 3D structure has major changes.	Undistinguishable in this dimension without fragmentation.	Expected to provide the same response.
Chirality	May be distinguishable in this dimension in special condition e.g. by using chiral counter ions or chiral stationary phases.	Undistinguishable in this dimension without fragmentation.	Expected to provide the same response.
Conformational isomers	May be resolved in chromatographic domain, but retention time are expected to be close, except when 3D structure has major changes.	Undistinguishable in this dimension without fragmentation.	Expected to provide the same response.

chromatograms are aligned is suboptimal in alignment of large dataset containing chromatograms with dissimilar molecular composition. Voss et al. [52] developed the simultaneous multiple alignment of LC-MS peak lists. This algorithm performs the pairwise matching of peak lists following a hierarchical-tree based alignment of subsequent chromatographic pairs using peak list similarity as sequence of alignments. Finally, the algorithm calculates a global retention time correction function using a multidimensional kernel function and uses maximum likelihood estimation to derive the common elution profile. It should be noted that the assumption of the existence of a global retention time profile of MS1 map set could be wrong e.g. in dataset that contains chromatogram obtained with different gradient programs due to orthogonality.

Many papers confuse time alignment with peak or feature matching step and use the word "feature alignment" or "peak alignment" for peak matching. The origin of this confusion may be that retention time shift correction algorithms need information from common compounds and one of the goals of shift correction algorithms is to find them. However, the goal of shift correction algorithms is not necessarily to find all common peaks (or signal of common compounds) between chromatograms, but to find them in a sufficient number, distribution and quality that allows to obtain a single monotonic shift correction function. After correction of shifts, the final peak matching algorithm is used to identify with highest accuracy all corresponding peaks across multiple chromatograms. The monotonicity aspect of shift correction means that the shift correction function cannot change the elution order of the peaks and provides one-to-one correspondences between chromatograms, while peak matching should deal with the remaining non-monotonic shift. The accuracy of the peak matching step will be dependent on how close the algorithm should look for corresponding partners in the two chromatograms, which distance will be smaller in case of data that was successfully corrected for monotonic shift compared to data where considerable monotonic shift is present. Many algorithms combine time alignment and feature matching in one module. PEPPeR, IDEAL-Q, SIMA [52], LWBMatch [62] and algorithm developed by Wandy et al. [63] which include grouping of peaks of related compounds are examples of algorithms which combine time alignment with peak matching within a single module.

Datasets with considerable peak elution order inversion (orthogonality) was aligned by Bloemberg et al. [64] using mass-trace optimized PTW. However, PTW does not change the elution order of the peaks, since it derive monotonic retention time correction function, and cannot deal properly with LC-MS(/MS) pairs with significant elution order inversion. It is also obvious that the retention mechanism of analytes/stationary phase that lead to elution order inversion i.e. orthogonality in two chromatograms does not solely depend on the *m/z* of the compound, but rather on other parameters and from complex retention mechanism of the eluting compounds. This approach providing different retention time correction function for different *m/z* traces does not take into account peak elution order inversion within a mass trace.

Non-monotonic shift assessment algorithms. Metrics to measure the amplitude of orthogonality were solely developed for retention time dimensions and was used to assess the difference and similarity in chromatography systems. This assessment is based on joint peak capacity in two-dimensional liquid (2D-LC) or gas chromatography systems. The goal in 2D-LC is to maximise orthogonality between the first and second separation dimensions and concomitant peak capacity of the chromatographic system, therefore those algorithms deal with large orthogonality. One of the first metrics for orthogonality was introduced by Gilar et al. [23,65]. This metric measures the occupancy of bins of common peaks determined based on identified peptide sequences in the retention space of the two chromatograms. Recently Camenzuli et al. [22] introduced a generic measure of orthogonality that uses spread of peaks along 4 equations enclosing 45° of angle and crossing in the middle of normalized retention time that range between values of 0 and 1. The latter approach is independent on the density distribution of peaks providing an accurate measure of orthogonality. Gilar et al. [24] compared 4 different measures of orthogonality using binning of retention times (correlation coefficients, mutual information, box-counting dimensionality, and surface fractional coverage with different hulls) and concluded that except correlation all orthogonality metrics are related to each other and are suitable to optimise peak capacity in two dimensional chromatography. Schure et al. [25] recently summarized the 20 metrics of orthogonality and assessed their performance using 47 two-dimensional LC chromatograms. This article pointed out that there are many metrics to measure orthogonality. Principal component analysis of the different orthogonality metrics shows that despite the fact that the studied metrics are correlated they do capture different aspects of the data. However so far there is no approach published that assesses orthogonality at the lower end i.e. small orthogonality between chromatographic separations. Developing metrics to measure small orthogonality is important, since orthogonality causes uncertainty to predict where a compound will elute in the other chromatogram and therefore determines the search domain to look for corresponding peaks by the peak matching algorithm using *rt* and *m/z* coordinates. Many peak matching algorithms try to find corresponding peak at all cost by allowing wide range to search for corresponding partners, which implementation may lead to mismatched peaks and subsequent statistical error. For this reason, we have developed an approach that assesses the extent of non-monotonic shift corresponding to the maximal retention time matching domain after alignment with single monotonic function. The algorithm determines the uncertainty region used to identify corresponding peaks in LC-MS(/MS) chromatogram pair of interest and LC-MS(/MS) chromatogram pair acquired subsequently in the same analyzis batch, where no peak elution order occurs and compare these regions on the basis of orthogonal residuals to assess the presence of peak elution order inversion or orthogonality [33].

Orthogonality between chromatograms will also have an effect on the accuracy of retention time normalisation algorithms such as iRT [66,67] or RePLiCal [68], which use the standardized retention time of reference standard set obtained with a standard mixture or spiked QconCAT proteins. In this case orthogonality will decrease the accuracy of normalised retention times or even may lead to completely false results in case of mismatching the reference standard peaks between chromatograms.

3.2. Mass to charge ratio dimension

The shifts in the *m/z* dimensions are mainly monotonic and may be caused e.g. by small change in temperature in the room where the instrument is installed in case of high resolution Orbitrap and time of flight mass analyzers or space-charge effect in case of low resolution three dimensional ion trap mass analyser [39]. Due to well-known physics of ion separation in theory no orthogonality in *m/z* dimension could happen except for a charge state shift of compounds, which may introduce orthogonality because the different compounds depending on their charge affinity have

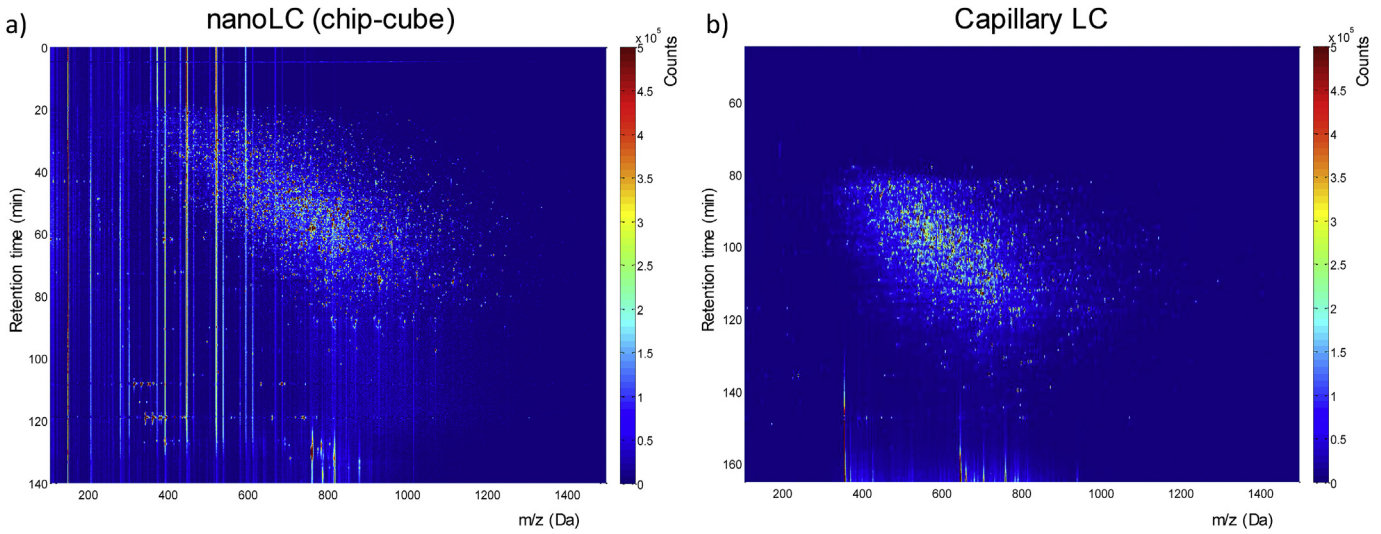


Fig. 5. Effect of charge state distribution in MS1 map. Image of an MS1 map of the same human serum depleted from the 6 most abundant proteins acquired with an Agilent ion trap LC-MS platform using nanoLC integrated in a microfluidic device (image a) and using capillary LC (image b). nanoLC was operated with an eluent flow rate of 300 nl/min, electrospray for peptide ionization and the injected sample amount was 5 pmol, while capillary LC analysis was performed using ionspray (electrospray enhanced with pneumatic nebulisation), 20 μl/min of flow rate and the injected sample amount was 140 pmol. Pneumatic nebulisation in ionspray provides additional charging of peptides resulting in shift of charge state of compounds, which effect can be different for the different peptides resulting in orthogonality in *m/z* dimension. Figure adapted from Horvatovich et al. [69].

different charge state distribution changes. Shifts of charge distribution is unconventional, which happens at discrete *m/z* values, compared to conventional shifts such as retention time shift, which has continuously scale. During electrospray process, ionisation parameters have a large influence on the charge distribution of analytes. For example, ionspray combining electrospray with pneumatic nebulisation used with normal or capillary LC column results in more charges on the same analytes due to triboelectric effect compared to electrospray ionisation regime. The effect of charge is dependent from the chemical composition of analytes, therefore its effect is different for the different analytes resulting in orthogonality. Fig. 5 shows the considerable charge shift in MS1 map obtained with analysis of the same human blood sample depleted from the 6 most abundant proteins on a LC-MS platforms differing in the used LC column diameter, the injected sample amount and electrospray ionisation type (ionspray and electrospray) [69]. No orthogonality measure was so far developed for the *m/z* dimension, but "orthogonality" due to charge state shifts can be corrected in compound lists by calculating the neutral mass of compounds and summing up the intensity of the different charge states. Other aspects of orthogonality may relate to adduct formation of the same analytes. Adduct formation is often taken into account in untargeted label-free metabolomics LC-MS data pre-processing workflows, and correction for them is performed by summing up intensities that belongs to the different adduct forms of the same metabolite. However, the detector response may be dependent from *m/z* range and adducts may alter the ionisation efficiency and therefore the measured signal for a given amount of analytes. These changes in detector signal are generally not taken into account when different types of ion signal are summed up in current data pre-processing pipelines.

Mass recalibration algorithms. Several algorithms were developed to correct for monotonic shift in *m/z*, with the goal to enhance mass accuracy, which becomes essential for modern high resolution mass spectrometers. Space-charge effect in low resolution three dimensional ion trap instruments may cause shift in *m/z* which stays monotonic within a mass spectrum. Space-charge effect are caused by the presence of high abundant compounds close in *m/z* to other ions that results in ion repulsion, which effect may be particularly strong in ions trapped in three dimensional space [70]. To correct for shifts in *m/z* domain, routine calibration of the mass spectrometers based on spiked internals standards [31,39] or ubiquitous background ions and contaminants [71] are performed at regular intervals of time or for each acquired mass spectrum. The most widely used approach to device a single monotonic mass shift correction function is based on regression using polynomial function of 2–5°. Generally one monotonic function is used for all MS spectra of the MS1 map, but it become more common to use MS spectra specific monotonic corrections function especially when calibrants are present in all spectra such as co-infused compounds or background ions. Methods that utilise prior knowledge of the sample being analysed in combination to multidimensional non-parametric regression have shown to decrease standard deviations of *m/z* errors by 1.8–3.7 fold [31]. Mass correction algorithm that takes part of bioinformatics toolbox of Matlab (available from version R2007a) eliminates the monotonic shift in *m/z* trace caused by space-charge effect by using advanced data binning algorithms that synchronize all the spectra in a dataset to a common mass/charge grid [72–74] (Fig. 6a and b). Space charging effect influenced by the eluent and co-eluting compound composition is strong in ion trap data, where the order of peaks stays the same but the monotonic shift can differ between different *m/z* traces. This allows to use different monotonic correction functions for individual *m/z* trace in contrast to *rt* domains where single monotonic correction used for all mass trace and compound is justified. Removal of mass measurement error is not only required for MS1 data processing, but also for correction of precursor mass error in the assignment of peptide identifications. One way to correct monotonic shift in *m/z* dimension is to obtain monotonic correction function for the difference between the measured *m/z* of the precursor ion and the theoretical *m/z* of the identified peptide (Fig. 6c) [75]. Petyuk et al. [31] have corrected mass measurement errors for covariates of *m/z*, such as retention time, ion intensity and other parameters using a multidimensional,

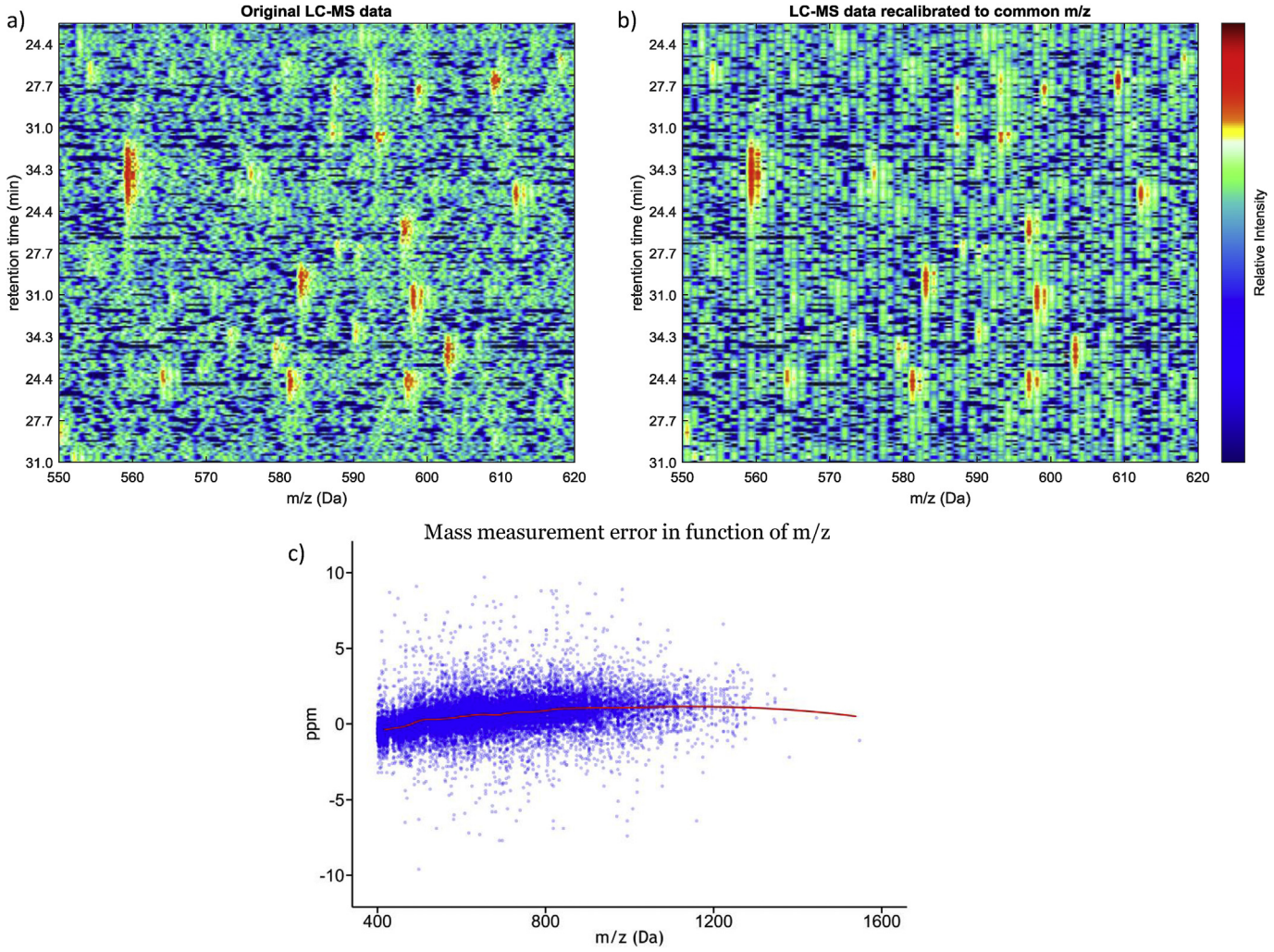


Fig. 6. Correction of monotonic shifts in *m/z* dimension of low resolution ion trap and high resolution Orbitrap LC-MS(/MS) data. Image representation of a raw ion trap MS1 LC-MS map, which shows the fluctuation of *m/z* due to space-charge effect in three-dimensional low resolution ion trap data (image a). This fluctuation results in small monotonic shifts, which does not change the order of peaks in *m/z* dimension and therefore could be corrected with binning algorithms that synchronizes all spectrum in a LC-MS chromatogram to a common *m/z* grid (image b). Scatter plot of mass error (difference of measured precursor *m/z* and theoretical *m/z* calculated from the sequence of identified peptide), showing non-linear monotonic shift and orthogonality in *m/z* dimension of high resolution Orbitrap LC-MS/MS data (plot c). Correction for monotonic shifts enhances the peptide identification rate, which option is implemented in some data pre-processing workflows. Images a and b were obtained with and LCQ ion trap LC-MS platform analysing a mix of 7 proteins obtained from Sashimi data repository (file 7MIX_STD_110802_1 from http://sashimi.sourceforge.net/repository.html). Plot c was obtained from proteomics analysis of HeLa cell using QExecutive Orbitrap LC-MS/MS platform and 1 h of gradient program.

nonparametric regression model. Based on the results from the study, the authors expected to reduce the number of false identifications by 2–4 fold after correcting for mass measurement error [31]. Lommen et al. [32] showed the dependency of mass error in function of retention time and ion intensity and the correction for these shifts allowed to reach sub ppm accuracy for steroid metabolites in UHPLC-Orbitrap platform. These studies show that minor interaction between MS1 dimensions exits and have effect on the accuracy of pre-processed LC-MS(/MS) data.

3.3. Ion intensity readout

Experimental variability such as fluctuation of ionization efficiency in complex samples e.g. due to ion suppression, changing eluent composition, difference in electrospray interface and parameter settings, and differences in sample preparation can influence quantified peptide/protein levels [76]. Ion suppression is a source of orthogonality in LC-MS(/MS) data in *iin* readout, since intensity of compounds may differ based on the composition of co-eluting compounds [77]. Ion suppression is larger in ionspray which combines electrospray with pneumatic nebulisation to ionise compounds at high eluent flow rate. Pneumatic nebulisation provides triboelectric effect which results in additional charging of compounds depending on their charge affinity [69]. However, ion suppression becomes less important at lower flow rate regimes where electrospray only dominates and this effect disappears at very low flow rates of a few nl/min [78]. In *iin* domain, methods used to correct monotonic shifts are known as normalisation and approach to assess orthogonality is unknown. When ion suppression effect is taken into consideration normalisation should be performed using the same set of compounds that have the same quantity in the two samples and have sufficiently even distribution in the full dynamic range of the detector. The best practice is to use an internal standard mixture for normalisation purpose, with

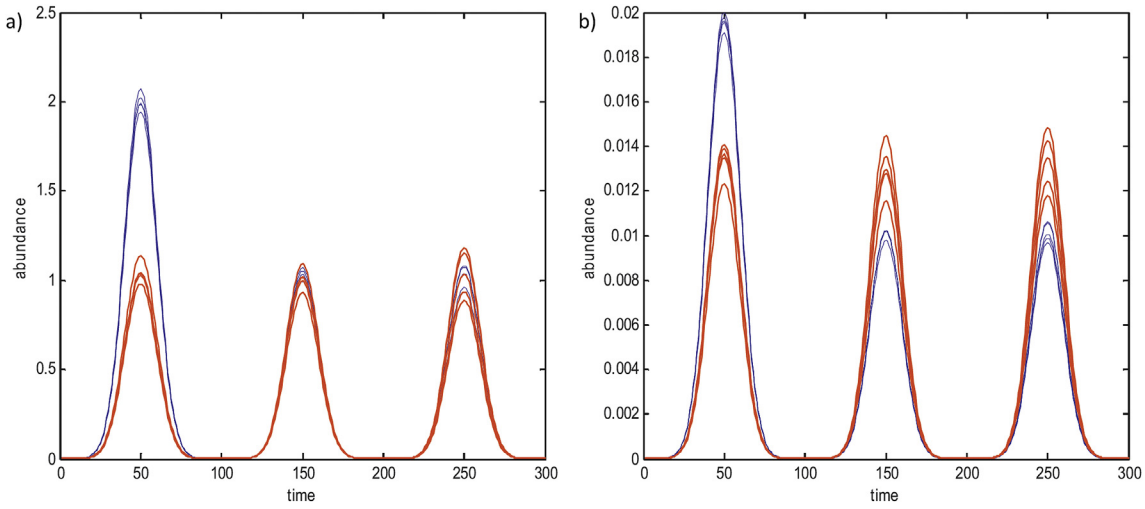


Fig. 7. Principle of "effect size" using simulated data of three peaks and two sample groups (red and blue traces). Effect size occurs when one sample class has large changes of one compound (first peak in blue traces) or part of the compounds only and the other peaks does not change (last two peaks in blue traces) compared to peaks in the other sample group (all peaks in red traces) where the amount of these peaks stays the same. The original situation is shown in plot a), while normalized data using the total sum of peak area (or compound quantity) results in lowering the fold change of the peak that has the major quantity change and introduces smaller fold changes in the two peaks that is present with the same quantity. This type of normalization leads to error in subsequent differential statistical analysis. Figure adopted from Filzmoser et al. [86]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

known absolute concentration of all analytes.

Normalisation approaches. The normalization step has the aim to correct monotonic shifts in *iin* readout. Commonly applied normalisation approaches use mean, median or some global fixed value to correct constant shift in intensity in each sample [79]. Such normalisation methods remove systematic bias across samples and assume that all peptides behave similarly and independently of their abundances across multiple samples. Constant value are often calculated from a set of unique peptides originating from known house-keeping proteins that are supposed to be tightly regulated and to have similar concentration in biological samples [80]. Global adjustment can correct for differences in the amounts of material loaded on the LC-MS(/MS) system for each sample, but cannot capture more complex (e.g., non-linear and intensity-dependent) biases. LOWESS regression approach applied in the ion intensity domain or quantile normalisation that makes distribution of peaks intensity similar across multiple samples [79,81] can correct for such non-linear bias [79], however these approaches assume that the majority of the compounds are the same and have very similar quantity across samples [76]. ANOVA and regression models can effectively remove systematic differences when their sources are known [82]. In order to normalise and model data obtained from varied sample groups, such as disease versus control, a method called normalized spectral index (SIN) was developed. SIN combines three MS abundance features: peptide count, spectral count and fragment-ion (MS/MS) intensity [83]. Most normalization methods used for label-free proteomics data, such as normalisation to various central tendencies (e.g. mean, median), LOWESS regression and quantile normalization, have originated in microarray studies [79,84]. Specific LC-MS(/MS) based data normalisation methods have also been developed which applies probability based model for imputing missing events in order to avoid severe biases due to compounds present below the detection limit in the statistical analysis [85]. All of the above described approaches do not change the *iin* order of peaks originating from the same compounds that have the same quantity in chromatograms i.e. they perform monotonic transformations.

Improper normalisation may introduce bias in the statistical analysis for example when one subclass of compound differs considerably in one sample group while the remaining compounds remains unchanged between samples (so-called non-closed data) and normalisation is performed using a fixed value such as sum of ion intensity, median fold change, sum of peptide-spectrum-matches or injected sample amount (Fig. 7). This effect is called size-effect and ratio based normalisation approach should be used to avoid such error [86]. The application of pairwise normalisation allowed to identify synergistic RAS and CIP2A signalling in HeLa cells before and after phosphopeptide enrichment. In this dataset there is a major shift in phosphopeptide composition before and after phosphopeptide enrichment and before and after stimulation of cells leading to major bias in statistical analysis of the phosphopeptide enriched samples without taking into account the enrichment effect. The enrichment effect was corrected using pairwise normalisation, which calculate a global factor using the median ratio of phosphopeptides that are present in samples both before and after phosphopeptide enrichment steps [87].

3.4. Order of correction for monotonic shift in rt and m/z dimensions and iin readout of MS1 map pairs

Order of correction for monotonic shifts in the *rt*, *m/z* dimensions and in the *iin* readout and the position of these modules in LC-MS(/MS) pre-processing workflows may influence the quality of LC-MS(/MS) pre-processing. In general correction for monotonic shift in *m/z* and *rt* dimensions should be made before peak matching step, since peak matching step require accurate *rt* and *m/z* coordinate of compounds. Normalisation in *iin* readout is generally performed after the peak matching step (Fig. 2). In general orthogonality is rare in *m/z* dimension, therefore it is advantageous to perform first mass recalibration before retention time alignment. Many retention time alignment algorithms uses *m/z* of compounds in peak list or in raw data, therefore this alignment order ensures that more accurate *m/z* values are used to identify common compounds, which drive the time alignment process.

4. Conclusion

Monotonic and non-monotonic shifts were generally considered separately and orthogonality was exclusively considered in retention time dimension. In this tutorial we have demonstrated that these two types of shifts should be considered separately along the *rt* and *m/z* dimensions and the *iin* readout of MS1 part of label-free LC-MS(/MS) data. This has the benefit to assess the quality of MS1 map in the *rt* and *m/z* dimensions and in the *iin* readout with the same mathematical model (i.e. correctable monotonic and non-correctable non monotonic shift). Accurate quantification of multiple MS1 map is possible when monotonic shift and non-monotonic shift due to LC-MS(/MS) pre-processing error are present in an LC-MS(/MS) data set. It should be noted that signals obtained with other separation methods and spectroscopy/spectrometry techniques suffer from similar problems and there are many algorithms that can be adapted to accurately align and pre-process LC-MS(/MS) data. It is obvious that mass spectrometry coupled to other separation techniques such as capillary electrophoresis (CE-MS) and gas chromatography (GC-MS) present similar behaviours of monotonic and non-monotonic shifts and orthogonality to those of LC-MS(/MS) data. For example peak elution order inversion was reported in GC-MS and GC×GC-MS data, which was obtained with different acquisition parameters [88–91]. Signals in two-dimensional gel electrophoresis, NIR or NMR shows joint presence of monotonic and non-monotonic shifts with orthogonality component. One example of algorithm that could be adopted to pre-process LC-MS(/MS) is the generalized fuzzy Hough transform algorithm, which has been used to process NMR spectra acquired in one batch. This algorithm follows NMR signals that change gradually resulting in peak elution order inversion in acquisition-time-sorted NMR spectra [92]. Similar algorithm could be adapted to model gradually changing of orthogonality in retention time in LC-MS(/MS) data, which can be used to determine corresponding peaks in datasets where gradual changes in retention time and elution order occur.

Assessment of small orthogonality in LC-MS(/MS) data is important when peak identity is transferred with accurate mass and time tag approach (AMT). AMT uses solely the *m/z* and *rt* coordinates of peaks and the increase of erroneous identification transfer due to peak elution order inversion was demonstrated by Tarasova et al. [35]. When orthogonality in the *rt* dimension is present, the transfer of peak identity will suffer from uncertainty, and may lead to false positives and negatives peak annotation. Therefore, it is necessary to accurately assess the presence of orthogonality between peptide identification in LC-MS/MS chromatograms. The extent of the orthogonality will determine the accuracy of identification transfer from LC-MS/MS data to LC-MS(/MS) data and will determine the quality of the annotated and quantitative pre-processed MS1 LC-MS(/MS) maps.

In future more effort should be made to develop accurate modelling of orthogonality in the *rt* and *m/z* dimensions and *iin* readout of MS1 maps such as models used to predict accurately retention time of peptides or metabolites. For example linear solvent strength theory in liquid chromatography and three dimensional structure of peptides were successfully used to predict retention time of peptides even when different linear elution programs were used [36,93–95]. However, modelling comes with more experimental effort and cost. For example, retention time prediction of peptides measured with different linear gradient programs and eluent flow rates require to measure peptide standards in different conditions to parametrise properly the retention time prediction model. Similar models should be developed for example to simulate ion suppression process, charge and adduct distribution changes of compounds in ionspray or electrospray regimes. Accurate modelling of orthogonality would reduce the effect of peak-elution order change which determine the uncertainty to match peaks solely using *m/z* and *rt* coordinates and will results in smaller analytical variance in *iin* readout.

In many LC-MS(/MS) profiling studies the data is acquired in one small analysis batch where orthogonality is absent or limited, however orthogonality becomes important when data originate from multiple batches/instruments or when data is acquired in large batches, which will become more and more common in future due to the need for large clinical proteomics and metabolomics studies. We also hope that our tutorial highlight the importance to assess small orthogonality and that data generator and evaluator users known the adverse consequences that orthogonality can have on the outcome of quantitative LC-MS(/MS) profiling studies.

Acknowledgement

We thank the Netherlands Proteomics Center NPC. We thanks the comments and detailed discussion with Frank Suits researcher at IBM Watson Center.

References

[1] S. Nahnsen, C. Bielow, K. Reinert, O. Kohlbacher, Tools for label-free peptide quantification, Mol. Cell. Proteomics 12 (2013) 549–556, https://doi.org/10.1074/mcp.R112.025163.
[2] J.R. Yates, A. Gilchrist, K.E. Howell, J.J.M. Bergeron, Proteomics of organelles and large cellular structures, Nat. Rev. Mol. Cell Biol. 6 (2005) 702–714, https://doi.org/10.1038/nrm1711.
[3] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, B. Kuster, Quantitative mass spectrometry in proteomics: a critical review, Anal. Bioanal. Chem. 389 (2007) 1017–1031, https://doi.org/10.1007/s00216-007-1486-6.
[4] C. a Luber, J. Cox, H. Lauterbach, B. Fancke, M. Selbach, J. Tschopp, S. Akira, M. Wiegand, H. Hochrein, M. O'Keeffe, M. Mann, Quantitative proteomics reveals subset-specific viral recognition in dendritic cells, Immunity 32 (2010) 279–289, https://doi.org/10.1016/j.immuni.2010.01.013.
[5] J. Cox, M. Mann, Quantitative, high-resolution proteomics for data-driven systems biology, Annu. Rev. Biochem. 80 (2011) 273–299, https://doi.org/10.1146/annurev-biochem-061308-093216.
[6] Y. Zen, D. Britton, V. Mitra, A. Brand, S. Jung, C. Loessner, M. Ward, I. Pike, N. Heaton, A. Quaglia, Protein expression profiles of chemo-resistant mixed phenotype liver tumors using laser microdissection and LC–MS/MS proteomics, EuPA Open Proteomics 1 (2013) 38–47, https://doi.org/10.1016/j.euprot.2013.10.001.
[7] K.S. Booksh, B.R. Kowalski, Theory of analytical chemistry, Anal. Chem. 66 (1994) 782A–791A, https://doi.org/10.1021/ac00087a718.
[8] Z. Li, R.M. Adams, K. Chourey, G.B. Hurst, R.L. Hettich, C. Pan, Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos, J. Proteome Res. 11 (2012) 1582–1590, https://doi.org/10.1021/pr200748h.
[9] R. Smith, D. Ventura, J.T. Prince, LC-MS alignment in theory and practice: a comprehensive algorithmic review, Brief. Bioinform (2013), https://doi.org/10.1093/bib/bbt080.
[10] D. a Megger, T. Bracht, H.E. Meyer, B. Sitek, Label-free quantification in clinical proteomics, Biochim. Biophys. Acta 1834 (2013) 1581–1590, https://doi.org/10.1016/j.bbapap.2013.04.001.
[11] X. Lai, L. Wang, F. a Witzmann, Issues and applications in label-free quantitative mass spectrometry, Int. J. Proteomics 2013 (2013) 756039, https://doi.org/10.1155/2013/756039.
[12] S.-E. Ong, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, Mol. Cell. Proteomics 1 (2002) 376–386, https://doi.org/10.1074/mcp.M200025-MCP200.
[13] S.P. Gygi, B. Rist, S. a Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, Nat. Biotechnol. 17 (1999) 994–999, https://doi.org/10.1038/13690.
[14] J. Kellermann, ICPL–isotope-coded protein label, Methods Mol. Biol. 424 (2008) 113–123, https://doi.org/10.1007/978-1-60327-064-9_10.
[15] P. Mortensen, J.W. Gouw, J. V Olsen, S. Ong, K.T.G. Rigbolt, J. Bunkenborg, L.J. Foster, A.J.R. Heck, B. Blagoev, J.S. Andersen, M. Mann, MSQuant, an open source platform for mass spectrometry-based quantitative proteomics

research articles, 2010, pp. 393–403.
[16] O. Kohlbacher, K. Reinert, C. Gröpl, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, M. Sturm, TOPP—the OpenMS proteomics pipeline, Bioinformatics 23 (2007) e191–e197, https://doi.org/10.1093/bioinformatics/btl299.
[17] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinforma. 11 (2010) 395, https://doi.org/10.1186/1471-2105-11-395.
[18] L.N. Mueller, O. Rinner, A. Schmidt, S. Letarte, B. Bodenmiller, M.-Y. Brusniak, O. Vitek, R. Aebersold, M. Müller, SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling, Proteomics 7 (2007) 3470–3480, https://doi.org/10.1002/pmic.200700057.
[19] C. a Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (2006) 779–787, https://doi.org/10.1021/ac051437y.
[20] M.-Y. Brusniak, B. Bodenmiller, D. Campbell, K. Cooke, J. Eddes, A. Garbutt, H. Lau, S. Letarte, L.N. Mueller, V. Sharma, O. Vitek, N. Zhang, R. Aebersold, J.D. Watts, Corra: computational framework and tools for LC-MS discovery and targeted mass spectrometry-based proteomics, BMC Bioinforma. 9 (2008) 542, https://doi.org/10.1186/1471-2105-9-542.
[21] L.P. Freedman, I.M. Cockburn, T.S. Simcoe, F. Collins, L. Tabak, C. Begley, J. Ioannidis, M. Macleod, S. Michie, I. Roberts, U. Dirnagl, I. Chalmers, L. Freedman, M. Gibson, J. Hartshorne, A. Schachner, C. Begley, L. Ellis, F. Prinz, T. Schlange, K. Asadullah, N. Vasilevsky, M. Brush, H. Paddock, L. Ponting, S. Tripathy, P. Glasziou, E. Meats, C. Heneghan, S. Shepperd, D. Christakis, F. Zimmerman, O. Steward, P. Popovich, W. Dietrich, N. Kleitman, J. Kimmelman, J. Mogil, U. Dirnagl, J. Ioannidis, S. Greenland, M. Hlatky, M. Khoury, M. Macleod, E. Sena, H. van der Worp, P. Bath, D. Howells, M. Macleod, L. Freedman, J. Inglese, C. Manski, J. Chakma, G. Sun, J. Steinberg, S. Sammut, R. Jagsi, A. Stern, A. Casadevall, R. Steen, F. Fang, W. Gunn, K. Roth, A. Cox, S. Landis, S. Amara, K. Asadullah, C. Austin, R. Blumenstein, D. Baker, K. Lidster, A. Sottomayor, S. Amor, S. Manolagas, H. Kronenberg, J. Furman, S. Stern, J. Lorsch, F. Collins, J. Lippincott-Schwartz, P. Hughes, D. Marshall, Y. Reid, H. Parkes, C. Gelber, G. Buehring, E. Eby, M. Eby, J. Farrell, T. Simcoe, E. Ostrom, L. Berte, A. Daley, The economics of reproducibility in preclinical research, PLoS Biol. 13 (2015), e1002165, https://doi.org/10.1371/journal.pbio.1002165.
[22] M. Camenzuli, P.J. Schoenmakers, A new measure of orthogonality for multi-dimensional chromatography, Anal. Chim. Acta 838 (2014) 93–101, https://doi.org/10.1016/j.aca.2014.05.048.
[23] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, Orthogonality of separation in two-dimensional liquid chromatography, Anal. Chem. 77 (2005) 6426–6434, https://doi.org/10.1021/ac050923i.
[24] M. Gilar, J. Fridrich, M.R. Schure, A. Jaworski, Comparison of orthogonality estimation methods for the two-dimensional separations of peptides, Anal. Chem. 84 (2012) 8722–8732, https://doi.org/10.1021/ac3020214.
[25] M.R. Schure, J.M. Davis, Orthogonal separations: comparison of orthogonality metrics by statistical analysis, J. Chromatogr. A 1414 (2015) 60–76, https://doi.org/10.1016/j.chroma.2015.08.029.
[26] J. Listgarten, R.M. Neal, S.T. Roweis, P. Wong, A. Emili, Difference detection in LC-MS data for protein biomarker discovery, Bioinformatics 23 (2007) e198–204, https://doi.org/10.1093/bioinformatics/btl326.
[27] J. Listgarten, A. Emili, Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry, Mol. Cell. Proteomics 4 (2005) 419–434, https://doi.org/10.1074/mcp.R500005-MCP200.
[28] M. Vandenbogaert, S. Li-Thiao-Té, H.-M. Kaltenbach, R. Zhang, T. Aittokallio, B. Schwikowski, Alignment of LC-MS images, with applications to biomarker discovery and protein identification, Proteomics 8 (2008) 650–672, https://doi.org/10.1002/pmic.200700791.
[29] K.M. Åberg, E. Alm, R.J.O. Torgrip, The correspondence problem for metabonomics datasets, Anal. Bioanal. Chem. 394 (2009) 151–162, https://doi.org/10.1007/s00216-009-2628-9.
[30] J.M. Davis, D.R. Stoll, P.W. Carr, Dependence of effective peak capacity in comprehensive two-dimensional separations on the distribution of peak capacity between the two dimensions, Anal. Chem. 80 (2008) 8122–8134, https://doi.org/10.1021/ac800933z.
[31] V.A. Petyuk, N. Jaitly, R.J. Moore, J. Ding, T.O. Metz, K. Tang, M.E. Monroe, A. V Tolmachev, J.N. Adkins, M.E. Belov, A.R. Dabney, W. Qian, D.G. Camp, R.D. Smith, Elimination of systematic mass measurement errors in liquid chromatography - mass spectrometry based proteomics using regression models and a priori partial knowledge of the sample content 80 (2010) 693–706.
[32] A. Lommen, A. Gerssen, J.E. Oosterink, H.J. Kools, A. Ruiz-Aracama, R.J.B. Peters, H.G.J. Mol, Ultra-fast searching assists in evaluating sub-ppm mass accuracy enhancement in U-HPLC/Orbitrap MS data, Metabolomics 7 (2011) 15–24, https://doi.org/10.1007/s11306-010-0230-y.
[33] V. Mitra, A. Smilde, H. Hoefsloot, F. Suits, R. Bischoff, P. Horvatovich, Inversion of peak elution order prevents uniform time alignment of complex liquid-chromatography coupled to mass spectrometry datasets, J. Chromatogr. A 1373 (2014) 61–72, https://doi.org/10.1016/j.chroma.2014.10.101.
[34] L.R. Snyder, J.W. Dolan, High-performance Gradient Elution, John Wiley & Sons, Inc, Hoboken, NJ, USA, 2006, https://doi.org/10.1002/0470055529.
[35] I.A. Tarasova, T.Y. Perlova, M.L. Pridatchenko, A.A. Goloborod'ko, L.I. Levitsky, V.V. Evreinov, V. Guryca, C.D. Masselon, A.V. Gorshkov, M.V. Gorshkov, Inversion of chromatographic elution orders of peptides and its importance for proteomics, J. Anal. Chem. 67 (2012) 1014–1025, https://doi.org/10.1134/S1061934812130102.
[36] V. Spicer, M. Grigoryan, A. Gotfrid, K.G. Standing, O.V. Krokhin, Predicting Retention Time Shifts Associated with Variation of the Gradient Slope in Peptide RP-HPLC, 2010.
[37] E. Lange, R. Tautenhahn, S. Neumann, C. Gröpl, Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements, BMC Bioinforma. 9 (2008) 375, https://doi.org/10.1186/1471-2105-9-375.
[38] T.-H. Tsai, M.G. Tadesse, C. Di Poto, L.K. Pannell, Y. Mechref, Y. Wang, H.W. Ressom, Multi-profile Bayesian alignment model for LC-MS data analysis with integration of internal standards, Bioinformatics 29 (2013) 2774–2780, https://doi.org/10.1093/bioinformatics/btt461.
[39] C. Christin, R. Bischoff, P. Horvatovich, Data processing pipelines for comprehensive profiling of proteomics samples by label-free LC-MS for biomarker discovery, Talanta 83 (2011) 1209–1224, https://doi.org/10.1016/j.talanta.2010.10.029.
[40] Z. Tang, L. Zhang, A.K. Cheema, H.W. Ressom, A new method for alignment of LC-MALDI-TOF data, Proteome Sci. 9 (Suppl 1) (2011) S10, https://doi.org/10.1186/1477-5956-9-S1-S10.
[41] N. Hoffmann, M. Keck, H. Neuweger, M. Wilhelm, P. Högy, K. Niehaus, J. Stoye, Combining peak- and chromatogram-based retention time alignment algorithms for multiple chromatography-mass spectrometry datasets, BMC Bioinforma. 13 (2012) 214, https://doi.org/10.1186/1471-2105-13-214.
[42] C.-C. Tsou, C.-F. Tsai, Y.-H. Tsui, P.-R. Sudhir, Y.-T. Wang, Y.-J. Chen, J.-Y. Chen, T.-Y. Sung, W.-L. Hsu, IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation, Mol. Cell. Proteomics 9 (2010) 131–144, https://doi.org/10.1074/mcp.M900177-MCP200.
[43] K. Podwojski, A. Fritsch, D.C. Chamrad, W. Paul, B. Sitek, K. Stühler, P. Mutzel, C. Stephan, H.E. Meyer, W. Urfer, K. Ickstadt, J. Rahnenführer, Retention time alignment algorithms for LC/MS data must consider non-linear shifts, Bioinformatics 25 (2009) 758–764, https://doi.org/10.1093/bioinformatics/btp052.
[44] N. Etxebarria, O. Zuloaga, M. Olivares, L.J. Bartolomé, P. Navarro, Retention-time locked methods in gas chromatography, J. Chromatogr. A 1216 (2009) 1624–1629, https://doi.org/10.1016/j.chroma.2008.12.038.
[45] C. Christin, A.K. Smilde, H.C.J. Hoefsloot, F. Suits, R. Bischoff, P.L. Horvatovich, Optimized time alignment algorithm for LC-MS data: correlation optimized warping using component detection algorithm-selected mass chromatograms, Anal. Chem. 80 (2008) 7012–7021, https://doi.org/10.1021/ac800920h.
[46] X. Lai, L. Wang, H. Tang, F. a. Witzmann, A novel alignment method and multiple filters for exclusion of unqualified peptides to enhance label-free quantification using peptide intensity in LC-MS/MS, J. Proteome Res. 10 (2011) 4799–4812, https://doi.org/10.1021/pr2005633.
[47] R.G. Sadygov, F.M. Maroto, A.F.R. Hühmer, ChromAlign: a two-step algorithmic procedure for time alignment of three-dimensional LC-MS chromatographic surfaces, Anal. Chem. 78 (2006) 8207–8217, https://doi.org/10.1021/ac060923y.
[48] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Peak alignment using wavelet pattern matching and differential evolution, Talanta 83 (2011) 1108–1117, https://doi.org/10.1016/j.talanta.2010.08.008.
[49] M. Ghanat Bari, X. Ma, J. Zhang, PeakLink: a new peptide peak linking method in LC-MS/MS using wavelet and SVM, Bioinformatics 30 (2014) 2464–2470, https://doi.org/10.1093/bioinformatics/btu299.
[50] X. Lai, L. Wang, H. Tang, F.A. Witzmann, A novel alignment method and multiple filters for exclusion of unqualified peptides to enhance label-free quantification using peptide intensity in LC-MS/MS, J. Proteome Res. 10 (2011) 4799–4812, https://doi.org/10.1021/pr2005633.
[51] A. Kassidas, J.F. MacGregor, P.A. Taylor, Synchronization of batch trajectories using dynamic time warping, AIChE J. 44 (1998), https://doi.org/10.1002/aic.690440412.
[52] B. Voss, M. Hanselmann, B.Y. Renard, M.S. Lindner, U. Köthe, M. Kirchner, F.A. Hamprecht, SIMA: simultaneous multiple alignment of LC/MS peak lists, Bioinformatics 27 (2011) 987–993, https://doi.org/10.1093/bioinformatics/btr051.
[53] C. Christin, H.C.J. Hoefsloot, A.K. Smilde, F. Suits, R. Bischoff, P.L. Horvatovich, Time alignment algorithms based on selected mass traces for complex LC-MS data, J. Proteome Res. 9 (2010) 1483–1495, https://doi.org/10.1021/pr9010124.
[54] N.P. V Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, J. Chromatogr. A 805 (1998) 17–35, https://doi.org/10.1016/S0021-9673(98)00021-1.
[55] F. Suits, J. Lepre, P. Du, R. Bischoff, P. Horvatovich, Two-dimensional method for time aligning liquid chromatography-mass spectrometry data, Anal. Chem. 80 (2008) 3095–3104, https://doi.org/10.1021/ac702267h.
[56] P.H.C. Eilers, Parametric time warping, Anal. Chem. 76 (2004) 404–411,

https://doi.org/10.1021/ac034800e.
[57] A.M. van Nederkassel, M. Daszykowski, P.H.C. Eilers, Y. Vander Heyden, A comparison of three algorithms for chromatograms alignment, J. Chromatogr. A 1118 (2006) 199–210, https://doi.org/10.1016/j.chroma.2006.03.114.
[58] A.M. van Nederkassel, C.J. Xu, P. Lancelin, M. Sarraf, D.A. Mackenzie, N.J. Walton, F. Bensaid, M. Lees, G.J. Martin, J.R. Desmurs, D.L. Massart, J. Smeyers-Verbeke, Y. Vander Heyden, Chemometric treatment of vanillin fingerprint chromatograms. Effect of different signal alignments on principal component analysis plots, J. Chromatogr. A 1120 (2006) 291–298, https://doi.org/10.1016/j.chroma.2005.11.134.
[59] E. Lange, E. Lange, C. Gröpl, C. Gröpl, O. Schulz-Trieglaff, O. Schulz-Trieglaff, A. Leinenbach, A. Leinenbach, C. Huber, C. Huber, K. Reinert, K. Reinert, A geometric approach for the alignment of liquid chromatography-mass spectrometry data, Bioinformatics 23 (2007) i273–i281, https://doi.org/10.1093/bioinformatics/btm209.
[60] J.D. Jaffe, D.R. Mani, K.C. Leptos, G.M. Church, M.A. Gillette, S.A. Carr, PEPPeR, a platform for experimental proteomic pattern recognition, Mol. Cell. Proteomics 5 (2006) 1927–1941, https://doi.org/10.1074/mcp.M600222-MCP200.
[61] B.L. LaMarche, K.L. Crowell, N. Jaitly, V. a Petyuk, A.R. Shah, A.D. Polpitiya, J.D. Sandoval, G.R. Kiebel, M.E. Monroe, S.J. Callister, T.O. Metz, G. a Anderson, R.D. Smith, MultiAlign: a multiple LC-MS analysis tool for targeted omics analysis, BMC Bioinforma. 14 (2013) 49, https://doi.org/10.1186/1471-2105-14-49.
[62] J. Wang, H. Lam, Graph-based peak alignment algorithms for multiple liquid chromatography-mass spectrometry datasets, Bioinformatics 29 (2013) 2469–2476, https://doi.org/10.1093/bioinformatics/btt435.
[63] J. Wandy, R. Daly, R. Breitling, S. Rogers, Incorporating peak grouping information for alignment of multiple liquid chromatography-mass spectrometry datasets, Bioinformatics (2015), https://doi.org/10.1093/bioinformatics/btv072 btv072-.
[64] T.G. Bloemberg, J. Gerretzen, H.J.P. Wouters, J. Gloerich, M. van Dael, H.J.C.T. Wessels, L.P. van den Heuvel, P.H.C. Eilers, L.M.C. Buydens, R. Wehrens, Improved parametric time warping for proteomics, Chemom. Intell. Lab. Syst. 104 (2010) 65–74, https://doi.org/10.1016/j.chemolab.2010.04.008.
[65] P. Horvatovich, B. Hoekman, N. Govorukhina, R. Bischoff, Multidimensional chromatography coupled to mass spectrometry in analysing complex proteomics samples, J. Sep. Sci. 33 (2010) 1421–1437, https://doi.org/10.1002/jssc.201000050.
[66] C. Escher, L. Reiter, B. MacLean, R. Ossola, F. Herzog, J. Chilton, M.J. MacCoss, O. Rinner, Using iRT, a normalized retention time for more targeted measurement of peptides, Proteomics 12 (2012) 1111–1121, https://doi.org/10.1002/pmic.201100463.
[67] R. Bruderer, O.M. Bernhardt, T. Gandhi, L. Reiter, High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation, Proteomics 16 (2016) 2246–2256, https://doi.org/10.1002/pmic.201500488.
[68] S.W. Holman, L. McLean, C.E. Eyers, RePLiCal: a QconCAT protein for retention time standardization in proteomics studies, J. Proteome Res. 15 (2016) 1090–1102, https://doi.org/10.1021/acs.jproteome.5b00988.
[69] P. Horvatovich, N.I. Govorukhina, T.H. Reijmers, A.G.J. van der Zee, F. Suits, R. Bischoff, Chip-LC-MS for label-free profiling of human serum, Electrophoresis 28 (2007) 4493–4505, https://doi.org/10.1002/elps.200600719.
[70] D. Guo, Y. Wang, X. Xiong, H. Zhang, X. Zhang, T. Yuan, X. Fang, W. Xu, Space charge induced nonlinear effects in quadrupole ion traps, J. Am. Soc. Mass Spectrom. 25 (2014) 498–508, https://doi.org/10.1007/s13361-013-0784-9.
[71] R.A. Scheltema, A. Kamleh, D. Wildridge, C. Ebikeme, D.G. Watson, M.P. Barrett, R.C. Jansen, R. Breitling, Increasing the mass accuracy of high-resolution LC-MS data using background ions - a case study on the LTQ-Orbitrap, Proteomics 8 (2008) 4647–4656, https://doi.org/10.1002/pmic.200800314.
[72] S. Purvine, N. Kolker, E. Kolker, Spectral quality assessment for high-throughput tandem mass spectrometry proteomics, OMICS 8 (2004) 255–265, https://doi.org/10.1089/omi.2004.8.255.
[73] W. Yu, X. Li, J. Liu, B. Wu, K.R. Williams, H. Zhao, Multiple peak alignment in sequential data analysis: a scale-space-based approach, IEEE/ACM Trans. Comput. Biol. Bioinforma. 3 (2006) 208–219, https://doi.org/10.1109/TCBB.2006.41.
[74] N. Jeffries, Algorithms for alignment of mass spectrometry proteomic data, Bioinformatics 21 (2005) 3066–3073, https://doi.org/10.1093/bioinformatics/bti482.
[75] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372, https://doi.org/10.1038/nbt.1511.
[76] Y. V Karpievitch, A.R. Dabney, R.D. Smith, Normalization and missing value imputation for label-free LC-MS analysis, BMC Bioinforma. 13 (Suppl 1) (2012) S5, https://doi.org/10.1186/1471-2105-13-S16-S5.
[77] A. Furey, M. Moriarty, V. Bane, B. Kinsella, M. Lehane, Ion suppression; A critical review on causes, evaluation, prevention and applications, Talanta 115 (2013) 104–122, https://doi.org/10.1016/j.talanta.2013.03.048.
[78] E.T. Gangl, M. Annan, N. Spooner, P. Vouros, Reduction of signal suppression effects in ESI-MS using a nanosplitting device, Anal. Chem. 73 (2001) 5635–5644, https://doi.org/10.1021/ac010501i.
[79] S.J. Callister, R.C. Barry, J.N. Adkins, E.T. Johnson, W.J. Qian, B.J.M. Webb-Robertson, R.D. Smith, M.S. Lipton, Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics, J. Proteome Res. 5 (2006) 277–286, https://doi.org/10.1021/pr050300l.
[80] C. Colantuoni, G. Henry, S. Zeger, J. Pevsner, Local mean normalization of microarray element signal intensities across an array surface: quality control and correction of spatially systematic artifacts, Biotechniques 32 (2002) 1316–1320.
[81] L. Käll, O. Vitek, Computational mass spectrometry–based proteomics, PLoS Comput. Biol. 7 (2011), e1002277, https://doi.org/10.1371/journal.pcbi.1002277.
[82] E.G. Hill, J.H. Schwacke, S. Comte-Walters, E.H. Slate, A.L. Oberg, J.E. Eckel-Passow, T.M. Therneau, K.L. Schey, A statistical model for iTRAQ data analysis, J. Proteome Res. 7 (2008) 3091–3101, https://doi.org/10.1021/pr070520u.
[83] N.M. Griffin, J. Yu, F. Long, P. Oh, S. Shore, Y. Li, J. a Koziol, J.E. Schnitzer, Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis, Nat. Biotechnol. 28 (2010) 83–89, https://doi.org/10.1038/nbt.1592.
[84] Y.V. Karpievitch, T. Taverner, J.N. Adkins, S.J. Callister, G.A. Anderson, R.D. Smith, A.R. Dabney, Normalization of peak intensities in bottom-up MS-based proteomics using singular value decomposition, Bioinformatics 25 (2009) 2573–2580, https://doi.org/10.1093/bioinformatics/btp426.
[85] P. Wang, H. Tang, H. Zhang, J. Whiteaker, A.G. Paulovich, M. Mcintosh, Normalization regarding non-random missing values in high-throughput mass spectrometry data, Pac. Symp. Biocomput (2006) 315–326, https://doi.org/10.1142/9789812701626_0029.
[86] P. Filzmoser, B. Walczak, What can go wrong at the data normalization step for identification of biomarkers? J. Chromatogr. A 1362 (2014) 194–205, https://doi.org/10.1016/j.chroma.2014.08.050.
[87] O. Kauko, T.D. Laajala, M. Jumppanen, P. Hintsanen, V. Suni, P. Haapaniemi, G. Corthals, T. Aittokallio, J. Westermarck, S.Y. Imanishi, J. Brognard, T. Hunter, J. Zhang, P.L. Yang, N.S. Gray, K. Rajalingam, R. Schreck, U.R. Rapp, S. Albert, I.A. Prior, P.D. Lewis, C. Mattos, F.G. Haluska, T. Hunter, A.A. Sablina, W.C. Hahn, P.J. Eichhorn, M.P. Creyghton, R. Bernards, J. Westermarck, W.C. Hahn, T.I. Zack, J. Chen, B.L. Martin, D.L. Brautigan, M.R. Junttila, A. Laine, W.C. Hahn, A. Rangarajan, S.J. Hong, A. Gifford, R.A. Weinberg, A.A. Sablina, M. Hector, N. Colpaert, W.C. Hahn, J.J. Zhao, N. Naetar, D.P. Mathiasen, J.V. Olsen, O.N. Jensen, M.R. Larsen, T.E. Thingholm, O.N. Jensen, P. Roepstorff, T.J. Jorgensen, Y. Zhang, B.R. Fonslow, B. Shan, M.C. Baek, J.R. Yates, K. Engholm-Keller, M.R. Larsen, K. Sharma, S.Y. Imanishi, T. Ohman, S.E. Ong, P.L. Ross, Y.T. Wang, E.J. Soderblom, M. Philipp, J.W. Thompson, M.G. Caron, M.A. Moseley, F. Yang, A. Montoya, L. Beltran, P. Casado, J.C. Rodriguez-Prados, P.R. Cutillas, E.L. de Graaf, P. Giansanti, A.F. Altelaar, A.J. Heck, N.P. Manes, A. Lundby, P. Casado, F. Gnad, N. Dephoure, P.R. Sudhir, M. Swingle, L. Ni, R.E. Honkanen, C. Bialojan, A. Takai, M. Mumby, J. Omerovic, D.E. Hammond, M.J. Clague, I.A. Prior, T. Taus, V. Suni, S.Y. Imanishi, A. Maiolica, R. Aebersold, G.L. Corthals, E. Sontag, S. Andrabi, O.V. Gjoerup, J.A. Kean, T.M. Roberts, B. Schaffhausen, K.F. Chen, X.W. Zhou, B. Winblad, Z. Guan, J.J. Pei, R. Sears, R.T. Peterson, B.N. Desai, J.S. Hardwick, S.L. Schreiber, P.P. Roux, J. Downward, H. Horn, Y. Xue, P.V. Hornbeck, D.M. Gougopoulou, E. Birkeland, C. Letourneux, G. Rocher, F. Porteu, S.B. Quintaje, C.M. Lucas, C. Bockelman, W. Li, M. Niemela, D. Gaglio, S. Kim, Y.Z. Lee, Y.S. Kim, Y.Y. Bahk, T. Young, B.F. Jin, H.A. Blomster, C. Guzman, M. Bagga, A. Kaur, J. Westermarck, D. Abankwa, M.S. Cline, J. Zhu, J.A. Vizcaino, Label-free quantitative phosphoproteomics with novel pairwise abundance normalization reveals synergistic RAS and CIP2A signaling, Sci. Rep. 5 (2015) 13099, https://doi.org/10.1038/srep13099.
[88] R.J. Pell, H.L. Gearhart, Elution order inversions observed on using different carrier gas velocities in temperature programmed gas chromatography, J. High. Resolut. Chromatogr. 10 (1987) 388–391, https://doi.org/10.1002/jhrc.1240100704.
[89] M. Mehran, W.J. Cooper, N. Golkar, M.G. Nickelsen, E.R. Mittlefehldt, E. Guthrie, W. Jennings, Elution order in gas chromatography, J. High. Resolut. Chromatogr. 14 (1991) 745–750, https://doi.org/10.1002/jhrc.1240141109.
[90] L.M. Blumberg, Temperature-programmed Gas Chromatography, John Wiley & Sons, 2010.
[91] 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.
[92] L. Csenki, E. Alm, R.J.O. Torgrip, K.M. Åberg, L.I. Nord, I. Schuppe-Koistinen, J. Lindberg, Proof of principle of a generalized fuzzy Hough transform approach to peak alignment of one-dimensional 1H NMR data, Anal. Bioanal. Chem. 389 (2007) 875–885, https://doi.org/10.1007/s00216-007-1475-9.
[93] H. Vu, V. Spicer, A. Gotfrid, O.V. Krokhin, A model for predicting slopes S in the basic equation for the linear-solvent-strength theory of peptide separation by reversed-phase high-performance liquid chromatography, J. Chromatogr. A 1217 (2010) 489–497, https://doi.org/10.1016/j.chroma.2009.11.065.
[94] D. Abate-Pella, D.M. Freund, Y. Ma, Y. Simón-Manso, J. Hollender, C.D. Broeckling, D.V. Huhman, O.V. Krokhin, D.R. Stoll, A.D. Hegeman, T. Kind, O. Fiehn, E.L. Schymanski, J.E. Prenni, L.W. Sumner, P.G. Boswell, Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods, J. Chromatogr. A 1412 (2015) 43–51, https://doi.org/10.1016/j.chroma.2015.07.108.
[95] J. Reimer, V. Spicer, O.V. Krokhin, Application of modern reversed-phase peptide retention prediction algorithms to the Houghten and DeGraw dataset: peptide helicity and its effect on prediction accuracy, J. Chromatogr. A 1256 (2012) 160–168, https://doi.org/10.1016/j.chroma.2012.07.092.

Vikram Mitra obtained an engineering degree (B.Eng) from Visvesvaraya Technological University, India in 2007 and then obtained a MSc. from University of Exeter (UK) in 2009. He then started a PhD at Rijksuniversiteit Groningen, the Netherlands. His PhD work involved development of data processing workflows for label-free LC-MS data. Currently, he is employed as a senior scientist (bioinformatics) at Proteome science plc. His research interests involve development of methods for data normalisation, quality control and statistical analysis of proteomics datasets for biomarker discovery. He is also working on development of functional enrichment routines to identify key molecular mechanisms in disease states.

Age K. Smilde is full professor of Biosystems Data Analysis at the Swammerdam Institute for Life Sciences at the University of Amsterdam and as of June 1, 2013 he holds a part-time position as professor at the Department of Food Science at the University of Copenhagen. He has published more than 230 peer-reviewed papers and has been the Editor-Europe of the Journal of Chemometrics during the period 1994–2002. He is a co-founder of the Netherlands Metabolomics Centre; a large Public/Private Consortium devoted to all aspects of metabolomics His research interest is data fusion and multiset methods. For more information: see www.bdagroup.nl.

Rainer Bischoff received his PhD in Chemistry from the University of Göttingen (Germany). After two postdoctoral positions at the Max-Planck-Institute for Experimental Medicine in Göttingen and the Department of Biochemistry at Purdue University (West-Lafayette, IN), he joint industry (Transgene (Strasbourg, France) and AstraZeneca (Lund, Sweden)) where he commenced protein-related research. He joined the University of Groningen (The Netherlands) as professor of Analytical Biochemistry in 2001. His research interests focus on biomarker discovery and validation, bioinformatics, biopharmaceutical proteins and the development of novel instrumental analytical techniques. He has authored over 200 peer-reviewed publications, book chapters and is inventor on 14 patents.

Péter Horvatovich received his PhD from University of Strasbourg (France) in food analytical chemistry in 2001. After 2 years spend in pharmaceutical industry at Sanofi-Synthelabo (Budapest, Hungary) and one and half year at Bundesinstitute für Risikobewertung (Berlin, Germany), he joined Analytical Biochemistry group at University of Groningen (Groningen, The Netherlands), where he is working in the last 11 years currently in the position of Associate Professor. His research interests focus on computational mass spectrometry, proteogenomics data integration and biomarker discovery. He authored more than 60 peer-reviewed publications and book chapters, and he is editorial board member of Journal of Proteomics.