analytical chemistry

pubs.acs.org/ac

Article

MSIWarp: A General Approach to Mass Alignment in Mass Spectrometry Imaging

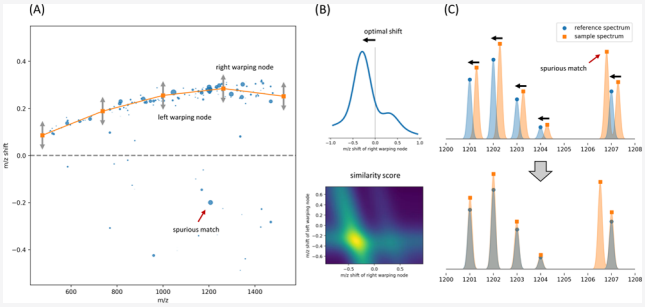
Jonatan O. Eriksson, Alejandro Sánchez Brotons, Melinda Rezeli, Frank Suits, György Markó-Varga, and Peter Horvatovich*

Cite This: *Anal. Chem.* 2020, 92, 16138–16148

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Mass spectrometry imaging (MSI) is a technique that provides comprehensive molecular information with high spatial resolution from tissue. Today, there is a strong push toward sharing data sets through public repositories in many research fields where MSI is commonly applied; yet, there is no standardized protocol for analyzing these data sets in a reproducible manner. Shifts in the mass-to-charge ratio (m/z) of molecular peaks present a major obstacle that can make it impossible to distinguish one compound from another. Here, we present a label-free m/z alignment approach that is compatible with multiple instrument types and makes no assumptions on the sample's molecular composition. Our approach, MSIWarp (https://github.com/horvatovichlab/MSIWarp), finds an m/z recalibration function by maximizing a similarity score that considers both the intensity and m/z position of peaks matched between two spectra. MSIWarp requires only centroid spectra to find the recalibration function and is thereby readily applicable to almost any MSI data set. To deal with particularly misaligned or peak-sparse spectra, we provide an option to detect and exclude spurious peak matches with a tailored random sample consensus (RANSAC) procedure. We evaluate our approach with four publicly available data sets from both time-of-flight (TOF) and Orbitrap instruments and demonstrate up to 88% improvement in m/z alignment.

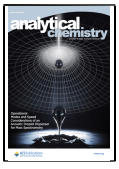


INTRODUCTION

Mass spectrometry (MS) is a widespread analytical technique used to detect and quantify ionized molecules, and it has many applications in biology, chemistry, and medicine. In MS imaging (MSI), molecular ions are sampled from different locations on a surface area, such as a tissue section, allowing the mass spectrometer to serve as a molecular imaging device. The ability to determine the spatial distribution of thousands of biological compounds in a single experiment makes MSI a powerful tool for tissue characterization. There have been extensive developments in the MSI field during the last decades, resulting in new experimental workflows, improved ionization and sampling methods, and advances in both the spatial and mass resolutions of instruments.[1−3] The availability of a wide variety of ionization techniques such as secondary-ion MS (SIMS), desorption electrospray ionization (DESI), and matrix-assisted laser desorption/ionization (MALDI) allows ionization of many compound classes with both targeted and untargeted approaches.[1] High-performance Orbitrap and Fourier transform ion cyclotron resonance (FT-ICR) mass analyzers can scan tissue sections with a subcellular spatial resolution and a mass resolution exceeding 500 000. Low sensitivity has been a longstanding obstacle for high-spatial-resolution MSI but has recently been improved by optimizing the laser wavelength in MALDI to increase ionization efficiency, or by post-ionizing neutral molecules to increase ion yield.[2] Novel sample preparation workflows have led to enhanced quantification and identification of metabolites, peptides, and proteins. These include protein extraction methods, in situ protease digestion of proteins, and the use of chemical derivatization such as labeling with photocleavable mass tags to enhance low-intensity molecule signals.[3,4] Altogether, this leads to highly complex data sets that demand accurate preprocessing and sophisticated bioinformatic analysis to maximize their utility in biological and clinical research.[5]

A persisting issue in MSI is systematic mass misalignment, leading to slight shifts in the m/z ratio of molecule peaks across spectra. These shifts can result in misidentified peaks or an increased risk of mixing peaks from different molecules with similar masses in the same ion image. Mass misalignment is typically more severe for time-of-flight (TOF) instruments than for FT-ICR, Orbitrap, or other Fourier transform (FT) instruments.[6] Variations in temperature throughout the

Received: September 9, 2020
Accepted: November 16, 2020
Published: December 2, 2020



https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

experiment, contractions and dilatations of the ion tube, contamination of the ion source, and tissue topography are some factors that are related to mass shifts in spectra generated by TOF instruments. For FT instruments, the most common source of mass misalignment is the space-charge effect, which causes mass shifts that increase with the number of ions in the trap.[7,8] The mass shifts in FT instruments can often be limited using automatic gain control (AGC), but can be considerable if suboptimal AGC settings are used.

When discussing mass misalignment and mass shifts, it is important to distinguish between relative mass alignment and absolute mass accuracy. Here, the former refers to how tightly peak masses are distributed across spectra, while the latter refers to the difference between a molecule's theoretical peak mass and its observed peak mass. A common approach to correct mass shifts is to perform either external or internal calibration by comparing the measured masses of predefined peaks to their expected theoretical masses. External calibration is performed by depositing a calibration standard outside the tissue region and comparing the measured peak masses to the known masses of the calibrants. The same calibration function is then applied to all tissue spectra. While this approach is simple, it does not reduce the relative misalignment or correct mass shifts introduced during the experiment. Internal calibration, on the other hand, relies on identifying known peaks in each tissue spectrum. The known peaks can be either prominent molecule peaks intrinsic to the tissue or peaks corresponding to calibrants sprayed on the whole tissue area. Internal calibration can improve both the relative alignment between spectra and absolute mass accuracy but performs poorly for spectra in which the calibration peaks are missing or incorrectly assigned.[9]

An alternative approach is to align all spectra to a common reference spectrum. Given that the absolute mass accuracy of the reference spectrum is high, this can reduce relative misalignment and improve absolute mass accuracy simultaneously. A simple approach to aligning one spectrum to another is fitting a polynomial recalibration function to the mass difference of their shared peaks and then using this recalibration function to recalibrate all peak masses. While this approach is flexible, it is highly sensitive to spurious peak matches. Bocker and Makinen[10] introduced a linear curve-fitting approach that is robust to spurious peak matches, and Kulkarni et al.[11] later used this approach, together with spatial information, to further improve mass alignment. More recently, there has been increased focus on mass alignment for TOF instruments. Ráfols et al.[6] proposed an alignment algorithm that uses the cross-correlation between two spectra in the upper and lower parts of the mass range to estimate mass shift. They then use the upper and lower shifts to recalibrate the mass axis of one of the spectra to that of the other. Boskamp et al.[9] elegantly exploit statistical properties of the peptide background signal to improve mass alignment and absolute mass accuracy. They estimate the mass shift across the m/z range by comparing observed to theoretical peak masses on the Kendrick mass scale. Unlike Ráfols et al.'s method, their method can correct nonlinear mass shifts, but the dependency on the peptide background limits its generalizability.

Another aspect of mass alignment is at which stage in the processing pipeline it is performed. It can be performed in the time domain for TOF spectra, in the frequency domain for FT spectra, using profile mass spectra, or using centroided mass spectra.[6,10,12,13] In principle, aligning spectra at an early stage is advantageous in the sense that errors are not accumulated in subsequent processing steps. In practice, however, time or frequency spectra are often inaccessible as vendor software typically only provide mass spectra, and the majority of data sets uploaded to repositories are processed to some extent. FT spectra, in particular, are often centroided to reduce their otherwise impractical size.[14] It is impossible to recover a raw spectrum from one that has been processed. Hence, an approach must be able to perform alignment with processed spectra to be compatible with most MSI data sets. This compatibility is essential; data sets generated independently in other labs are frequently used to validate biological findings or novel methods. A public data set may be generated with any instrument and additional processing is sometimes required to ensure its quality. This compatibility requirement, together with the growing popularity of public MSI data set repositories such as MetaboLights[15] or METASPACE,[16] creates a demand for algorithms that can perform accurate mass alignment on spectra acquired with multiple instrument types and regardless of whether they are already partially processed.

In this work, we adapt the correlation optimized warping (COW)[17] algorithm to perform label-free MSI mass alignment using a custom benefit function and show that we can greatly reduce variation in peak masses between spectra. COW aligns a pair of signals by performing local warpings on one signal so that the global similarity relative to the other is maximized. We have previously shown that COW is effective in reducing misalignment in the time dimension between liquid chromatography−mass spectrometry (LC-MS) sample runs.[18,19] Here, we instead use COW to reduce mass misalignment between spectra by warping the mass dimension.

Crucially, our method finds the optimal warping of one spectrum relative to another using only a list of centroided peaks from each spectrum. We model centroided peaks as Gaussians whose widths vary with m/z, and define the similarity between two spectra as the total overlap of their shared peaks. To assess and further improve COW's robustness for particularly peak-sparse or misaligned spectra, we include an optional outlier detection step in the form of a tailored random sample consensus (RANSAC)[20] procedure. The concept of our method is similar to that of Ráfols et al.,[6] but differs in two key aspects. First, we find the mass recalibration function by maximizing the product (overlap) of centroided peaks instead of the cross-correlation of continuous spectra. Second, our method can, since it is derived from COW, correct nonlinear mass shifts. Thereby, our method utilizes information about peak height and width (not simply mass location) while remaining compatible with most MSI data sets. We demonstrate the effectiveness and generalizability of our method, named MSIWarp, by applying it to four publicly available data sets. MSIWarp performs accurate and robust alignment with centroided spectra, is compatible with multiple instrument types, and makes no assumptions on the molecular composition of the sample. We provide a fast C++ implementation of MSIWarp together with a Python binding at https://github.com/horvatovichlab/MSIWarp.

THEORY

The core of MSIWarp is a spectrum-to-spectrum similarity score that is used to find an optimal warping function between the mass axis of one spectrum and that of another. To align an entire MSI data set, all spectra, the sample spectra, are warped to a common reference spectrum that can be selected from the

https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

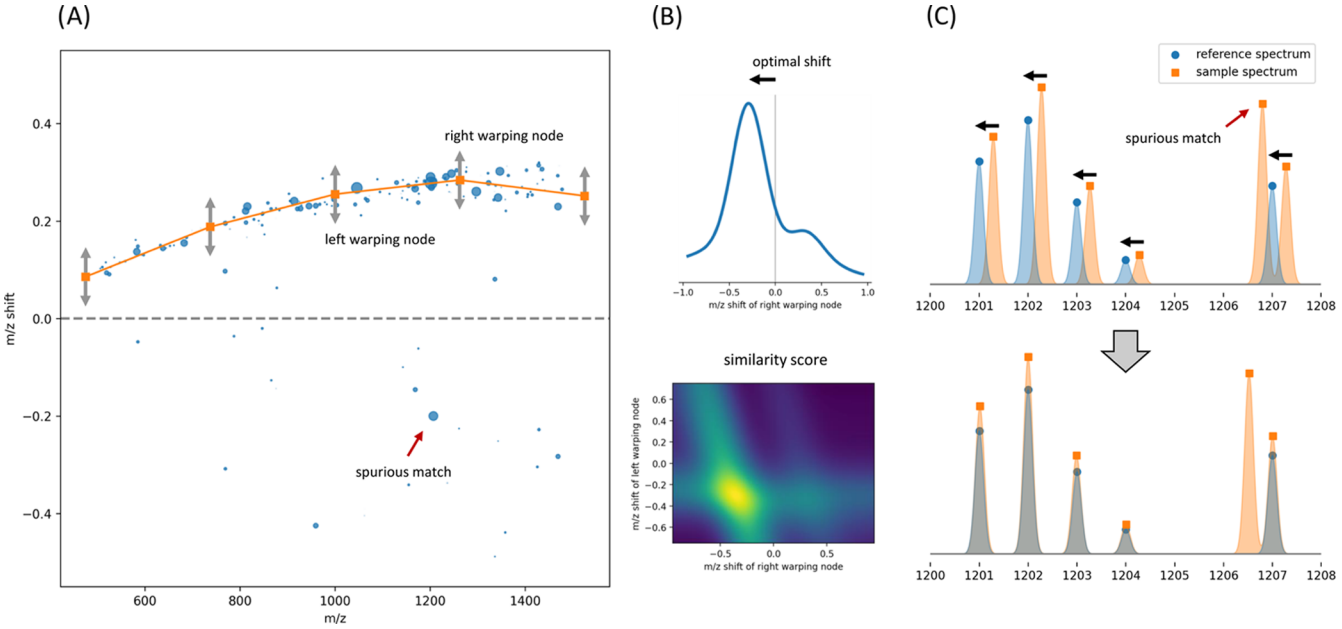


Figure 1. Conceptual description of MSIWarp. After matching the peaks in the sample spectrum against those in the reference spectrum, the sample spectrum is warped so that its similarity to the reference spectrum is maximized. (A) Scatter of the mass shift between matched peaks across the m/z range. The shifted warping nodes are marked with vertical arrows (the actual search space is centered around zero and extends beyond the arrows). The orange curve shows the estimated mass shift based on our similarity score. (B) The similarity score is evaluated for the set of candidate warpings, and the warping resulting in the highest score is used to align the sample spectrum. (C) Zoom-in of the spectra with the centroided peaks modeled as Gaussians. Two sample spectrum peaks are matched to the reference spectrum peak at 1207 m/z, but the spurious match (also marked in (A)) has no effect on the warping.

data set, or be a composite spectrum constructed from multiple spectra. Similar to our previous work,[19] MSIWarp deliberately relies on centroided, i.e., peak-picked, spectra to perform alignment. Peak-picking can improve alignment by retaining most compound-related signals while removing background noise that degrades performance. More importantly, however, an alignment method that takes centroided spectra as input can trivially be used to align profile spectra, but not vice versa. MSIWarp relies on centroid spectra instead of profile spectra and is thereby readily applicable to almost any MSI data set.

MSIWarp can be summarized in three steps (Figure 1): first, we match the peaks of the sample spectrum against those of the reference spectrum. Second, based on the peak matches from step one, we split the m/z range in a manner that ensures there is sufficient shared information in all segments. Finally, we find an optimal warping function with the peak matches and the partitioning of the m/z range obtained in steps one and two, respectively, and use this function to recalibrate the peak masses of the sample spectrum.

Mass Alignment. Our method aligns a pair of spectra by warping one in the mass dimension so that its similarity to the other is maximized. Provided that the type of mass spectrometer used to generate the spectra is known, it requires only a list of peak heights and m/z locations for each spectrum. We model peak intensity as a Gaussian function of m/z with centroid mass μ and height H. With this peak model, we can then compute the similarity of two centroided spectra as the sum of all pairwise peak overlaps, and use this similarity score as a measure of alignment quality. To model peak width, σ, we use known theoretical relationships between peak width and m/z, together with the mass resolution of the data set. The theoretical relationships depend on instrument type and are summarized in Suits et al.[21] If the mass resolution is unknown, a good estimate of a single peak's full width at half-maximum (FWHM) is enough to model the width of all other peaks in the data set. While not a true representation of a mass spectrum peak, the Gaussian peak model is a sufficient approximation for the purpose of alignment. Similarly, the modeled peak width does not have to match the true width exactly, since its main purpose is to provide some freedom when matching and aligning peaks.

Equations 1−4 formally define our Gaussian peak model p, the overlap between two peaks I, and the similarity between two spectra, B. The intensity of a peak varies with m/z according to

$$p(mz) = H \exp\left(-\frac{(mz-\mu)^2}{2\sigma^2}\right) \tag{1}$$

The overlap of two peaks is

$$I(p_i, p_j) = \int_{-\infty}^{\infty} H_i \exp\left(-\frac{(mz-\mu_i)^2}{2\sigma_i^2}\right) \times H_j \exp\left(-\frac{(mz-\mu_j)^2}{2\sigma_j^2}\right) dmz \tag{2}$$

The integral in eq 2 can be solved analytically to yield

$$I(p_i, p_j) = H_i H_j \cdot \sqrt{2\pi} \cdot \frac{\sigma_i \sigma_j}{\sqrt{\sigma_i^2 + \sigma_j^2}} \exp\left(\frac{1}{2\sigma_i^2\sigma_j^2}(\alpha-\beta)\right) \tag{3}$$

where

https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

$$\alpha = \frac{(\sigma_j^2 \mu_i + \sigma_i^2 \mu_j)^2}{\sigma_j^2 + \sigma_i^2}$$

and

$$\beta = \sigma_j^2 \mu_i^2 + \sigma_j^2 \mu_j^2$$

The similarity score, B, between two spectra, S_i and S_j, is the sum overlap of all matched peaks

$$B(S_i, S_j) = \sum_{|\mu_i - \mu_j| < \epsilon} I(p_i, p_j) \quad (4)$$

To compute the similarity score between a pair of spectra, the set of peak pairs that satisfy the condition in eq 4 must be found. A peak in the first spectrum is matched to one in the second spectrum if their m/z locations are within a small distance threshold, ϵ, of each other. Note that a peak in one spectrum can be matched to multiple peaks in the other spectrum. The threshold ϵ in eq 4 is proportional to the modeled peak width and therefore increases with m/z.

With our definition of pairwise similarity between two centroided spectra, we can search for an optimal warping from a set of candidate warpings in a similar way as the original COW implementation. This involves dividing the m/z range into segments, evaluating B for all candidate warpings for each segment independently, and then finding the optimal combination of segment warpings. The analytical form of the integral in eq 2 enables fast computation of the similarity score, which is critical since it must be repeated for each candidate warping. Here, we denote the lower and upper edges of an m/z segment n_l and n_r, respectively, and refer to them as the warping nodes of the segment. Note that each warping node, except those at the lowest and highest ends of the m/z range, are shared by two segments. To generate the set of candidate warpings for an m/z segment, the warping nodes at its edges are shifted a fixed number of steps upward and downward in m/z. The set of warpings for that segment then corresponds to all possible combinations of shifts of n_l and n_r. Computing B for a particular warping and segment is then performed by warping the peaks in the segment and then computing B with the warped peaks and the peaks from the corresponding segment in the reference spectrum. Peaks are warped by updating their mass with linear interpolation according to

$$\mu' = x \cdot (n_r' - n_l') + n_l' \quad (5)$$

where

$$x = (\mu - n_l)/(n_r - n_l)$$

μ is the original peak mass, μ' is the warped peak mass, and n_l' and n_r' are the shifted positions of n_l and n_r, respectively. Note that while segments are compressed, stretched, and/or shifted, a peak is never warped out of its segment and its width and height are left untouched. Finally, like in the original COW implementation,[17] the optimal combination of warping node moves is found with dynamic programming. The shift of a warping node is bounded by the slack parameter: $|n' - n| \leq m_n$. The slack is reflected by the amplitude of the warping function and should be sufficiently large to capture the largest shifts. Like the peak matching threshold (ϵ in eq 4), m is proportional to FWHM and is computed for each warping node individually.

By maximizing our similarity score, we find a piecewise linear mass recalibration function with a degree of freedom determined by the number of warping nodes. A large number of short segments gives a flexible warping function that can correct local m/z shifts, whereas a small number of long segments generally results in a more stable, but less flexible, warping function. The risk of overfitting the warping function to noise or random variations in peak mass is smaller with segments that have many matched peaks. Due to this, we prefer long segments that accommodate a sufficient number of peak matches (at least 10−20) to short segments that are potentially more flexible.

Placement of Warping Nodes. In many MSI data sets, there is a large variation in peak density across the m/z range. For such data sets, the placement of the warping nodes can greatly influence the warping quality. The same warping nodes can be used for all spectra, or they can be placed uniquely for each spectrum. The goal of the warping node placement is to have a segment length that is adapted to the amount of shared information, i.e., peak matches, between the sample and reference spectra in all parts of the m/z range. This can be achieved by generating a density estimate (smooth histogram) of the peak matches between the sample and reference spectra over the m/z range and then placing the warping nodes between the peaks of the density curve. If the warping nodes are uniquely placed for each spectrum, they can alternatively be placed so that the number of peak matches is the same in all segments. A third option is to use segments with uniform lengths. This may work well for spectra with a high peak density throughout the m/z range but can result in segments without peak matches for peak-sparse spectra.

RANSAC Outlier Detection. Unlike alignment methods that rely solely on the difference in mass between matched peaks, MSIWarp is naturally robust to spurious matches, as long as there are sufficiently many true matches. To confirm this, we use a custom RANSAC procedure to detect spurious matches, perform alignment both with the full set of peak matches and with that obtained after having removed spurious matches, and compare the results to evaluate whether spurious matches degrade the alignment quality of MSIWarp in practice. Generally, RANSAC fits a model to a minimal subset of data points that may contain outliers. The subset is resampled numerous times and the model is fit to each subset. The best model, given some criteria, is then selected and all data points that fit the model are included in the "inlier" set. We combine RANSAC with our method to separate true matches (inliers) from spurious matches in the following way:

(i) Generate a list of preliminary peak matches with a permissive distance threshold proportional to peak FWHM.
(ii) Randomly sample two matches from the preliminary set of matches for each segment, and fit a trial warping model to the sampled matches. Warp all other preliminary matches according to the trial model.
(iii) Add peak matches whose mass distance after alignment is below a strict threshold to the inlier set.
(iv) Repeat steps (i)−(iii) n times and return the largest set of inliers. Given an estimate of the fraction of inliers among the peak matches, n can be set to obtain a desired probability of an outlier-free candidate model.

We use ϵ from eq 4 as the threshold in (i) and 0.3 times peak FWHM as that in (iii). When using a large number of

https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

segments, the warping function found using the subset of peak matches in (ii) is highly unstable and can sometimes fit a large number of spurious matches by chance. Therefore, we use only one or two warping segments in the RANSAC step. After removing the spurious matches, more warping segments can be added in parts of the m/z range that are supported by the number of true matches. The final warping is then searched for using all, or a large fraction, of the true matches.

MATERIALS AND METHODS

Data Sets. To evaluate MSIWarp, we applied it to four publicly available data sets that together represent the most common MSI experimental setups. The first data set was generated from two mouse kidney sections with a rapifleX MALDI TOF/TOF instrument (Bruker Daltonics).[22] The second data set was generated from human cancer spheroids with an ultrafleXtreme MALDI-TOF/TOF instrument (Bruker Daltonics).[23] The third data set was generated from a rat liver section with a MALDI LTQ Orbitrap XL instrument (Thermo Fischer Scientific).[24] The fourth data set was generated from a colorectal adenocarcinoma sample using a home-built motorized DESI ion source and an LTQ XL Orbitrap Discovery instrument (Thermo Fischer Scientific).[25] The data sets, referred to as the TOF kidney, TOF spheroids, Orbitrap liver, and Orbitrap DESI data sets, are summarized in Table 1. A more detailed description of the data sets is available in the Supporting Information, and total ion current (TIC) images for each data set are shown in Figure S1. Before performing mass alignment, we filtered out peaks whose intensity was below a signal-to-noise ratio (SNR) of 2.5 from the mouse kidney TOF data set and centroided all data sets, except for the Orbitrap DESI, with the parabolic centroiding algorithm by Robichaud et al.[26] We downloaded the Orbitrap DESI data set in centroid mode from the MetaboLights repository. The data sets were preprocessed with in-house Python scripts.

Table 1. Summary of the Data Sets[a]

	TOF kidney	TOF spheroids	Orbitrap liver	Orbitrap DESI
ionization technique	MALDI	MALDI	MALDI	DESI
species	mouse	human	rat	human
no. spectra	33 242	1114	23 823	20 286
raster size (μm)	100	50	50	100
m/z range (Da)	500−2500	800−4500	150−1000	200−1000
resolution	2600	17 500	60 000	60 000
avg. no. peaks	244	57	730	701

[a]Both TOF data sets were uploaded to ProteomeXchange smoothed and with their baseline removed. Resolutions for the Orbitrap data sets were calculated at 400 m/z.

Data Analysis. To measure the effect of alignment in each data set, we calculated the mass dispersion around a set of reference peaks. We obtained the mass dispersion of a reference peak by binning all spectra around its m/z and then calculating the standard deviation of peak masses within the resulting mass bin. By binning we mean isolating peaks across spectra at predefined m/z locations, and we refer to the isolation windows as mass bins. We used a bin width two times the FWHM of the reference peak. As reference peaks, we used the most intense peaks of the mean spectrum (100 for the TOF kidney, Orbitrap liver, and Orbitrap DESI data sets and 50 for the TOF spheroid data set), some matrix peaks, and peaks that were identified in the papers that originally published the data sets. The mean spectrum was generated after alignment with MSIWarp, and we performed the binning and calculated mass dispersion both before and after alignment.

To further assess the quality of the alignment, we generated scatter plots of peak mass and spectrum acquisition time for the mass bin of each reference peak (Figures S3−S13). The scatter plots provide a clear view of the mass shift before and after alignment and serve as valuable quality control for the alignment of specific peaks. Despite the previous intensity filtering of the TOF kidney data set based on SNR, some mass bins were still contaminated with faint background peaks. To reduce the influence of these, we applied an intensity threshold to each mass bin. The threshold was defined as the lower intensity quartile of all of the peaks in the mass bin, and peaks whose intensity was below this threshold were excluded when calculating mass dispersion.

RESULTS AND DISCUSSION

To reiterate, MSIWarp aligns a data set by maximizing the similarity between each spectrum and a common reference spectrum. Like any method that performs pairwise alignment, it relies on shared information. In the ideal case, all spectra share numerous peaks with the reference spectrum throughout the m/z range. In a more challenging case, there are few shared peaks overall and/or wide gaps in the m/z range without any shared peaks. Figure 2 shows a pair of spectra from the Orbitrap data set, another from the TOF kidney data set, and the m/z difference between preliminary matched peaks. The Orbitrap spectra are homogeneous, with shared peaks throughout the m/z range, and the mass shifts are small (<1.5 ppm). In contrast, the TOF spectra are heterogeneous, the mass shifts are significantly larger (>200 ppm), and there is a part of the m/z range with almost no shared peaks (1000−1400 m/z). Note that the mass differences between shared peaks for a pair of aligned spectra are expected to be distributed around zero throughout the m/z range. However, in these examples, the misalignment is apparent; the mass shift of matched peaks consistently increases with m/z for both pairs. To accommodate large mass shifts, we used a peak matching threshold (ϵ in eq 4) of approximately two times the FWHM when matching peaks. With the low mass resolution in the TOF kidney data set, this meant matching peaks within a window of ±0.76 m/z at 1000 m/z. A wide matching window increases the number of spurious matches; the scatter in Figure 2b contains numerous examples, most notably those clustered around 850 and 1050 m/z. Finally, we chose to place the warping nodes based on the density estimate of peak matches, since it generally resulted in a smoother partitioning of the m/z range than when placing them so that all segments contained the same number of peak matches. We generated the density estimate by performing a Kernel density estimation (KDE) of peak m/z and then placed the warping nodes between the peaks of the density curve, resulting in 4−10 warping segments for the four data sets. Increasing and decreasing the bandwidth of the KDE is a flexible way to adjust the number of warping segments and the number of peak matches in each segment. We used a bandwidth of 15 Da for the TOF kidney and Orbitrap data sets, and a bandwidth of 100 Da for the TOF

https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

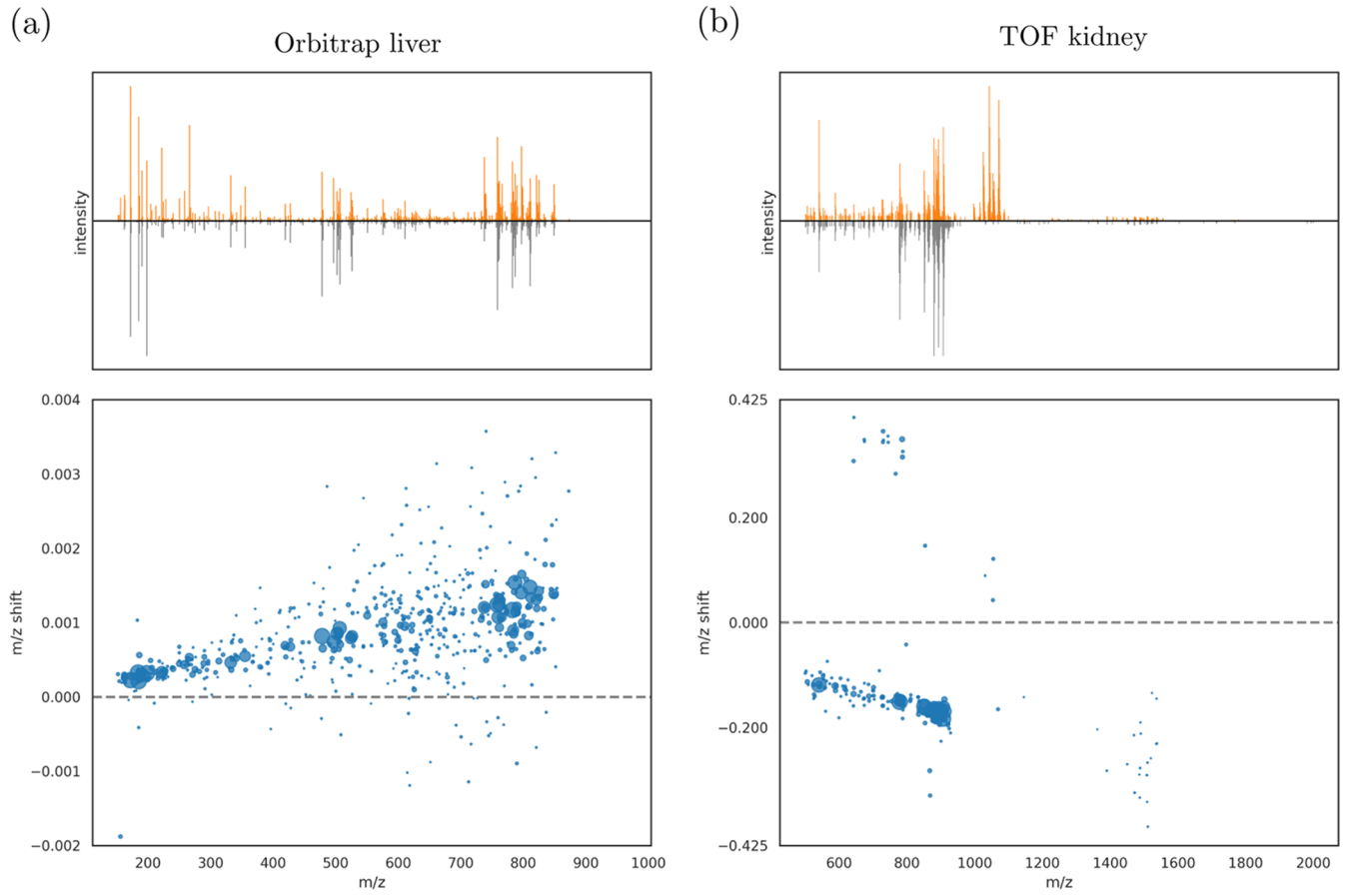


Figure 2. Top: pair of raw spectra from the Orbitrap liver data set (a) and another from the TOF kidney (b). Bottom: scatters of m/z difference between matched peaks with point size scaled by intensity. The Orbitrap spectra share peaks across the entire m/z range. In contrast, the TOF spectra share no peaks in a large part of the m/z range (1000−1400 m/z). This example also highlights the severity of the mass shift in the TOF kidney data set: at 800 m/z, the shift is close to 0.18 m/z (219 ppm) between the TOF pair compared to 0.001 m/z (1.25 ppm) between the Orbitrap pair.

spheroid data set since it was significantly less peak-dense than the other data sets.

Spurious peak matches are largely inconsequential to MSIWarp as long as spectra are reasonably peak-dense; in fact, most spectra are aligned accurately and reliably without RANSAC, even those in the TOF data sets. We hypothesized that the importance of identifying true matches increases when aligning peak-sparse spectra. For this reason, we aligned the TOF data sets both with and without RANSAC. When combining RANSAC with COW, it is important to use a small number of warping segments in this step to avoid overfitting the candidate models to spurious peak matches; here, we used two segments for both TOF data sets in the RANSAC step. Manual inspection of scatter plots like those in Figure 3 suggests that RANSAC confidently filters out spurious peak matches in almost all spectra across both TOF data sets. We used RANSAC to filter out spurious peak matches before searching for the optimal warping. Then, we added more warping nodes in the peak-dense parts of the m/z range to gain more flexibility. Thereby, we obtained a flexibility in the warping function that was adapted to the number of shared peaks throughout the m/z range. We provide some examples of how RANSAC finds the true peak matches (Figure S2) along with an animation in the Supporting Information.

When aligning a data set by aligning each spectrum to a common reference spectrum, the quality of that reference spectrum is essential. We tried an approach similar to that of Kulkarni et al.,[11] where the reference spectrum is continuously updated with each aligned sample spectrum, but observed no significant improvement over aligning against a constant reference spectrum of high quality. The spectrum with the highest TIC was sufficient in terms of peak coverage for all four data sets, and we therefore chose to use it as a reference. Although the spectra with the highest TIC were appropriate references for the alignment of the data sets that we discuss here, a composite spectrum may be needed to fully cover the m/z range in other data sets.

How we dealt with segments without shared peaks also deserves mention. We chose to interpolate the warping function in empty regions, as is evident in Figure 3b at around 1200 m/z. This is reasonable under the assumption that some relevant data set peaks are missing in the reference spectrum, which is often the case, so they can be present in a sample spectrum without being shared with the reference. An alternative approach is to leave the empty parts of the m/z region unaligned, which can be more appropriate if the reference spectrum has a very high peak coverage throughout the m/z range. When this is the case, the sample spectrum likely has little or no information in regions where it does not share peaks with the reference spectrum, and aligning those regions is unnecessary.

https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

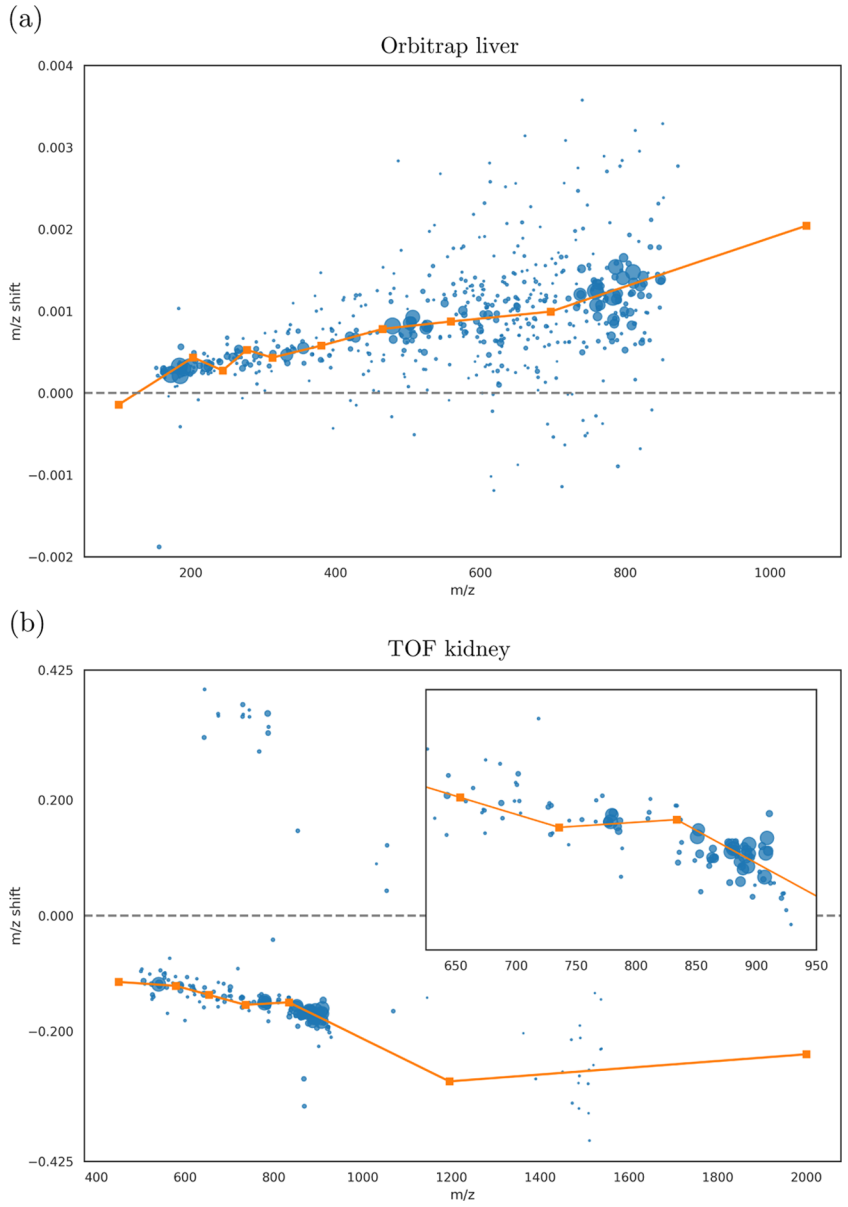


Figure 3. (a, b) Mass shift estimated by MSIWarp (orange line) overlaid on the peak match scatter from Figure 2. (b) We use more warping nodes in the peak-dense part of the m/z range than in the peak-sparse part; the zoom-in shows that the warping function closely follows the local shifts between 700 and 900 m/z.

Reduction in Mass Misalignment. After alignment with MSIWarp, mass dispersion is reduced considerably in all four data sets. In Table 2, the mass dispersions of the mean spectrum peaks before and after alignment are reported in both ppm and FWHM. The full list of dispersions of the mean spectrum peaks is available in the Supporting Information spreadsheet. Interestingly, we observed no significant improvement when aligning the TOF data sets with RANSAC compared to aligning it without RANSAC. This suggests that the inherent robustness of COW is sufficient in dealing with spurious peak matches for the majority of spectra. To assess the sensitivity of MSIWarp to the modeled peak width, σ, and the peak matching threshold, ϵ, we reran the analysis of the Orbitrap liver and TOF kidney data sets for various values of these parameters (Table S1 and S2). This parameter sensitivity analysis suggests that MSIWarp can perform well even when σ over- or underestimates the experimental peak width by a factor of up to 2. It also suggests that a large peak matching threshold is better than a small one, which is further evidence that MSIWarp is robust to spurious peak matches and that the most important factor is that true matches are captured

Table 2. Median Mass Dispersion of Mean Spectrum Peaks before (Raw) and after Alignment (Warped) for the Four Data Sets

	TOF kidney	TOF spheroids	Orbitrap liver	Orbitrap DESI
raw (ppm)	106.39	35.63	0.63	0.52
warped (ppm)	12.73	6.48	0.18	0.17
raw (FWHM)	0.27	0.63	0.03	0.03
warped (FWHM)	0.03	0.11	0.01	0.01
reduction (%)	88.03	81.82	72.03	68.00

https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

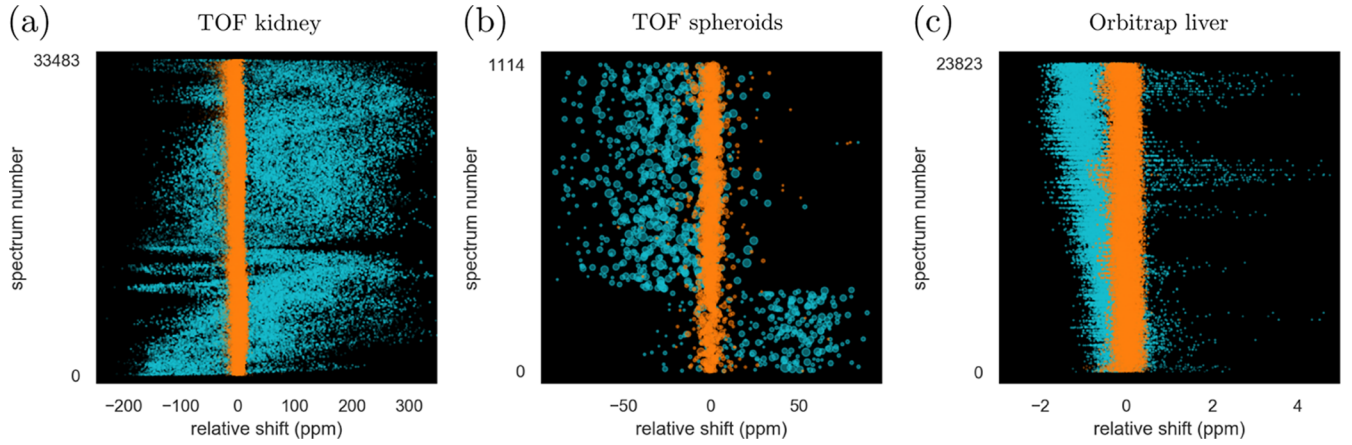


Figure 4. Scatter plots of mass shift relative to the reference peak (y-axis) and spectrum index ordered according to acquisition time (y-axis) before (cyan) and after alignment (orange). (a−c) Scatters around reference peaks at m/z 850.80 (PI 36:7), 1403.10 (unknown), and 172.04 (matrix) in the TOF kidney, TOF spheroids, and Orbitrap liver data sets, respectively.

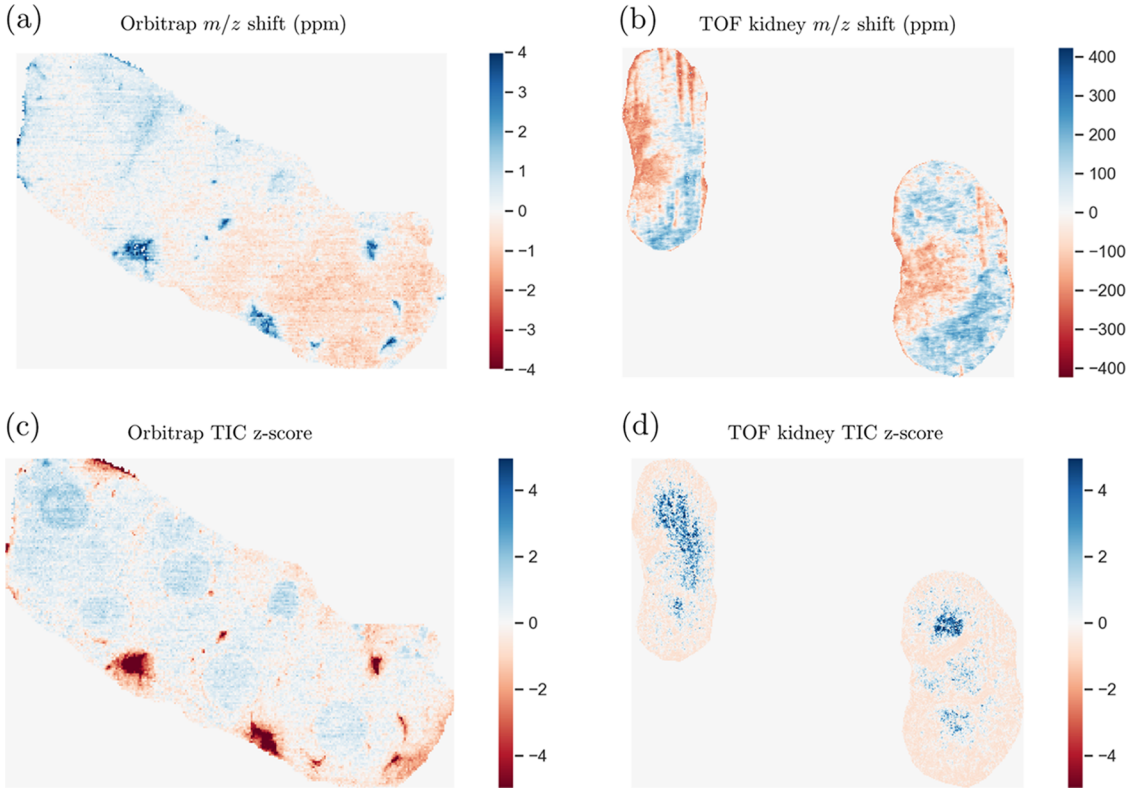


Figure 5. Mass shift and TIC images from the Orbitrap liver and TOF kidney data sets. Left: the mass shift (a) of the matrix peak at 172.04 m/z in the Orbitrap data set is correlated to TIC (c). Right: the mass shift (b) of the lipid peak (PI 38:2) in the TOF kidney data set appears to be related to tissue structures rather than to TIC (d) or acquisition time.

throughout the m/z range. In addition to using the spectrum with the highest TIC as a reference, we also aligned the TOF kidney data set to each of the 11 spectra with TICs closest to the data set median, resulting in median mass dispersions of mean spectrum peaks between 11.64 and 16.10 ppm (avg. 12.53). For 9 out of 11 spectra, the median mass dispersion was lower than when using the spectrum with the highest TIC as a reference (12.73 ppm). In comparison, using the spectrum with the lowest TIC as a reference resulted in a median mass dispersion of 73.45 ppm. Altogether, this indicates that a minimum TIC threshold can be used to filter out unsuitable reference spectra, but otherwise, the TIC provides little or no information about a spectrum's suitability as a reference.

Visualization of how the peak mass varies across the experiment gives a good overview of alignment; Figure 4a−4c shows the scatter of the relative peak mass before and after alignment with MSIWarp and spectrum acquisition time for three example peaks. The same pattern is seen for the three peaks: aligning with MSIWarp visibly tightens the peak scatter and removes the systematic shifts related to spectrum acquisition time. These scatter plots are effective in visualizing time-dependent mass shift but do not provide a clear picture of how mass shift relates to tissue location. To better visualize this

https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

relationship, we show the relative mass shift of a matrix peak ($[M - H_2O + H]^+$, 172.04 m/z) from the Orbitrap data set as a function of tissue location in Figure 5a. In this plot, it is evident that the peak masses decrease with time (the tissue was scanned from top to bottom), but also that some shifts are related to tissue location. The blue spots in the bottom half of the section break the trend related to acquisition time; in these spots, peak masses are increased by more than 4 ppm, compared to an average decrease of approximately 1 ppm in the bottom half of the section. The blue spots appear to be correlated to the TIC of the spectra, which could be due to the space-charge effect. Figure 5b shows the mass shift image of the peak at 890.8 m/z (PI 38:2) from the TOF kidney data set. In contrast to the matrix peak from the Orbitrap data set, the mass shift of this peak appears to be unrelated to both spectrum acquisition time and TIC (Figures 4a and 5b). Instead, there is a strong relationship to tissue location: in both kidney sections, peak masses appear to be increased in the cortex and decreased in the medulla. The "stripes" at the top part of the left section are likely an experimental artifact (due to tissue folding or damage) rather than being related to any tissue structure.

The mass dispersion of some identified compounds and matrix peaks in the Orbitrap liver and TOF kidney data sets are shown in Tables 3 and 4, respectively. The monoisotopic peak

Table 3. Mass Dispersion (ppm) of Five Matrix (α-Cyano-4-hydroxycinnamic Acid) Peaks, the Monoisotopic Peak of Two Spiked-In Compounds, and the Peak of Phosphocholine before (Raw) and after Alignment (Warped) in the Orbitrap Liver Data Set

compound	m/z	disp. raw (ppm)	disp. warped (ppm)
$[M - H_2O + H]^+$	172.039	0.644	0.159
$[M + H]^+$	190.050	0.743	0.086
$[M + Na]^+$	212.032	0.724	0.222
$[M - H + 2Na]^+$	234.014	0.998	0.435
$[2M + H]^+$	379.092	0.699	0.343
phosphocholine	184.073	0.706	0.036
ipratropium	332.222	0.675	0.271
dasatinib	488.164	0.475	0.364

Table 4. Mass Dispersion (ppm) of Some Identified Lipid Peaks before (Raw) and after Alignment (Warped) in the TOF Kidney Data Set[a]

compound	m/z	disp. raw (ppm)	disp. warped (ppm)
LPS 18:0	524.19	126.83	15.24
PA 34:1	673.68	128.19	24.99
PA 36:1	701.76	123.15	16.80
PS 36:3	784.38	118.82	16.13
PI 36:7	850.80	103.38	5.70
PS 42:5	864.82	105.18	7.23
PI 40:6	907.87	103.49	6.51
PI 40:1	919.90	87.83	32.69

[a]All lipids are represented as $[M - H]^-$ ions.

of the phosphatidylcholine head group (184.07 m/z) in the Orbitrap data set has a notably lower dispersion after alignment (0.036 ppm) than those of the other compounds (0.086−0.435 ppm). The intensity of this peak is several orders of magnitude larger than that of all other peaks. As a consequence, it dominates the similarity score and effectively acts as a lock mass for the warping function. Importantly, this does not appear to increase dispersion for lower intensity peaks close in m/z; the matrix peaks at 172.04 and 190.05 m/z contribute insignificantly to the similarity score in comparison, but still have lower dispersion than most other peaks after alignment. This suggests that the shift in low-intensity peaks can be estimated accurately with the shift of nearby high-intensity peaks. In the TOF kidney data set, dispersion is reduced consistently by more than 80 percent, except for the peak at m/z 919.90 (PI 40:1), whose dispersion remains relatively high after alignment (32.69 ppm). By looking at the mass scatter of this peak, however, it is evident that it has been mixed with another peak in the same mass bin and that this causes the relatively high dispersion after alignment (Figure S5, Supporting Information).

An interesting example of the utility of MSIWarp is shown in Figure 6. Like the mass bin at 919.90 m/z, other mass bins in the TOF kidney data set contain mixed peaks as well, including that of the unidentified reference peak at 904.90 m/z. Before alignment, the two peaks in this bin are indistinguishable, but after alignment, the peaks are separated sufficiently to enable us to generate a distinct ion image for each. The distribution of peak mass within the bin is considerably narrower after alignment (Figure 6b). Crucially, the density curve of the warped peak masses reveals, albeit just barely, the second peak as distinct from the first one. By re-binning the spectra with two tighter windows (±10 ppm), whose respective centers are at the main and shoulder peaks of the density estimate, two distinct ion images can be generated (Figure 6d,6e) instead of one in which the two compounds are mixed (Figure 6c). Despite the low mass resolution of the TOF kidney data (the FWHM is approximately 380 ppm), these two peaks, which are 25 ppm apart, can still be separated by looking at the mass locations of their centroids. Note that while we separate them on a data set level, we do not separate them in a single spectrum; when the compounds corresponding to the peaks are present in the same spectrum, i.e., tissue spot, the peak of the more intensive compound masks that of the other. This is evident in the ion images of the two peaks (Figure 6d,6e): only the more intense peak (904.878 m/z) is visible in the pixels where the two peaks overlap.

Implementation and Processing Time. Although our method involves repeating the similarity score computation for a large number of candidate warpings to align a single spectrum, we keep the processing time for a whole data set low by implementing the core part of MSIWarp in efficient C++. Aligning the TOF kidney data set, consisting of 33 242 spectra with 244 peaks on average, took approximately 150 s when MSIWarp was run without RANSAC and in parallel mode on a laptop with an Intel i7-6700HQ CPU (2.6 GHz) and 16 GB RAM. Aligning the Orbitrap liver data set with the same settings took approximately 300 s. The parameter that has the largest impact on processing time is the number of steps by which the warping nodes are shifted when searching for the optimal warping. Given that the slack has been set to capture the mass shift for all spectra, the number of steps corresponds to the resolution of the warping function. We found that a step size of 0.05−0.10 times the peak FWHM gives a sufficient alignment resolution. If the search space of the warping function is ±2 FWHM, this results in 40−80 steps for each warping node. The source code of MSIWarp along with Python bindings are available at https://github.com/horvatovichlab/MSIWarp. Our goal is to make it possible to

https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

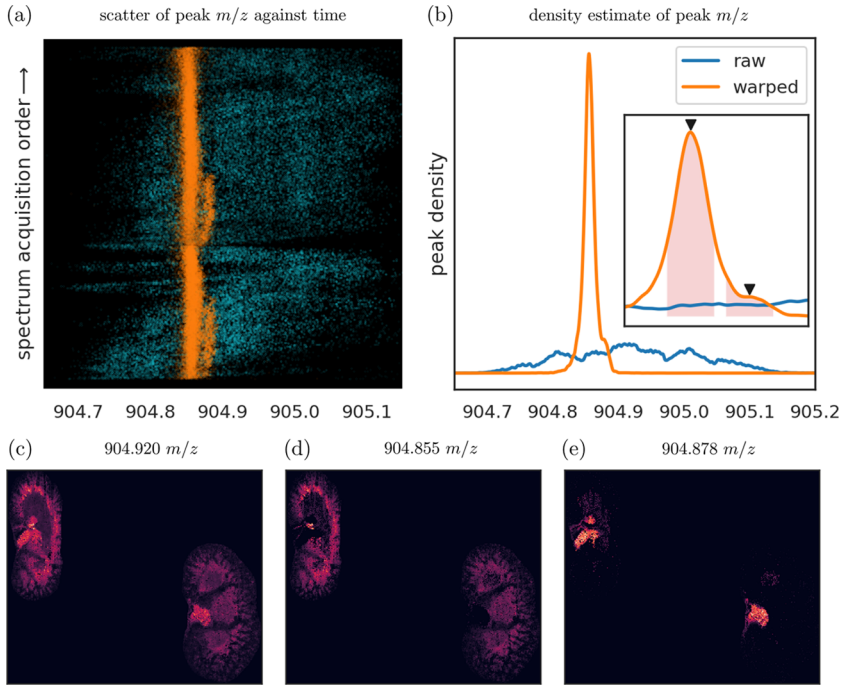


Figure 6. Mass bin from the TOF kidney data set exemplifies how severe misalignment leads to mixed ion images. The scatter of peak mass and acquisition time in (a) reveals two peaks after alignment (orange) that are indistinguishable before alignment (cyan). Alignment similarly reveals the two peaks in the density estimate of peak mass within the bin (b): before alignment (raw), the variation in peak mass across the spectra is too large to separate the two peaks, but after alignment (warped), the peaks appear as the main and shoulder peaks of the density curve. The left ion image (c) was generated from the raw (unaligned) spectra with a wide bin window (±200 ppm), while the center (d) and right (e) ion images were generated by binning the warped spectra with a narrow window (±10 ppm) around 904.855 and 904.878 m/z, respectively. The median mass of the two peaks and the narrow windows are marked in the zoom-in on the density curve in (b). The ion images in (c)−(e) were generated by summing peak intensities in the bin for each spectrum/pixel.

interface MSIWarp with existing MSI analysis packages such as MALDIquant,[27] Cardinal,[28] and rMSIproc.[29]

CONCLUSIONS

We have presented an approach, MSIWarp, that readily improves relative alignment in both TOF and Orbitrap data sets that together represent a large variety of MSI experimental setups. Even the severe misalignment in the TOF kidney data set is brought down to a level that enables separation of peaks close in m/z. With a median mass dispersion of 6.48 ppm (and below 5 ppm for more than half of the peaks) in the TOF spheroid data set, MSIWarp matches the alignment performance of methods that rely on profile spectra using only centroided spectra. While the largest improvements in relative mass alignment can be gained for TOF spectra, our results suggest that MSIWarp can further improve the already high mass alignment in Orbitrap data sets. By investigating the effect of spurious peak matches with RANSAC, we have also shown that MSIWarp is robust and performs well even for most peak-sparse spectra. Finally, we believe that a careful assessment of mass alignment is critical when analyzing MSI data sets. Tools such as scatter plots of peak mass and acquisition time, scatter plots of mass shift between individual pairs of spectra, and images of mass shift as a function of tissue location for individual peaks provide a way to do this in a simple manner.

Although MSIWarp has demonstrated significant benefits in analyzing MSI data sets, there are several research paths that could yield additional improvements. Currently, the output of MSIWarp is an m/z recalibration function for each data set spectrum. It is important to highlight that even though this function is found by searching for the optimal alignment between a pair of centroided spectra, it can also be used to align the corresponding profile spectra. Conceptually, the recalibration function could also be found by directly computing the correlation integral between the profile spectra, if available, instead of using our analytical expression based on the overlap of centroided peaks. This would allow MSIWarp to utilize subtle features in the profile data, such as peak shape, that may be lost in the peak-picking step. Another possible enhancement is to create a hybrid approach by combining MSIWarp with existing calibration strategies to improve absolute mass accuracy in addition to relative mass alignment. To do this, a spectrum with accurate masses should be used as a reference, and if no such spectrum is available, the reference spectrum can be chosen based on other criteria and calibrated prior to alignment. Together, these approaches represent a rich area of research that would allow interesting comparisons of related techniques and potential for even further improvements in performance.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c03833.

Description of the TOF kidney, TOF spheroids, Orbitrap liver, and Orbitrap DESI data sets. Tables S1 and S2, median mass dispersions of mean spectrum peaks from the TOF kidney and Orbitrap liver data sets after alignment with various values of σ and ϵ. Figure S1,

https://dx.doi.org/10.1021/acs.analchem.0c03833
Anal. Chem. 2020, 92, 16138−16148

TIC images for the four MSI data sets. Figure S2, RANSAC outlier detection results for three example spectra from the TOF kidney data set. Figures S3−S13, scatter plots of peak mass and acquisition time for mean spectrum peaks from the top 100, 50, 100, and 100 mean spectrum peaks' TOF kidney, TOF spheroids, Orbitrap liver and Orbitrap DESI data sets (PDF)
Ransac outlier detection animation (ZIP)
Mean spectrum dispersions (XLSX)

AUTHOR INFORMATION

Corresponding Author

Peter Horvatovich − *Department of Biomedical Engineering, Lund University, Lund 221 00, Sweden; Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, University of Groningen, 9713 AV Groningen, The Netherlands;* orcid.org/0000-0003-2218-1140; Email: p.l.horvatovich@rug.nl

Authors

Jonatan O. Eriksson − *Department of Biomedical Engineering, Lund University, Lund 221 00, Sweden*

Alejandro Sánchez Brotons − *Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, University of Groningen, 9713 AV Groningen, The Netherlands*

Melinda Rezeli − *Department of Biomedical Engineering, Lund University, Lund 221 00, Sweden;* orcid.org/0000-0003-4373-5616

Frank Suits − *IBM Research - Australia, Southbank, VIC 3006, Australia*

György Markó-Varga − *Department of Biomedical Engineering, Lund University, Lund 221 00, Sweden*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c03833

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the support from Fru Berta Kamprads Stiftelse. This research was part of the Netherlands X-omics Initiative and partially funded by NWO, project 184.034.019. We also thank the reviewers for their valuable feedback.

REFERENCES

(1) Buchberger, A. R.; DeLaney, K.; Johnson, J.; Li, L. *Anal. Chem.* **2018**, *90*, 240−265.
(2) Gilmore, I. S.; Heiles, S.; Pieterse, C. L. *Annu. Rev. Anal. Chem.* **2019**, *12*, 201−224.
(3) Han, J.; Permentier, H.; Bischoff, R.; Groothuis, G.; Casini, A.; Horvatovich, P. *TrAC, Trends Anal. Chem.* **2019**, *112*, 13−28.
(4) Ryan, D. J.; Spraggins, J. M.; Caprioli, R. M. *Curr. Opin. Chem. Biol.* **2019**, *48*, 64−72.
(5) Verbeeck, N.; Caprioli, R. M.; Van de Plas, R. *Mass Spectrom. Rev.* **2020**, *39*, 245−291.
(6) Ràfols, P.; del Castillo, E.; Yanes, O.; Brezmes, J.; Correig, X. *Anal. Chim. Acta* **2018**, *1022*, 61−69.
(7) Gorshkov, M. V.; Good, D. M.; Lyutvinskiy, Y.; Yang, H.; Zubarev, R. A. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1846−1851.
(8) Kharchenko, A.; Vladimirov, G.; Heeren, R. M.; Nikolaev, E. N. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 977−987.
(9) Boskamp, T.; Lachmund, D.; Casadonte, R.; Hauberg-Lotte, L.; Kobarg, J. H.; Kriegsmann, J.; Maass, P. *Anal. Chem.* **2020**, 1301.
(10) Bocker, S.; Makinen, V. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2008**, *5*, 91−100.
(11) Kulkarni, P.; Kaftan, F.; Kynast, P.; Svatoš, A.; Böcker, S. *Anal. Bioanal. Chem.* **2015**, *407*, 7603−7613.
(12) Tracy, M. B.; Chen, H.; Weaver, D. M.; Malyarenko, D. I.; Sasinowski, M.; Cazares, L. H.; Drake, R. R.; Semmes, O. J.; Tracy, E. R.; Cooke, W. E. *Proteomics* **2008**, *8*, 1530−1538.
(13) Barry, J. A.; Robichaud, G.; Muddiman, D. C. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1137−1145.
(14) Alexandrov, T. *Annu. Rev. Biomed. Data Sci.* **2020**, *3*, 61.
(15) Haug, K.; Salek, R. M.; Conesa, P.; Hastings, J.; De Matos, P.; Rijnbeek, M.; Mahendraker, T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; et al. *Nucleic Acids Res.* **2013**, *41*, D781−D786.
(16) Alexandrov, T.; Ovchnnikova, K.; Palmer, A.; Kovalev, V.; Tarasov, A.; Stuart, L.; Nigmetzianov, R.; Fay, D.; Contributors, K. M. *BioRxiv* **2019**, No. 539478.
(17) Nielsen, N.-P. V.; Carstensen, J. M.; Smedsgaard, J. *J. Chromatogr. A* **1998**, *805*, 17−35.
(18) Christin, C.; Smilde, A. K.; Hoefsloot, H. C.; Suits, F.; Bischoff, R.; Horvatovich, P. L. *Anal. Chem.* **2008**, *80*, 7012−7021.
(19) Suits, F.; Lepre, J.; Du, P.; Bischoff, R.; Horvatovich, P. *Anal. Chem.* **2008**, *80*, 3095−3104.
(20) Fischler, M. A.; Bolles, R. C. *Commun. ACM* **1981**, *24*, 381−395.
(21) Suits, F.; Hoekman, B.; Rosenling, T.; Bischoff, R.; Horvatovich, P. *Anal. Chem.* **2011**, *83*, 7786−7794.
(22) Noh, S. A.; Kim, S.-M.; Park, S. H.; Kim, D.-J.; Lee, J. W.; Kim, Y. G.; Moon, J.-Y.; Lim, S.-J.; Lee, S.-H.; Kim, K. P. *J. Proteome Res.* **2019**, *18*, 2803−2812.
(23) Mittal, P.; Price, Z. K.; Lokman, N. A.; Ricciardelli, C.; Oehler, M. K.; Klingler-Hoffmann, M.; Hoffmann, P. *Proteomics* **2019**, *19*, No. 1900146.
(24) Eriksson, J. O.; Rezeli, M.; Hefner, M.; Marko-Varga, G.; Horvatovich, P. *Anal. Chem.* **2019**, *91*, 11888−11896.
(25) Gerbig, S.; Golf, O.; Balog, J.; Denes, J.; Baranyai, Z.; Zarand, A.; Raso, E.; Timar, J.; Takats, Z. *Anal. Bioanal. Chem.* **2012**, *403*, 2315−2325.
(26) Robichaud, G.; Garrard, K. P.; Barry, J. A.; Muddiman, D. C. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 718−721.
(27) Gibb, S.; Strimmer, K. *Bioinformatics* **2012**, *28*, 2270−2271.
(28) Bemis, K. D.; Harry, A.; Eberlin, L. S.; Ferreira, C.; van de Ven, S. M.; Mallick, P.; Stolowitz, M.; Vitek, O. *Bioinformatics* **2015**, *31*, 2418−2420.
(29) Ràfols, P.; Heijs, B.; del Castillo, E.; Yanes, O.; McDonnell, L. A.; Brezmes, J.; Pérez-Taboada, I.; Vallejo, M.; García-Altares, M.; Correig, X. *Bioinformatics* **2020**, *36*, 3618−3619.