

The correspondence problem for metabonomics datasets

K. Magnus Åberg · Erik Alm · Ralf J. O. Torgrip

Received: 31 October 2008 / Accepted: 15 January 2009 / Published online: 7 February 2009
© Springer-Verlag 2009

Abstract In metabonomics it is difficult to tell which peak is which in datasets with many samples. This is known as the correspondence problem. Data from different samples are not synchronised, i.e., the peak from one metabolite does not appear in exactly the same place in all samples. For datasets with many samples, this problem is nontrivial, because each sample contains hundreds to thousands of peaks that shift and are identified ambiguously. Statistical analysis of the data assumes that peaks from one metabolite are found in one column of a data table. For every error in the data table, the statistical analysis loses power and the risk of missing a biomarker increases. It is therefore important to solve the correspondence problem by synchronising samples and there is no method that solves it once and for all. In this review, we analyse the correspondence problem, discuss current state-of-the-art methods for synchronising samples, and predict the properties of future methods.

Keywords Alignment · Warping · Chromatography · Metabolic profiling · NMR · Mass spectrometry (MS)

Introduction

This critical review focuses on the correspondence problem and its properties for metabonomics datasets. Starting from the properties of NMR and chromatography–mass spectrometry data, a selection of current state-of-the-art synchronisation methods are discussed. This review is intended as a guide to the problem and to the current attempts at



Magnus Åberg holds a Ph.D. in chemometrics and has been a researcher in the BioSysteMetrics Group at the Department of Analytical Chemistry at Stockholm University since 2006. His current research interests are developing algorithms and methods for maximizing information recovery from data, e.g. from metabonomics

solving it. Recent reviews dealing with this problem are Listgarten and Emili [1] and Vandenberg et al. [2]. The review of Listgarten has a wider scope—statistical methods for comparative proteomic profiling. Vandenberg reviews alignment of LC–MS images with focus on proteomics and detection of biomarkers. In this review we give a more in-depth description of the correspondence problem with focus on metabonomics data from NMR and LC–MS.

What is correspondence?

The correspondence problem is about arranging things in their proper place, i.e. putting the right values in the right rows and columns of a data table. An illustrative example is shown in Fig. 1. Suppose you want to compare suppliers of fruit baskets to your office and you have a preference for green apples. You would like to get the most fruit for your money but there must not be too few green apples. The fruit is sorted according to category and weighted. The weight data are summarized in a data table on which you are going to base your decision on which supplier to use. The table in Fig. 1 cannot be used for reliable decision-making because

K. M. Åberg (✉) · E. Alm · R. J. O. Torgrip
Department of Analytical Chemistry, BioSysteMetrics Group,
Stockholm University,
10691 Stockholm, Sweden
e-mail: magnus.berg@anchem.su.se





































Supplier No.	Orange	Red Apple	Green Apple	Pear	Banana
1					
2					
5					
3					
4					
6					
7					
8					

Fig. 1 Fruit data that illustrate the correspondence problem and different errors in peak alignment. The amount of fruit is proportional to the size of the image in the table

of errors with fruits in the wrong columns. The problem with statistical analysis of metabonomics datasets is fully analogous, but the errors are less obvious because you cannot tell the identity of a metabolite by looking at a peak. In *real* metabonomics datasets (to be distinguished from datasets constructed to *test* biomarker detection) there is no known ground truth and the integrity of the data table can only be checked for obvious errors by inspecting the raw data. A synchronisation method should be without obvious errors, although, in our experience, obvious errors appear with most methods. It is often not feasible to check all the columns of the data table. Obvious assignment errors reduce the confidence in the assignments that cannot be judged.

What is metabonomics?

Metabonomics is concerned with non-targeted analysis of biofluids to obtain quantitative or semi-quantitative information about as many metabolites as possible. The biofluids most commonly analysed are plasma, serum, and urine [3–5], although in the literature there are also reports of analyses of, e.g., cerebrospinal fluid [6], sweat [7], and saliva [8].

Whenever non-targeted data are generated, the analysis is less controlled compared with targeted analyses and the correspondence problem becomes relevant. Shotgun, or label-free, proteomics has the same problem with assigning correspondence.

Metabonomics data and its properties

The most commonly used analytical platforms of metabonomics are NMR, LC–MS, and GC–MS [9]. The data from

the different platforms are associated with their particular properties and problems. The correspondence problem is similar for all platforms but there are differences. Whenever samples are measured there will be positional uncertainty in the signals. When only a few analytes are targeted this can be handled by the experimental procedure. In non-targeted analysis, it is more difficult to optimise the experimental procedure. The samples cannot be prepared by purification in the same way and, therefore, the sample matrix will have a greater effect on the observed data.

1D ^1H NMR

One-dimensional ^1H NMR spectra of blood plasma or urine takes the form of a forest of peaks in the region between 0 and 9 ppm. The widths of the peaks are mainly dependent on the field strength of the magnet (measured in MHz). Peak shapes are distorted from the ideal Lorentzian shape to something less symmetric by an inhomogeneous magnetic field or incomplete phase correction. The positions of peaks along the ppm axis are sensitive to, e.g., temperature, pH, and ionic strength [10]. It is, therefore, standard procedure to buffer the samples and control the temperature during data acquisition. More about the practical and instrumental issues of NMR for metabonomics can be found in a recent review by Fan and Lane [10].

Important properties of 1D ^1H NMR data:

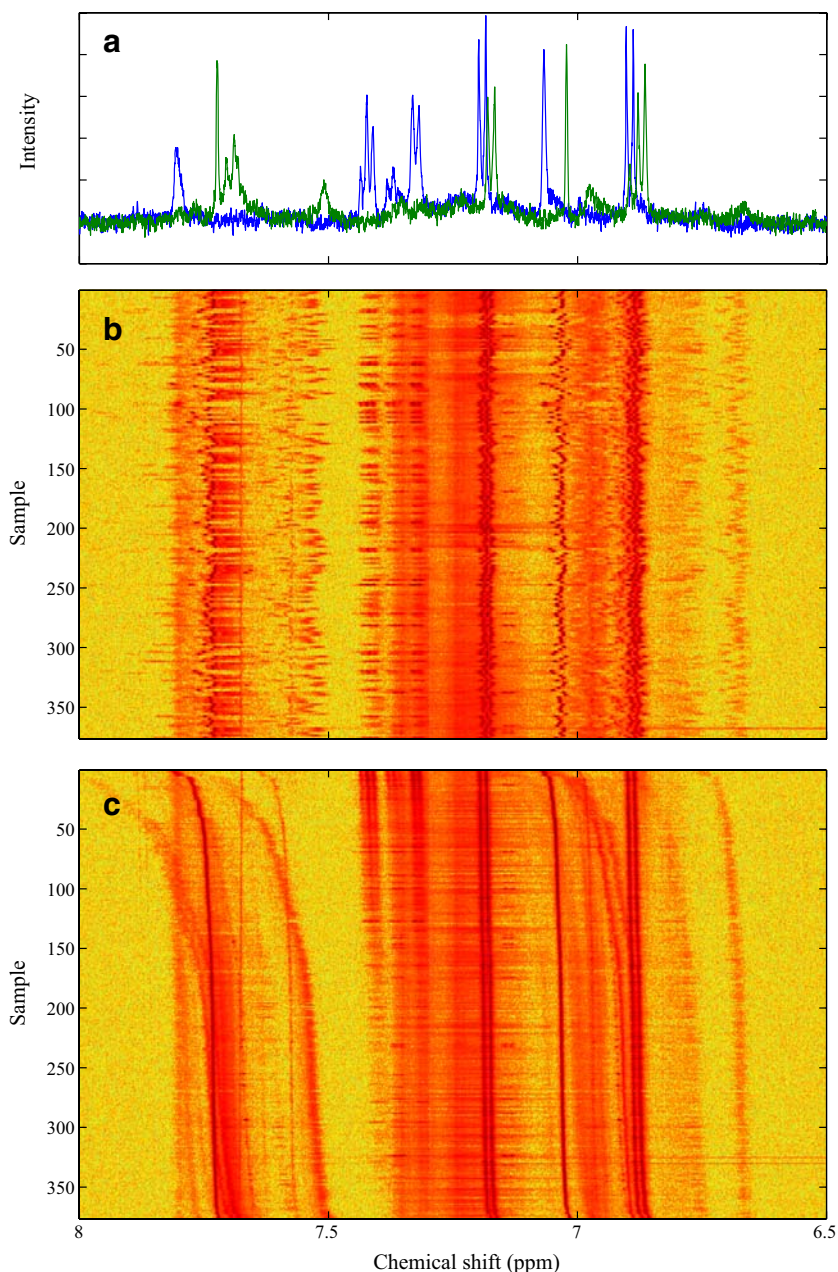
- Peak shifts are so large that peaks change order along the ppm axis.
- Peak-shape distortion by an inhomogeneous magnetic field may cause problems with peak detection—deconvolution may give false-positive peaks if a symmetrical peak shape is assumed.
- Limit of detection is quite high compared with, e.g., mass spectrometry.
- A metabolite usually has many different protons giving rise to several peaks in the NMR spectrum; many peaks will be also be split into multiplets.

The problem of correspondence for NMR is illustrated in Fig. 2, in which two extreme spectra are shown in (a) and a heat map of spectra in order of acquisition is shown in (b). In (c), the samples of (b) have been ordered on the basis of the histidine peak at approximately 7.03 ppm. The sorting reveals that the shifts are structured and that some peaks change order along the ppm axis.

LC–MS and GC–MS (full-scan MS)

Metabonomics data from LC–MS instruments are often acquired with electrospray ionization and a high-resolution mass analyser, e.g. time-of-flight, orbitrap, or ion-cyclotron-resonance. With GC–MS it is common to use

Fig. 2 NMR spectra viewed as heat maps in which a row represents a sample and each column is a ppm value; the intensity is colour coded. **(a)** Spectra from the top and bottom rows of **(c)**. **(b)** Spectra in order of acquisition. **(c)** Spectra ordered by the position of the histidine peak at about 7.03 ppm



electron ionisation with a quadrupole mass analyser giving unit resolution in the m/z dimension. The data have two measurement dimensions: retention time and m/z , where the retention time dimension has the most peak-shift problems.

Metabolomics LC–MS data are closely related to shotgun or label-free proteomics data. The basic data structure and problems with the data are the same. A difference is that for metabolomics a protein precipitation step may be performed rather than protein digestion. The observed mass range may differ—normally approximately 50–1000 m/z in metabolomics whereas in proteomics data is acquired in higher mass regions. The low-mass region is

advantageous because mass uncertainty increases with increasing m/z .

Properties of chromatography–MS data:

- Peaks seldom change order along the retention time axis, at least not peaks with the same m/z . It can happen for peaks with different m/z (an example is given in Fig. 3).
- A metabolite can have multiple signals in the m/z dimension. These signals come from isotopes, adducts, and fragments.
- The limit of detection of MS is generally lower than that of NMR.

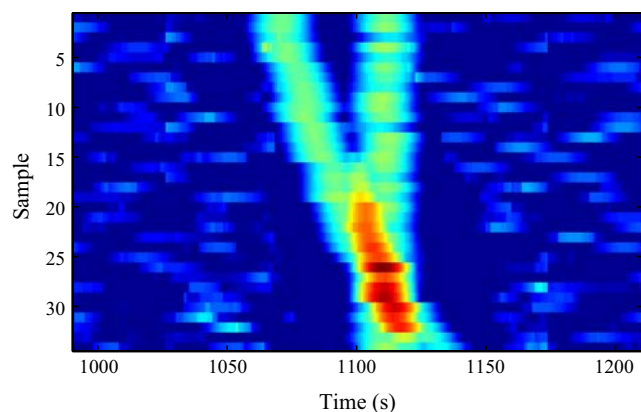


Fig. 3 Heatmap of 34 samples analysed by LC MS. The two peaks have different masses, $m/z = 316$ for the shifting peak and $m/z = 512$ for the non-shifting peak. Data from Ref. [11]

- Chromatography–MS instruments are less stable than NMR instruments which may cause increased run-to-run variability.
- The chromatographic separation can be affected by changes in pH, solvent composition, temperature, column ageing, etc. Also, the sample matrix is an important factor in retention-time differences between samples.

The correspondence problem for chromatography–MS data is less complicated than for NMR in respect of peaks changing order. On the other hand there are two dimensions to consider. Keeping isotopes, adducts, and fragments of a single metabolite together, while allowing peaks to change elution order may be difficult to combine in an algorithm.

Sources of peak shift

Peak shift can be attributed to three sources:

- instrument drift,
- the chemistry of the sample (matrix) and separation system (if any), or
- random variation.

Instrument drift as a source of peak shift should be relatively small; otherwise the experimental procedure could probably be improved. Nevertheless, there will always be an element of instrument drift in the data.

The largest source of peak shift is probably the chemistry of the sample matrix and the separation system. It is well known that peaks shift more in liquid chromatography than in gas chromatography. In GC the chemical processes are limited to the interaction between the stationary phase and the sample constituents. In LC there is an additional interaction with one or more solvents. The solvents may undergo changes with time. For instance, the pH may change slightly as a result of uptake of carbon

dioxide from air, and this could, in turn, affect the retention behaviour of an acidic compound with pK_a close to the pH of the solvent. This kind of instrument-related chemistry can be controlled. Thus, a good experimental procedure minimizes peak shift because of instrument-related chemistry. For both GC and LC, column degradation can cause peak shift and this cannot be prevented. There can also be an effect of column-to-column differences in the retention time of the same peak.

The chemistry of the sample cannot be controlled. The only way to obtain similar properties for different samples is to dilute them with buffer so that the original sample is a negligible fraction of the prepared sample. This ruins the possibility of detecting low-abundance metabolites irrespective of instrumental technique. The practical compromise is to add a small volume of buffer to the samples. This will not completely buffer the sample. A good example is the pH-sensitive citrate peaks which are notorious for shifting in 1D ^1H NMR spectra. They always shift, even when the samples are buffered.

The situation is not hopeless, chemistry is predictable and, therefore, there is a possibility of correcting shifts of chemical nature by the synchronisation method.

The part of the shift that cannot be attributed to either instrument drift or chemistry can be regarded as random. The random shift is, hopefully, small and can be handled with more or less any existing alignment or warping method.

1D ^1H NMR presents the greatest synchronisation challenge where peaks frequently change order along the ppm axis. In chromatography–MS data, the challenge is to effectively use the information provided by the mass axis in the best way, and to handle the larger amount of data.

The correspondence problem and methods for alignment

Ambiguous correspondence

Although most peaks may have an obvious correspondence, there are peaks for which there will be a question about which assignment is correct.

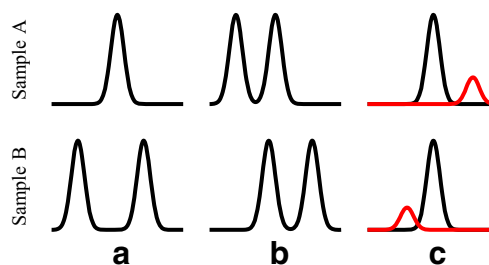


Fig. 4 Three cases of ambiguous assignments. (a) One peak in the first sample can match either of two peaks in the second sample. (b) Should both peaks be matched or just one? (c) Peaks changing order

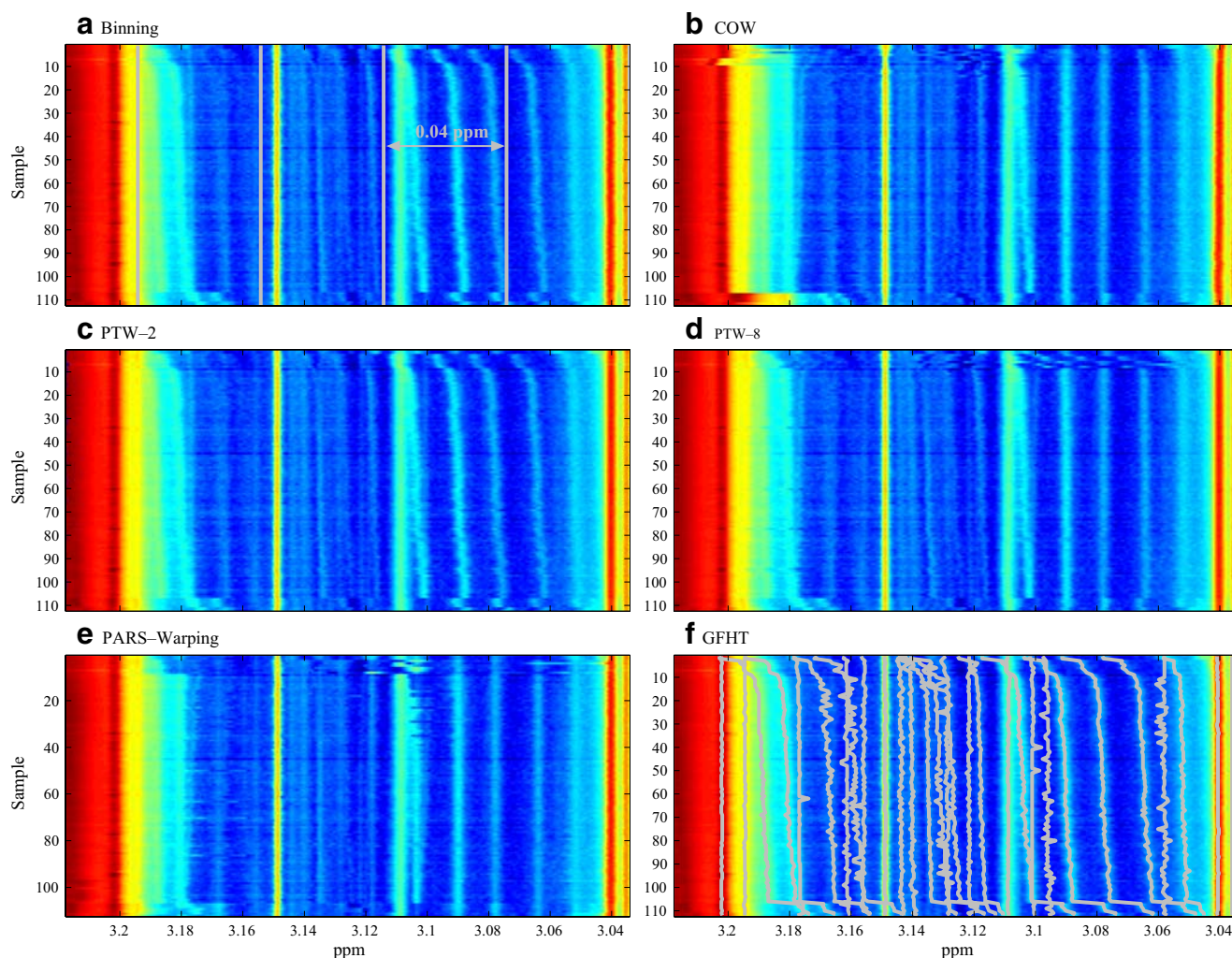


Fig. 5 Heatmaps of NMR spectra for 112 samples for comparison of alignment methods: (a) raw spectra (bin limits are shown as grey lines), (b) COW, (c) PTW-2, (d) PTW-8, (e) PARS warping, and (f) FGHT with corresponding peaks are connected by grey lines

We have identified three basic cases of ambiguous correspondence:

1. One peak in sample A can match either of two peaks in sample B (Fig. 4a).
2. Two peaks in sample A can match two peaks in sample B, but with little or no shifting of the peaks the last peak in A matches the first peak in B (Fig. 4b).
3. Peaks change order between samples A and B (Fig. 4c).

If there are only two samples one can probably afford to try all the different assignments. It is not feasible to try all assignments with multiple samples, because the number of possible combinations grows too rapidly.

The local environment of other nearby peaks may affect which assignment is made by the alignment or warping method. This can lead to the correct assignment but not

always—there are cases where the local environment makes a method assign the wrong correspondence (Fig. 5).

Information that can help (use of supporting information)

For two-dimensional data, the second dimension (m/z in chromatography–MS) can contain information that can facilitate correct peak assignment [12, 13]. Curve resolution is a good choice for GC–electron ionization MS in which several metabolites have mass peaks in common because of extensive fragmentation [13]. In LC–electrospray MS, in which fragmentation is limited, deisotoping and deadducting can help produce second-order support. Unfortunately, isotope ratios are similar for metabolites of similar mass and exactly the same for metabolites with the same elemental composition. Adduct formation may be more discriminative between different chemical species, although the discriminatory power is likely to be low. An advantage of deisotoping

and deadducting is that isotopes and adducts can be kept together as pseudo-metabolites. The risk that monoisotopic peak and peaks of higher isotopes or adducts are aligned differently is eliminated and signal-to-noise is also improved.

The sample dimension can also be helpful. It may be difficult to discover which the adduct peaks are, but differences in retention time and intensity between samples can help. Tentative adduct, fragment, and isotope peaks can be tested by correlating the retention time pattern and the intensity pattern with those of the monoisotopic peak. If correlations are high, the tentative peak belongs to the same pseudo-metabolite with high probability, otherwise not. An example, and more detail, are given in Ref. [14].

Inclusion of information about sample grouping in the alignment algorithms is recommended in Ref. [1]. This information can indeed improve the alignment results but needs to be used with care because it can introduce bias into subsequent statistical analysis. In the extreme, the peaks of a differently expressed metabolite may end up as two different metabolites because of the grouping of the intensities. The reader with a bias towards statistical subtleties might enjoy the book “Subset selection in regression” by Miller [15] in which he analyses the subject of bias in estimation.

In NMR, all the peaks of a multiplet will have the same shift. The multiplet structure of proton signals can be used to improve the possibility of correct assignment.

Methods for solving the correspondence problem

Most methods for solving the correspondence problem can be classified into one of the following different approaches, some more naive than others: binning, nearest-neighbour clustering, warping, and combinations of these.

Every method that compares different alternatives has an objective function that may be explicitly or implicitly defined. Explicit objective functions are to be preferred because they make the algorithms easier to understand mathematically. If the warping alternatives are evaluated by algorithmic rules, the objective function is implicitly defined. That may be the natural human way of trying to solve the problem by reasoning but it makes the methods less transparent and more difficult to understand.

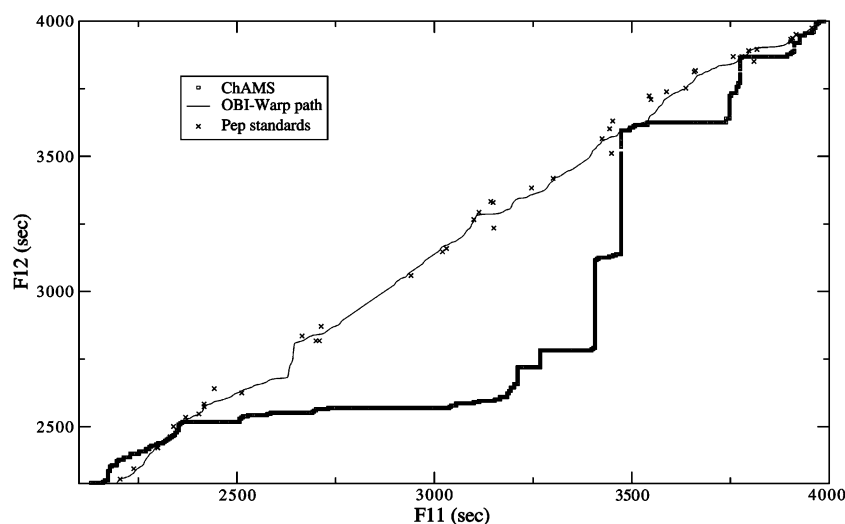
Binning

The simplest and most naive method is binning. In the past, binning has been used frequently, and it still is [16–21]. In binning, the measurement dimensions are divided into segments and each segment is assigned a single number that summarizes the data. The number can, for instance, be the integral of the intensity or the maximum intensity. Bin

widths are chosen to be greater than the expected peak shifts. For NMR metabonomics, the golden standard was binning with bins 0.04 ppm wide. Recent improvements to binning include mean filtering before using bilinear modelling (e.g. PCA) [20] and adaptive binning in which bin limits are located in minima of superpositioned data [19]. Binning can circumvent ambiguity of type 1 by combining the two peaks in B in the same bin. This will partly destroy the information about the individual peaks and hamper statistical analysis. If the same metabolite has many peaks, as in NMR, information about the individual peak intensities can be revealed by use of bilinear modelling. Bilinear modelling can only partially overcome the problem with multiple peaks in a single bin. For example, information about a small peak next to a large peak may be completely lost because of fluctuations of the intensity of the larger peak. Binning can also avoid ambiguity of type 2 by use of large enough bins. If both peaks in the two samples end up in the same bin, the problem is solved. In the same way, ambiguity of type 3 can be avoided. The problem with binning is the trade off between loss of information and solving the correspondence problem. Information is lost when one bin contains more than one peak. To solve the correspondence problem the bin limits must be chosen with care. It may even be that correspondence is impossible to achieve with reasonable bin sizes. The special case of a single bin always solves the correspondence problem but is useless for data analysis. Uniformly distributed bins of the same size will almost certainly split peaks so that they end up in two adjacent bins. Using minima of a superposition or another combination of data from all samples is not guaranteed to solve ambiguous correspondences correctly.

Binning can be excellent for quickly obtaining an overall picture of a dataset. One can only expect to find big changes using binning. Small but significant changes will probably be obscured because of loss of information. Many methods have been compared with binning and shown to be superior [22–24]. The reason why binning is still used today is that, so far, no method has proven to be sufficiently easy to use and sufficiently successful in producing better results. An example is seen in Prince et al. [25], in which the authors use the method ChAMS [26], which performs poorly with obvious and gross errors. It is obvious that ChAMS has a warp path which is too flexible (Fig. 6). Examples of this kind may scare the non-expert from using warping or peak alignment methods for sample synchronisation. The non-expert sticks with binning because it is the “fast food” of synchronisation (you know what you get and that you get it right away) rather than trying to be chefs of haute cuisine by using state-of-the-art methods. Binning is not the best approach but has acceptable worst-case performance.

Fig. 6 Comparison of OBI-Warp and ChAMS, where ChAMS is obviously too flexible. Known correspondences are marked with \times symbols. (Reproduced with permission from Anal Chem (2006) 78:6140–6152. Copyright 2006 American Chemical Society)



Nearest-neighbour clustering

Next to binning, perhaps the most common way of solving the correspondence problem is by nearest-neighbour clustering. In many papers different forms of clustering is used [7, 14, 27–30]. Some methods allow assignment of two or more peaks from one sample to a cluster [28, 30], resolving this “collision” later [28]. Not all methods needing collision resolution use it [31]. Clustering is normally performed on peak lists and is closely related to binning in minima of superpositioned data.

When it comes to handling the three cases of ambiguous correspondence, clustering is quite a naive method. For the first case the results will be sensitive to the exact position of the peak in sample A and the peak is likely to get different assignments in different samples. In the second case, the last peak in A will always match the first peak in B whether this is correct or not. The third case where peaks change order cannot be assigned correctly. At most, one of the peaks can be correctly assigned.

Warping

Warping is the term for transforming the measurement dimension of samples to achieve correspondence. The simplest example of warping is to offset the retention-time axis of one sample so that the retention times of the peaks better match those in a second sample, called the reference sample. More advanced transformations can, e.g., be linear [32], a second order polynomial [33], piece-wise linear [34–36], or based on b-splines [37]. Warping functions are normally required to be monotonous. Otherwise, loops can be created on the measurement axis. Loops are undesirable and lack physical meaning. Warping methods make sense physically and chemically because peak shifts in, e.g., chroma-

tography are often correlated. If one peak shifts to a later retention time, neighbouring peaks are likely to do the same, especially if the peak shift depends on instrument-related chemistry such as a slight difference in LC gradient or column degradation. For NMR the situation is more complicated. Some peaks may shift substantially whereas others with almost the same ppm do not shift at all (Fig. 5). The warping function can be estimated from the raw data [38, 39], from peak lists with pre-specified correspondences [40], or from peak lists without pre-specified correspondences [41].

Warping using raw data Many warping methods which use the raw data are based on dynamic programming to evaluate different solutions. Dynamic programming (DP) is a method of solving optimisation problems by dividing it into smaller subproblems and using recursion to construct the globally optimum solution [42]. The typical DP problem is to find the shortest path between two points in a connected graph and the solutions to the correspondence problem can be posed as a graph. Early examples of DP-based warping for synchronising samples in chemistry is dynamic time warping (DTW) [38] and correlation optimised warping (COW) [39]. A very good introduction to both methods is available elsewhere [34]. There are a number of recent uses and modifications of DTW [25, 26, 43] and COW [37, 44–46].

There are a few methods that are based on local segment-wise optimisations without trying to achieve a globally optimum solution [23, 47, 48]. Compared with the most closely related DP-based method COW, their only merit is perhaps their computational speed. For comprehensive separations, e.g. LC \times LC, this seems to be the only type of method available [49, 50].

Parametric time warping [33] and semi-parametric time warping [37] use continuous functions to warp the time

dimension by iteratively minimising the squared difference between a reference and a sample chromatogram.

The continuous profile model (CPM) method by Listgarten et al. [51, 52] is a hidden-Markov model for estimating a true unobserved chromatogram using expectation maximization. The method was initially designed to align replicate samples but has been extended to align samples of different origin [52]. The method is perhaps the only example of a warping method that does not use a reference sample—a very tractable feature.

Warping of peak lists Warping using data in the form of peak lists is perhaps the most diverse field of warping. Two different approaches can be identified—the landmark peaks approach and the tentative assignment approach. In the landmark peaks approach, the warping function is estimated from landmark peaks with known correspondence. After warping peak correspondences are reassigned. Some algorithms stop here [40], others refine the warping function iteratively. The first landmark peaks can be found by clustering [40, 53, 54] or, in proteomics, by LC–MS–MS identification [55–57]. The tentative assignment approach makes a list of possible assignments and uses this list to estimate the warping function by robust regression [35, 55].

Warp2D [46] extends COW into two dimensions with overlap between Gaussians in the objective function.

A few methods use successive pair-wise alignments to avoid specifying a reference sample [58–60].

Peak alignment followed by warping All methods that align peaks in one dimension can be used for computing a warping function. Johnson et al. [61] use nearest-neighbour clustering followed by piecewise linear interpolation between matched peaks to find a warping function. PARS [24] uses tentative assignments and constructs a graph problem that mimics DTW and is solved by dynamic programming. It is not originally a warping method but a peak alignment method. For PARS, linear interpolation between matched peaks produces a piecewise linear warping function.

Warping and ambiguous correspondence For the first case of ambiguous assignment, warping methods have a greater chance of finding the correct correspondence than clustering methods—if there is local support from nearby peaks with unambiguous correspondence. Then, warping methods can be expected to find the true solution. If the warping function is too flexible, however, the warping will collapse to a nearest-neighbour method.

The second case of ambiguous assignment can, often, also be handled by warping, again because of support from unambiguous matches.

The main drawback of warping methods is the necessary requirement of a monotonic warping function. Warping can,

therefore, not find the correct correspondence for the third case with peaks changing order. The exception is if peaks change order in LC–MS and the peaks have different masses and the mass channels are warped independently.

Warping does not always completely solve the correspondence problem. If the data are to be modelled by bi or trilinear methods warping the raw data is sufficient. For further statistical analysis of peak lists, peak detection may be needed and nearest-neighbour clustering is always needed. In many warping methods for peak lists, nearest-neighbour clustering is included in the method and need not be performed explicitly after warping.

Image-processing methods An interesting and promising, but currently immature, approach inspired by image registration is *amsrpm* [41] which uses robust point matching [62] to align peak lists. The method is very slow and it is, therefore, difficult to optimise its parameters [56]. *Amsrpm* uses fuzzy correspondence and simulated annealing to match samples to a reference. The method can also align total-ion chromatograms. Its ambiguity-resolving properties are similar to those of other warping methods.

The generalized fuzzy Hough transform method [22] (GFHT) is based on a method for detecting shapes and objects in images which has been adapted for NMR-data. The key features of the GFHT method is that it can solve all three ambiguities by pre-calibrating the model on the shifts of peaks with known correspondence.

Algorithm/method symmetry The methods derived from DTW use one sample as a target to which all other samples are aligned. This makes the algorithms simple to understand and implement. The drawback is that the method becomes asymmetric in relation to the samples. Choosing a different target sample may affect the alignment or warping results. It is a desirable feature that an algorithm solving the correspondence problem is symmetric so that there is a unique solution that does not depend on an arbitrary choice of target. Creating compound targets based on all samples is likely to degrade the results unless performed carefully. Shifting peaks will be blurred or even lost if averaging over NMR-spectra or TICs is used.

Today, there are a few symmetric methods available for more than two samples. The GFHT [22] and CPM [52] are examples of symmetric methods. Some warping methods that use tentative assignments of peak lists and regression to fit warping functions are symmetric, e.g. XCMS [40] has symmetric warping as an option.

Using raw data vs. peak lists There are alignment methods that use the raw data and methods that use peak lists. The main argument for using the raw data is that you are not dependent on a peak-detection step which might introduce

errors by failing to detect peaks. The reasons for using peak lists are that they present the relevant information more compactly and with improved signal-to-noise ratio from averaging when integrating a peak.

From experience, we have observed that the number of peaks increases with decreasing intensity approximately as $\#peaks \propto 1/intensity$. Pushing the limit of data analysis requires that you accurately detect, align, and further analyse peaks very close to the limit of detection. Accurate detection would mean that an experienced experimentalist confirms that you have very few false positive and very few false negative peaks in your peak list. A high fraction of false positives and false negatives will severely disturb the alignment results. Small peaks may be difficult to align using raw data, unless they are automatically aligned by surrounding larger peaks.

The ambiguity where peaks change order can possibly be resolved by aligning peak lists with support from an extra dimension. By warping the raw data, peaks can never be shifted around each other. For LC–MS, warping the raw data often requires binning of the m/z axis which destroys information that can be used for alignment.

Example of warping 1D 1H NMR data A dataset consisting of NMR spectra from 112 replicate quality-control samples of human plasma [22] is used to demonstrate different methods assigning correspondence in NMR data. In this data there are approximately 30 peaks and several examples of peaks changing order along the ppm axis. The methods demonstrated are: binning, PTW (second-order polynomial), PTW (eighth-order polynomial), COW, PARS-warping, and GFHT. The results are shown as heat maps in Fig. 5. Notice how 0.04-ppm-wide bins all contain more than one peak. A trained eye can find ten peaks in the third bin from the right. PTW-2 improves the alignment but is not flexible enough. With PTW-8 it seems that most peaks are properly aligned but there are a number of erroneous assignments. The same is true for COW, which performs well for most samples but gives strange results for some of the extreme samples. PARS-warping performs worse than COW by being too flexible, which results in alignment errors. PARS-warping is performed on peak list data with about 25 peaks per sample. The GFHT is calibrated on the pattern of peak shifts of the histidine peaks at 7.03 ppm and it finds the true correspondence in most cases. Note especially how it handles the peaks changing order along the ppm-axis. For NMR, it is important how the correspondence problem is solved. The large fraction of peaks changing places (up to 30%, not shown [22]) may be one explanation of why NMR is losing ground to LC–MS in metabonomics. It has been impossible to get the correspondence right for NMR datasets. Another explanation is the relatively low sensitivity of NMR.

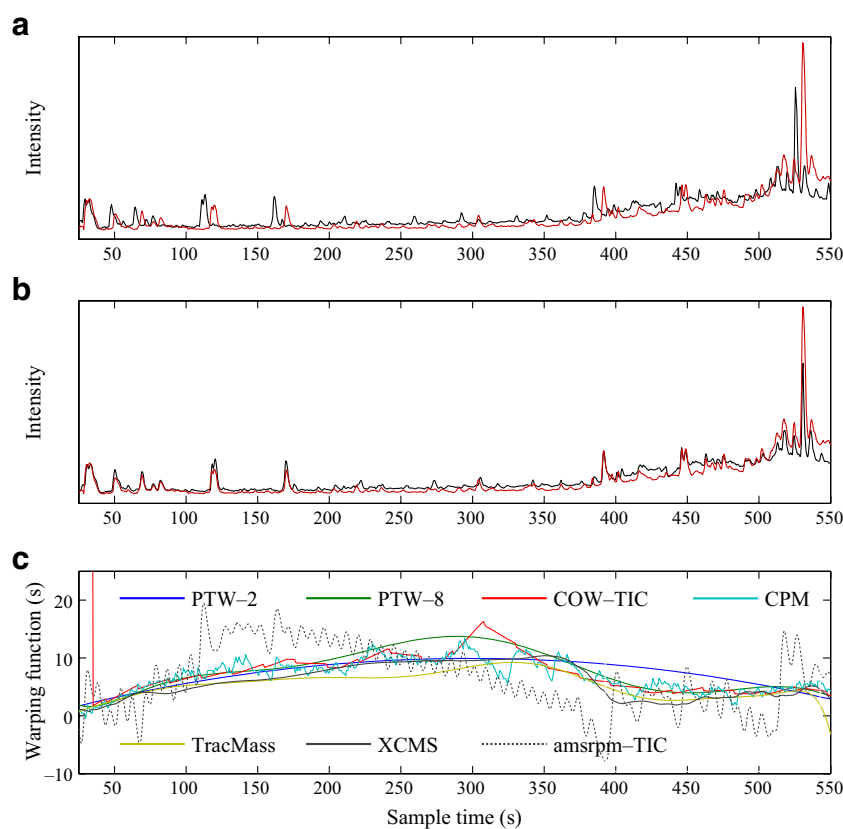
Example of warping LC–MS data The example of warping LC–MS data is limited to comparing the warping functions of seven different methods for warping two samples of blood plasma [14]. There is no ground truth other than that which can be guessed from the chromatograms. The point here is to show how similar the warping functions of different methods are. The methods included in this example are: 1. PTW-2, 2. PTW-8, 3. COW-TIC, 4. CPM, 5. TracMass [14], a nearest neighbour assignment on peak lists followed by least-squares fit of a twelfth-order polynomial (a naive warping method), 6. XCMS, which aligns approximately 3,000 peaks and computes a warping function using robust regression, and 7. amsrpm-TIC.

The warping results are presented in Fig. 7. TICs of a sample and a reference are overlaid in Fig. 7a, and the TICs of the sample warped by PTW-8 and the reference sample are overlaid in Fig. 7b. The warped time difference is shown in Fig. 7c (positive values mean peaks are shifted to later time points). It is obvious that the warping function is non-linear, and that most methods find more or less the same solution. PTW-2 is very good in the first half of the chromatogram but cannot follow the transition at 375 s. The differences between the other methods are relatively small. XCMS and TracMass compute a warping function which is too constrained; peak correspondence is much improved but is implied rather than there being complete overlap of peaks (not shown). It is definitely not good enough for bi or trilinear modelling. With COW, PTW-8, and CPM most high-intensity peaks overlap almost completely. Amsrpm is the only method that disagrees on the warping function. The chromatograms warped by amsrpm look quite good but amsrpm has been run with suboptimum parameters because the sampling rate differed between the example data and our data. We find the method is slow and therefore difficult to optimise.

Validation of alignment results

Alignment results are inherently difficult to validate. The best validation is to compare the assignments of a method to a known ground truth. Because the instruments create the correspondence problem it is impossible to obtain real data with known ground truth. The value of using synthetic data is limited, because of the difficulty in accurately reproducing all the peak-shift artefacts present in real data. Still, it may be one of the most illustrative ways of demonstrating the properties of a synchronisation method. The second best method is to use spike-in experiments where the spike-in mixture is analysed separately and samples are analysed both with and

Fig. 7 Comparison of warping functions for LC MS data. **(a)** Uncorrected sample (*black*) and target (*red*) TICs. **(b)** Warped sample (*black*) and target (*red*) TICs. **(c)** Warping functions computed by different methods



without the spike-in mixture [40, 52]. The problem with spike-in experiments is that the results for the alignment method become confounded with all other data-processing steps. Spike-in experiments are ideal for validating an entire pipeline for biomarker detection. Two public spike-in datasets are available—one for metabolomics by Nordström et al. [63] and one for proteomics by Listgarten et al. [52].

De Souza et al. [30] validate their results on a subset where the ground truth was manually defined by visual inspection. It is not a bad approach if the subset is representative of the whole dataset. Manual alignment can be as good as any other alignment method in use today.

Many papers use other indirect methods of validation that may not be completely relevant in that they are not based on the number of correct and erroneous assignments [24, 31, 54, 58, 64, 65].

In proteomics it is not uncommon to perform LC-MS-MS on some peaks. The results can be used by the alignment method [55–57] or for checking the alignment results from aligning only on the LC-MS part of the data [25].

For NMR (and LC-MS) the samples can be sorted on the position of a shifting, yet easily assigned, peak. This

can give a very good indication of the ground truth for peak correspondence, c.f. Fig. 2 and Fig. 5.

If I were hunting for biomarkers today (and not researching the correspondence problem) (Conclusions)

For NMR the correspondence problem is more difficult than for chromatography-MS because in NMR changes in peak order are common whereas in chromatography-MS they are uncommon. As far as we are aware, there is only one method that can handle changes in peak order for 1D data and that is the GFHT method. For chromatography-MS the competition is harder—many methods are almost equally good. A combination of non-linear warping and peak alignment is probably the way to go.

To identify biomarkers, we recommend that:

- Several methods are used, at least one working on peak lists and one working on the raw data (warping). Do not forget to use binning for a preliminary investigation and to protect against a worst-case scenario.
- The method variables are adjusted for the analytical platform in question using two or more data sets.

- With every new dataset the methods are run quick-and-easy with the pre-adjusted variables.
- Top candidate biomarkers from each method are checked manually using the raw data.

Future outlook

For NMR we expect the future holds methods competing with the generalized fuzzy Hough transform by being able to correctly assign peaks that change order along the ppm axis.

For chromatography–MS, we believe that more warping methods that warp both time and mass will appear. The field seems to progress toward multistage alignment methods where the warping function is iteratively refined by steps of warping and reassignment.

Acknowledgements The authors are thankful to AstraZeneca for financing and for access to metabolomics data from LC–MS and NMR. Helena Idborg is acknowledged for supplying the data for Fig. 3.

References

- Listgarten J, Emili A (2005) *Mol Cell Prot* 4:419–434
- Vandenbogaert M, Li-Thiao-Te S, Kaltenbach HM, Zhang RX, Aittokallio T, Schwikowski B (2008) *Proteomics* 8:650–672
- Nicholson JK, Wilson ID (1989) *Prog Nucl Magn Reson Spectrosc* 21:449–501
- Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HWL, Clarke S, Schofield PM, McKilligin E, Mosedale DE, Grainger DJ (2002) *Nat Med* 8:1439–1445
- Shockcor JP, Holmes E (2002) *Curr Top Med Chem* 2:35–51
- Wishart DS, Lewis MJ, Morrissey JA, Flegel MD, Jeroncic K, Xiong Y, Cheng D, Eisner R, Gautam B, Tzur D, Sawhney S, Bamforth F, Greiner R, Li L (2008) *J Chromatogr B* 871:164–173
- Dixon SJ, Breerton RG, Soini HA, Novotny MV, Penn DJ (2006) *J Chemom* 20:325–340
- Yan S-K, Wei B-J, Lin Z-Y, Yang Y, Zhou Z-T, Zhang W-D (2008) *Oral Oncol* 44:477–483
- Nicholson JK, Connelly J, Lindon JC, Holmes E (2002) *Nat Rev Drug Discov* 1:153–161
- Fan TWM, Lane AN (2008) *Prog Nucl Magn Reson Spectrosc* 52:69–117
- Idborg H (2007) Analysis of metabolites in complex biological samples using LC/MS and multivariate data analysis. PhD Thesis, Stockholm University, Stockholm
- Idborg-Björkman H, Edlund PO, Kvalheim OM, Schuppe-Koistinen I, Jacobsson SP (2003) *Anal Chem* 75:4784–4792
- Jonsson P, Johansson AI, Gullberg J, Trygg JAJ, Grung B, Marklund S, Sjöström M, Antti H, Moritz T (2005) *Anal Chem* 77:5635–5642
- Åberg KM, Torgrip RJO, Kolmert J, Schuppe-Koistinen I, Lindberg J (2008) *J Chromatogr A* 1192:139–146
- Miller AJ (1990) Subset selection in regression. Chapman and Hall, London
- Sun J, Schnackenberg LK, Holland RD, Schmitt TC, Cantor GH, Dragan YP, Beger RD (2008) *J Chromatogr B* 871:328–340
- De Meyer T, Sinnaeve D, Van Gasse B, Tshiporkova E, Rietzschel ER, De Buyzere ML, Gillebert TC, Bekaert S, Martins JC, Van Crieckinge W (2008) *Anal Chem* 80:3783–3790
- Anderson PE, Reo NV, DelRaso NJ, Doom TE, Raymer ML (2008) *Metabolomics* 4:261–272
- Davis RA, Charlton AJ, Godward J, Jones SA, Harrison M, Wilson JC (2007) *Chemom Intell Lab Syst* 85:144–154
- Danielsson R, Backstrom D, Ullsten S (2006) *Chemom Intell Lab Syst* 84:33–39
- Jonsson P, Bruce SJ, Moritz T, Trygg J, Sjöström M, Plumb R, Granger J, Maibaum E, Nicholson JK, Holmes E, Antti H (2005) *Analyst* 130:701–707
- Csenki L, Alm E, Torgrip RJO, Åberg KM, Nord LI, Schuppe-Koistinen I, Lindberg J (2007) *Anal Bioanal Chem* 389:875–885
- Forshed J, Schuppe-Koistinen I, Jacobsson SP (2003) *Anal Chim Acta* 487:189–199
- Torgrip RJO, Åberg M, Karlberg B, Jacobsson SP (2003) *J Chemom* 17:573–582
- Prince JT, Marcotte EM (2006) *Anal Chem* 78:6140–6152
- Prakash A, Mallick P, Whiteaker J, Zhang HD, Paulovich A, Flory M, Lee H, Aebersold R, Schwikowski B (2006) *Mol Cell Prot* 5:423–432
- Luedemann A, Strassburg K, Erban A, Kopka J (2008) *Bioinformatics* 24:732–737
- Duran AL, Yang J, Wang LJ, Sumner LW (2003) *Bioinformatics* 19:2283–2293
- Tibshirani R, Hastie T, Narasimhan B, Soltys S, Shi GY, Koong A, Le QT (2004) *Bioinformatics* 20:3034–3044
- De Souza DP, Saunders EC, McConville MJ, Likic VA (2006) *Bioinformatics* 22:1391–1396
- de Groot JCW, Fiers M, van Ham R, America AHP (2008) *Proteomics* 8:32–36
- Lange E, Gropl C, Schulz-Trieglaff O, Leinenbach A, Huber C, Reinert K (2007) *Bioinformatics* 23:1273–1281
- Eilers PHC (2004) *Anal Chem* 76:404–411
- Tomasi G, van den Berg F, Andersson C (2004) *J Chemom* 18:231–241
- Palmblad M, Mills DJ, Bindschedler LV, Cramer R (2007) *J Am Soc Mass Spectrom* 18:1835–1843
- Walczak B, Wu W (2005) *Chemom Intell Lab Syst* 77:173–180
- van Niderkassel AM, Daszykowski M, Eilers PHC, Heyden YV (2006) *J Chromatogr A* 1118:199–210
- Kassidas A, MacGregor JF, Taylor PA (1998) *Aiche J* 44:864–875
- Nielsen NPV, Carstensen JM, Smedsgaard J (1998) *J Chromatogr A* 805(1–2):17–35
- Smith CA, Want EJ, O’Maille G, Abagyan R, Siuzdak G (2006) *Anal Chem* 78:779–787
- Kirchner M, Saussen B, Steen H, Steen JAJ, Hamprecht FA (2007) *J Stat Soft* 18:4
- Dynamic programming. http://en.wikipedia.org/wiki/Dynamic_programming (Accessed 26 Sept 2008)
- Baran R, Kochi H, Saito N, Suematsu M, Soga T, Nishioka T, Robert M, Tomita M (2006) *BMC Bioinformatics* 7:530
- Christin C, Smilde AK, Hoefsloot H CJ, Suits F, Bischoff R, Horvatovich PL (2008) *Anal Chem* 80:7012–7021
- Sadygov RG, Maroto FM, Huhmer AFR (2006) *Anal Chem* 78:8207–8217
- Suits F, Lepre J, Du PC, Bischoff R, Horvatovich P (2008) *Anal Chem* 80:3095–3104

47. Lee GC, Woodruff DL (2004) *Anal Chim Acta* 513:413–416
48. Yao WF, Yin XY, Hu YZ (2007) *J Chromatogr A* 1160:254–262
49. Fraga CG, Prazen BJ, Synovec RE (2001) *Anal Chem* 73:5833–5840
50. Pierce KM, Wood LF, Wright BW, Synovec RE (2005) *Anal Chem* 77:7735–7743
51. Listgarten J (2006) Analysis of sibling time series data: alignment and difference detection. University of Toronto, Toronto
52. Listgarten J, Neal RM, Roweis ST, Wong P, Emili A (2007) *Bioinformatics* 23:E198–E204
53. Bellew M, Coram M, Fitzgibbon M, Igra M, Randolph T, Wang P, May D, Eng J, Fang RH, Lin CW, Chen JZ, Goodlett D, Whiteaker J, Paulovich A, McIntosh M (2006) *Bioinformatics* 22:1902–1909
54. Vorst O, de Vos CHR, Lommen A, Staps RV, Visser RGF, Bino RJ, Hall RD (2005) *Metabolomics* 1:169–180
55. Fischer B, Grossmann J, Roth V, Gruissem W, Baginsky S, Buhmann JM (2006) *Bioinformatics* 22:E132–E140
56. Fischer B, Roth V, Buhmann JM (2007) *BMC Bioinformatics* 8 (Suppl 10):S4
57. Jaffe JD, Mani DR, Leptos KC, Church GM, Gillette MA, Carr SA (2006) *Mol Cell Prot* 5:1927–1941
58. Åberg KM, Torgrip RJO, Jacobsson SP (2004) *J Chemom* 18:465–473
59. Sauve AC, Speed TP (2004) Normalization, baseline correction and alignment of high-throughput mass spectrometry data. *Proc Gensips*
60. Toppo S, Roveri A, Vitale MP, Zaccarin M, Serain E, Apostolidis E, Gion M, Maiorino M, Ursini F (2008) *Proteomics* 8:250–253
61. Johnson KJ, Wright BW, Jarman KH, Synovec RE (2003) *J Chromatogr A* 996:141–155
62. Chui H (2001) Non-rigid point matching: algorithms, extensions and applications. PhD Thesis, Yale University, New Haven
63. Nordström A, O'Maille G, Qin C, Siuzdak G (2006) *Anal Chem* 78:3289–3295
64. Skov T, van den Berg F, Tomasi G, Bro R (2006) *J Chemom* 20:484–497
65. Wu W, Daszykowski M, Walczak B, Sweatman BC, Connor SC, Haseldeo JN, Crowther DJ, Gill RW, Lutz MW (2006) *J Chem Inf Model* 46:863–875