

## Mass Spectrometry Image Correlation: Quantifying Colocalization

Liam A. McDonnell,\* Alexandra van Remoortere, René J. M. van Zeijl, and André M. Deelder

Biomolecular Mass Spectrometry Unit, Department of Parasitology, Leiden University Medical Center,  
2333ZA Leiden, The Netherlands

Received March 21, 2008

**Abstract:** A typical imaging mass spectrometry data set can contain 100+ images, each describing the distribution of a specific biomolecule. Multivariate and hierarchical clustering techniques have been developed to investigate the correlations within a data set, and have revealed the differential patterns associated with different organs/anatomical features. These methods do not quantify the correlations between the hundreds of molecular distributions produced in an imaging mass spectrometry experiment, and are extremely difficult to apply to multiple tissue section investigations. This latter aspect includes quantifying the correlation between the results of repeat imaging mass spectrometry experiments, a crucial aspect for determining the significance of any measured changes in distribution. To date, the large chemical background and pixel-to-pixel variation in the images has limited the quantification of correlation between imaging mass spectrometry results. Here, we demonstrate how to quantify the correlations between imaging mass spectrometry images, both within a data set and between data sets.

**Keywords:** MALDI • imaging mass spectrometry • bioinformatics • correlation

### Introduction

MALDI imaging mass spectrometry combines the biochemical information provided by present-day mass spectrometry with the spatial distribution of a microscope. Parallel spatial analysis of up to hundreds of distinct biomolecules without the need of a label, combined with the ability to distinguish between isoforms, post-translational modifications and metabolites, has seen the technique identified as a powerful new discovery tool and has led to a rapid growth of interest in this field.<sup>1,2</sup> It has been used to study the spatial distribution of lipids,<sup>3</sup> peptides,<sup>4</sup> proteins,<sup>5</sup> pharmaceuticals<sup>6</sup> and their metabolites<sup>7</sup> in tissue sections and cells to study the biochemical changes (in terms of levels and distribution) associated with several diseases,<sup>2</sup> including Alzheimer's disease,<sup>8</sup> Parkinson's disease,<sup>9,10</sup> multiple forms of cancer,<sup>1,11,12</sup> muscular dystrophy,<sup>13</sup> and obesity.<sup>4</sup>

One of the principal motivations for the increasing application of spatially resolved mass spectrometry in biomolecular research, both imaging and profiling (targeted mass spectrom-

etry analysis of user-defined areas), is the ability to correlate anatomical information provided by histology with the spatially resolved biochemical information provided by the imaging mass spectrometry experiment.<sup>14</sup> Cornett et al. have used histology-directed profiling to record the biomolecular signatures of specific cell types in a breast tumor tissue section and demonstrated how these signatures can be used to distinguish the different cell types.<sup>15</sup> Expanding this technique to imaging mass spectrometry allows the spatial variation of the molecular signatures throughout the entire tissue section to be traced. This molecular histology has already been shown to be able to differentiate anatomical and pathological features, but is also able to trace variations not detectable by classic histology or other imaging methodologies.<sup>16</sup> The present interest in imaging mass spectrometry lies with its potential to reveal changes that underlie the anatomical changes associated with a pathology, as well as to trace the distributions of the pharmaceuticals (and their metabolites) developed to treat/investigate a pathology.

Advanced data analysis algorithms are now beginning to be applied to MALDI imaging mass spectrometry data sets. PCA,<sup>16</sup> discriminant analysis,<sup>16</sup> support vector machines<sup>17</sup> and hierarchical clustering<sup>18</sup> (now available in Bruker's ClinProt Tools version 2.0 software) have all been used to distinguish which molecular distributions are correlated within a single imaging mass spectrometry data set and to determine the mass spectral patterns that differentiate these regions. These tools form the basis of molecular histology, as it has been shown that the differential clustering is associated with anatomical and pathological features in the tissue section.

Advanced data analysis algorithms for data processing and classification are fundamental to all biomarker discovery investigations in order to identify *reliable* differences between control and disease samples.<sup>19,20</sup> In some regards, it is possible to compare imaging mass spectrometry analysis with a classic biomarker discovery experiment. In classic biomarker discovery, a large number of control and diseased samples are examined followed by rigorous data processing and classification for determining *reliable* biochemical differences. An imaging mass spectrometry experiment of a tissue section containing a tumor (or lesions, plaques, etc.) consists of 5–95k pixels spread over control (healthy tissue remote from, for example, a tumor), diseased and "intermediate" tissue. As each pixel has its own associated mass spectrum, processing and classification of the 5–95k spectra of one imaging mass spectrometry experiment has some similarities to a conventional biomarker discovery experiment.

There is however a crucial difference between a conventional biomarker discovery experiment and an imaging mass spec-

\* To whom correspondence should be addressed. Dr Liam McDonnell, Biomolecular Mass Spectrometry Unit, Department of Parasitology, Leiden University Medical Center Albinusdreef 2 2333ZA Leiden The Netherlands.

trometry experiment. A conventional biomarker discovery experiment analyses a large number of samples in order to ensure the statistical significance of the conclusions across a range of samples.<sup>19,20</sup> The large number of measurements within an imaging mass spectrometry analysis of a tissue section addresses the measurement repeatability but not the sample-to-sample variation. To date, it has not been possible to quantify the repeatability of an imaging mass spectrometry experiment. Instead, several groups now combine a conventional biomarker discovery experiment, utilizing multiple control and diseased samples, with an imaging experiment in order to address this issue of sample variability.<sup>5,21</sup>

In this report, we introduce a new method of analyzing imaging mass spectrometry data sets. The method is based on calculating correlation coefficients between individual mass spectrometry images. A correlation matrix describes the correlation coefficients between all images generated in an imaging mass spectrometry experiment, and can be used to quickly establish which images are correlated, anticorrelated and largely independent of each other. This is quite distinct from the data analysis tools previously used in imaging mass spectrometry, which examine the entire data set and look for global variations. For example, PCA typically highlights the variations due to anatomical features first and the features-of-interest are hidden in the later principal components described by just a small amount of the variance. It is often desired to know directly which biomolecules are correlated with each other, or with biomolecules that are known to be associated with a disease or to investigate the localized effects of a pharmaceutical. Additionally, the reports of 3 distinct classes of protein in protein complexes—cores, modules and attachments—in which proteins belonging to a common core and module have similar cellular localization, half-life and expression characteristics,<sup>22</sup> could be expected to have similar distributions on a tissue scale as well as a cellular scale.

Determination of imaging mass spectrometry correlations provides a potential tool for examining the repeatability and significance of an experiment: the distributions for specific peaks can be compared between the different data sets, which then allows any changes in distribution to be tested against the variability between samples. Here, we demonstrate how reliable correlation parameters for MALDI images can be calculated, show the effects of different data processing strategies on the correlation parameters, and quantify the correlations between imaging mass spectrometry data sets.

## Experimental Section

**Mass Spectrometry.** All experiments were performed on a Bruker Autoflex III mass spectrometer equipped with a Smart-beam laser system, and were acquired in fully automated mode using the Flex Control, Analysis, and Imaging software suite. The imaging mass spectrometry data sets were acquired using 200  $\mu\text{m}$  pixel size, 600 laser shots per pixel (50 laser shots per position of a random walk in each pixel) and a laser intensity a few percent above threshold. Flex imaging automatically aligns the imaging mass spectrometry data sets to a user-defined optical image. In this case, a high resolution optical scanner was used to acquire 1200 dpi resolution optical images of the matrix-coated sample immediately prior to the imaging experiment.

**Sample Preparation.** Experimental procedures were in accordance with the European directives (86/609/EEC) and approved by the Commission on Laboratory Animal Experiments of the University Medical Center Utrecht. Male Wistar

rats (Cr:WU) weighing 350 g were obtained from Charles River. Rats were decapitated without prior anesthesia, and the brains were dissected and frozen in liquid isopentane, cooled to  $-50^\circ\text{C}$  on dry ice, and then stored at  $-80^\circ\text{C}$  until sectioning. The 12  $\mu\text{m}$  thick rat brain tissue sections were cut at interaural 7.2/bregma -1.8 mm using a cryomicrotome. Sections were thaw-mounted on indium-tin-oxide-coated glass slides and were stored at  $-80^\circ\text{C}$  until use. Prior to imaging mass spectrometry analysis, the tissue sections were slowly brought to room temperature in a desiccator. The samples were then washed in ice-cold 70% ethanol ( $2 \times 30$  s) and dried under a stream of nitrogen, and the MALDI matrix was applied. A Bruker ImagePrep was used to coat the tissue sections with matrix. A solution of 20 mg mL<sup>-1</sup> sinapinic acid in 1:1 AcN/0.1% trifluoroacetic acid was aerosolized such that the resulting mist of small droplets of matrix solution uniformly coated the tissue sections. The samples were then dried in a desiccator for 30 min. Immediately prior to loading the sample in the mass spectrometer, a 1200 dpi scan was recorded.

**Data Analysis.** Scheme 1 shows a schematic of the data acquisition and data analysis reported here. Following data acquisition, a data cube of position-correlated mass spectra is obtained. This data cube is then subject to an array of mass spectral, image processing and correlation calculations. Mass spectral processing of each pixel's mass spectrum includes both smoothing and baseline subtraction steps, and were performed with FlexAnalysis using a predefined script and parameter set developed specifically for imaging mass spectrometry data sets (details in Scheme 1).

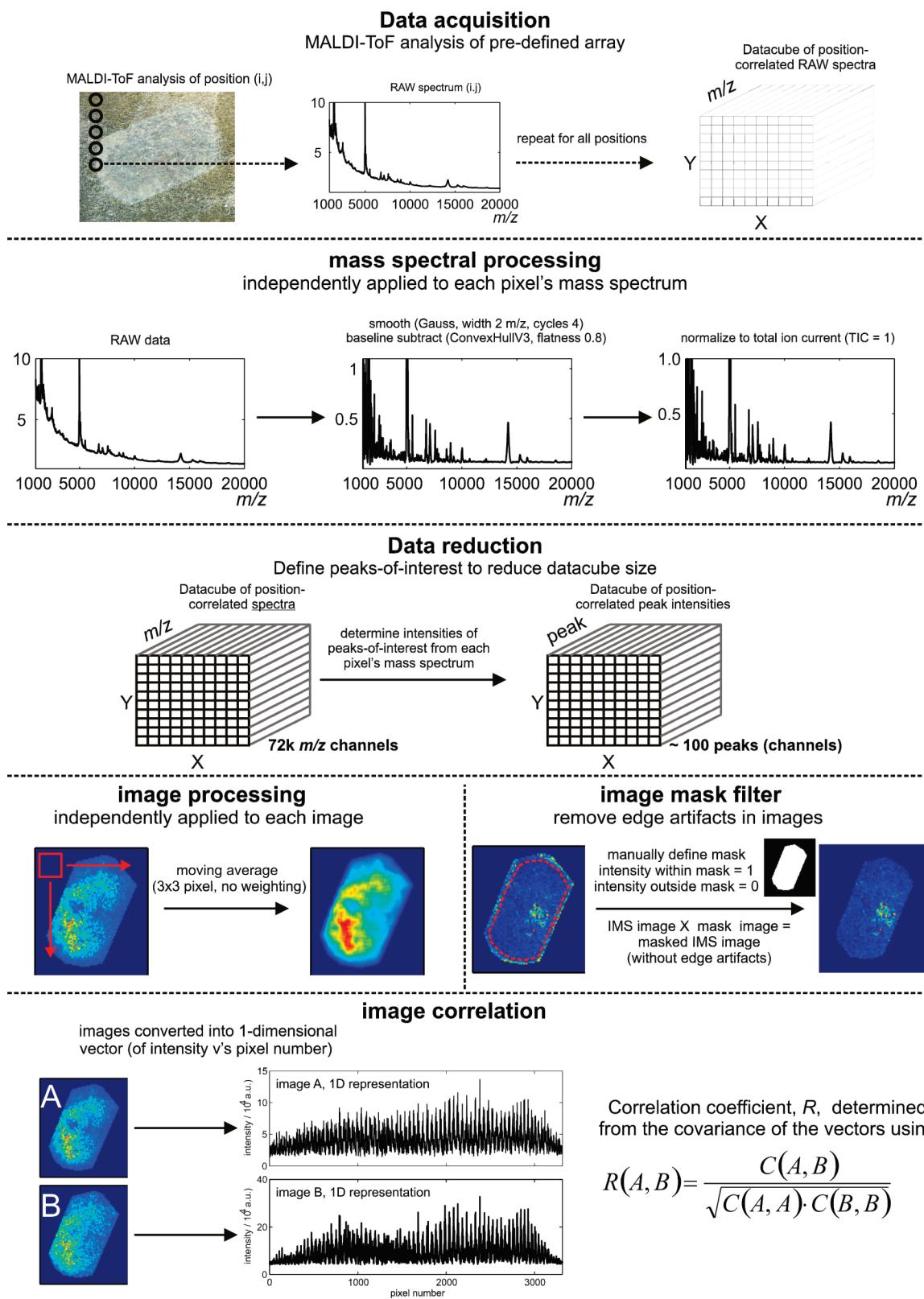
The data cube of position correlated spectra was too large to fit into memory (a common theme in imaging mass spectrometry data analysis). Data reduction was performed by creating a list of peaks-of-interest with FlexImaging. The intensity of each peak in this filter list was then extracted from each pixel's mass spectrum. The resulting image data cube (an image for each peak defined in the filter list) is much smaller than the mass spectral data cube, 18 Mb versus 3 Gb, respectively, thus making it much more amenable to further processing.

To reduce the pixel-to-pixel variation in the IMS images, the images were processed using a moving average filter of  $3 \times 3$  pixels, with a range of weighting algorithms. All results included here were obtained with a uniform weighting. This averaging is referred to as image processing.

An image mask filter was used to omit the high intensity areas frequently observed at the edge of tissue sections. In this process, the user defines which area of the imaging mass spectrometry data set to include. The program creates a binary image-mask (1 for pixels within the area-of-interest, 0 for all other areas). Multiplication of the IMS images with the image mask then provides the masked IMS images.

The final step of the calculation is image correlation. The IMS images were reshaped into one-dimensional vectors and the Pearson correlation between these vectors was calculated. This approach is similar to that used by Zubarev and coworkers for determining the similarity between MS/MS mass spectra of modified and unmodified peptides.<sup>23</sup>

The RAW data, processed data, mass filter lists, data processing lists and high resolution optical images were automatically imported into Matlab version 7.4.0. using custom-developed user programs. The program reduces the array of position-correlated mass spectra (total file size 3 Gb) to an image data cube and a total mass spectrum, in which the image cube contains all of the images defined by the mass filter list, and the total mass spectrum

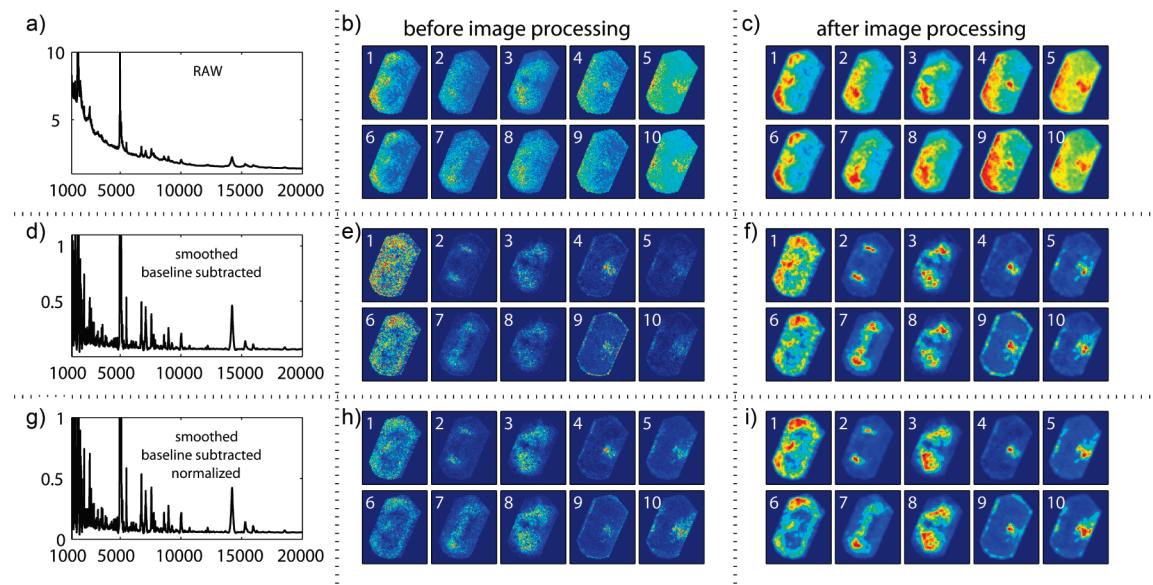
**Scheme 1.** Schematic of Data Acquisition, Data Processing and Correlation Strategies Used

is the mass spectrum accumulated over all positions. This data reduction, final file size 18 Mb, enabled the subsequent image calculations to be performed on a standard windows desktop computer (Intel dualcore 2.66 GHz, 4Gb RAM). Image processing and mask application were performed by exploiting functions supplied with the Image Processing Toolbox (*imfilter* and *roipoly*,

respectively). Image correlation exploited the *corrcoef* function supplied within Matlab.

## Results and Discussion

Imaging mass spectrometry of a tissue section using MALDI can provide the distributions of hundreds of molecular species.



**Figure 1.** MALDI analysis of rat brain tissue sections showing the effects of mass spectral and image processing on the total spectra and a selection of images. The nominal masses of the images are 1 ( $m/z$  1499, lipid dimer), 2 ( $m/z$  4618), 3 ( $m/z$  14 207), 4 ( $m/z$  1354), 5 ( $m/z$  15 954), 6 ( $m/z$  1475, lipid dimer), 7 ( $m/z$  6750), 8 ( $m/z$  7097), 9 ( $m/z$  1234), 10 ( $m/z$  15 275). Scheme 1 details the mass spectral and image processing steps used.

Initial attempts to correlate images provided correlation coefficients that were evidently too large (indicating highly correlated). Closer examination of the data revealed that the spatial distribution of the MALDI background is highly correlated, and its contribution is responsible for the inflated correlation coefficients. This background cannot be removed by varying the intensity scale and offset as the background is not constant across the sample. Instead, background subtraction must be performed in the mass spectral domain prior to generation of the images. Furthermore, previous work on secondary ion mass spectrometry images found that uncertainty in signal intensities could lead to underestimates of the correlation (data not shown).

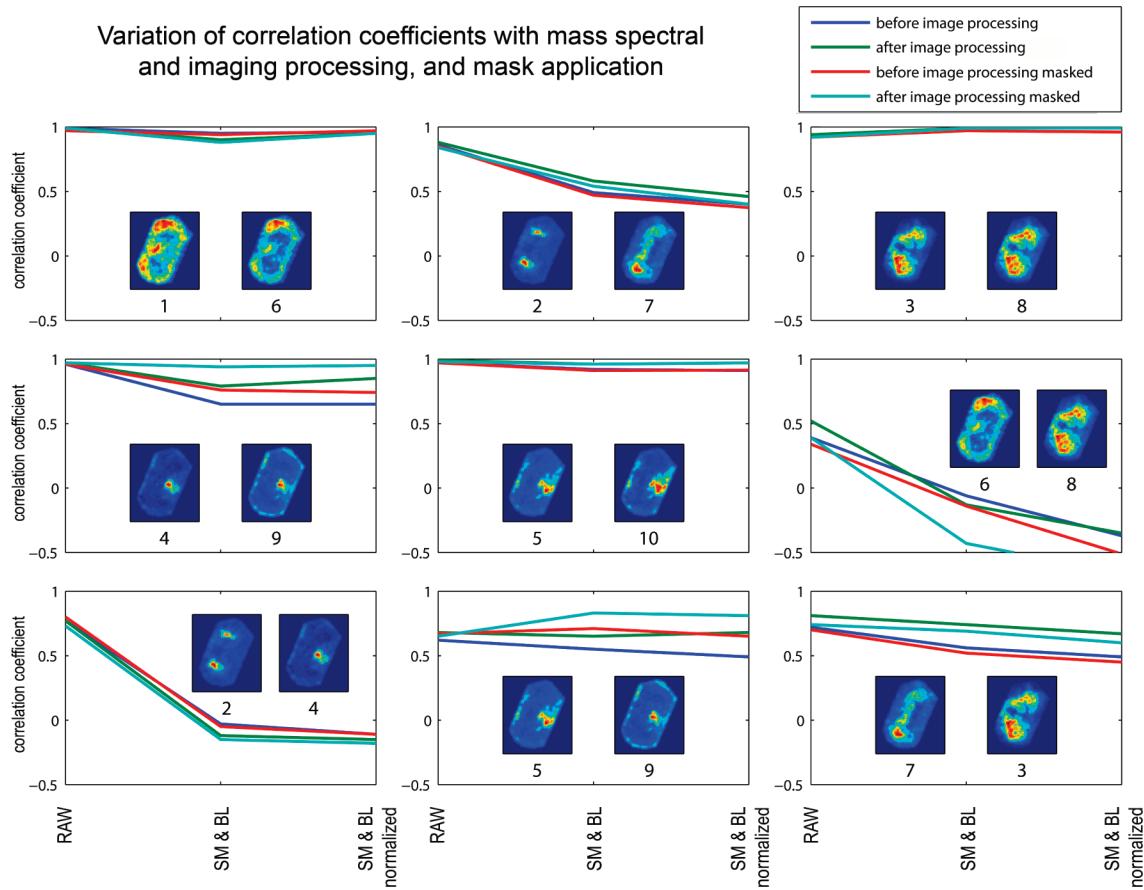
Figure 1 shows the improvements that can be made in MALDI mass spectrometry images by data processing in both the mass spectral and image domains. This example uses a MALDI imaging data set of rat brain tissue sections that includes a large background. Figure 1a shows the peptide and small protein region of the total mass spectrum prior to any processing (RAW data). The background is clearly visible as an almost exponentially decreasing nonspecific signal. Figure 1b shows a selection of 10 images extracted from this RAW data and Figure 1c shows the effects of image processing. The high background and high level of noise (pixel-to-pixel variation in signal) in images from the RAW data result in images that, at best, only hint at distributions. In this case, processing the images to reduce the noise provides more clarity, but images still contain the variation of the background.

Figure 1d shows the total mass spectrum after individually smoothing and baseline subtraction of each pixel's mass spectrum. Many of the lower intensity peaks are more clearly distinguished and now provide images much less tainted by the nonspecific background (Figure 1e). Image processing reduces the noise in the images to provide images that clearly reveal the biomolecular distributions, Figure 1f. Smoothing, background subtraction followed by normalization of each pixel's mass spectrum was used to generate Figure 1g–i. The total mass spectrum and images are quite similar to those

images without normalization (Figure 1d–f) except for subtle changes in the intensities of peaks and in the intensities (contrast) of the distributions. It is clear from Figure 1 that the combination of data processing in the mass spectral and image domains provides images with much more clearly distinguished distributions.

The effects of mass spectral and image processing on the correlation coefficients for a selection of image pairs are shown in Figure 2. Close examination reveals that, if the images are highly similar and possess sufficient signal (high intensity pixels), mass spectral and image processing do not significantly affect the correlation coefficients (image pairs 1 and 6, 3 and 8, and 5 and 10). However, it must be borne in mind that if two images are highly correlated and possess sufficient signal the additional positive correlation from the noise is proportionally less. For images that are dissimilar (image pairs 6 and 8, and 2 and 4), removal of background is crucial to get a more accurate value of their relative distributions. For images that have some similarity and some dissimilarity (image pairs 2 and 7, 4 and 9, and 7 and 3) the effects of the processing were less clear: removal of the background removes the positive correlation of the background while the image processing can increase the correlation as it reduces noise in the images (that would otherwise lower the correlation coefficients) and marginally diffuses the images. For example, for image pair 7 and 3, the correlation coefficient obtained after smoothing and baseline subtraction of each pixel's mass spectrum as well as image processing appears to overestimate the similarity of the images, and one might be tempted to conclude that no such image processing should be performed for MALDI imaging data sets. In contrast, the correlation coefficient for image pair 4 and 9 obtained after smoothing and baseline subtraction of each pixel's mass spectrum, but without image processing, appears to underestimate the similarity of the images. Closer inspection of image 9 reveals that this is probably due to the higher intensity at the edge of the tissue section.

The varying chemical nature of the tissue section exhibits itself as a spatially correlated background signal. This artifact



**Figure 2.** Effects of mass spectral processing, image processing and applying an image mask on the correlation parameters for selected image pairs. The nominal masses of the images are the same as Figure 1.

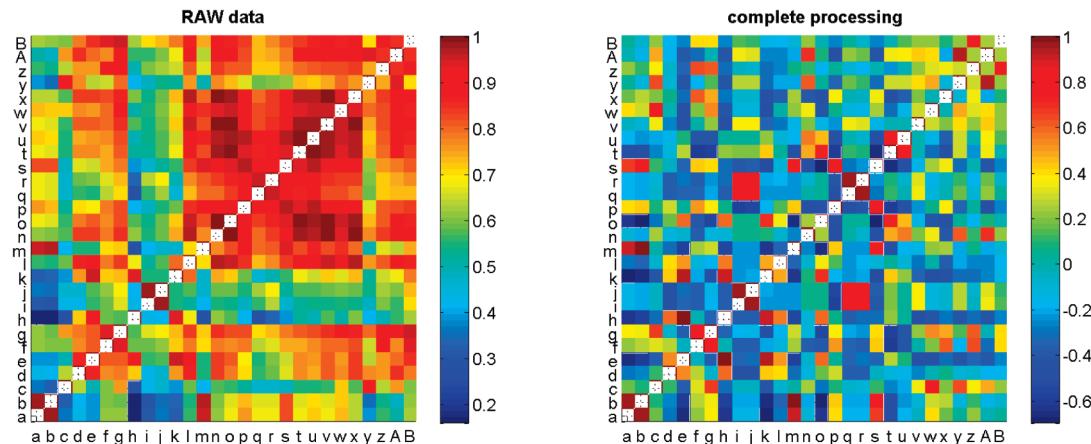
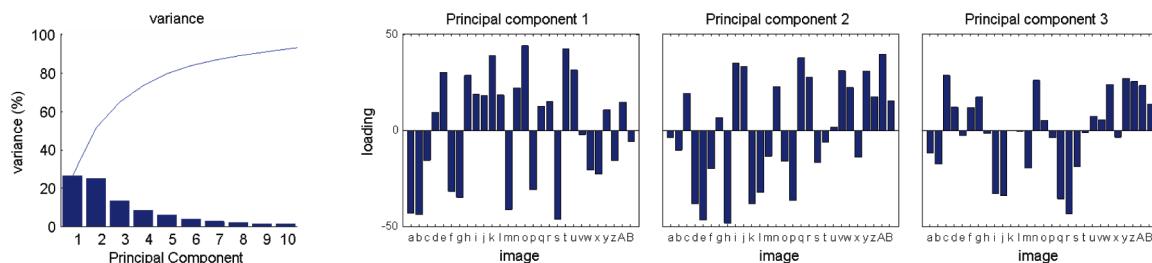
was reduced using background subtraction. Another artifact that can interfere with the correlation analysis is the very different chemical signals obtained from the tissue section and the sample support plate. In these experiments, the contribution of the sample support plate was reduced by defining the imaging mass spectrometry experiments to analyze only the tissue section and the area immediately surrounding the tissue section. However, in some experiments and only for some molecular species, for example in image 9, the border between the tissue section and the sample plate is associated with higher signal intensity. The different chemical environment of the sample plate and the tissue section in this border region can give rise to different matrix crystal growth, and subsequently different MALDI signals. The questions then arise whether this border area should be omitted when determining correlation parameters, and whether an image intensity offset should be used to further reduce the influence of residual background.

The red lines and light blue lines in Figure 2 show the correlation parameters calculated for the RAW and processed (in the mass spectral and image domains) images after an image mask has been applied to omit the high intensity edge area of image 9 from the complete image data cube (see Scheme 1 for image mask details). It is clear that for image pairs 4 and 9 (correlated), 6 and 8 (anticorrelated – offscale coefficient = -0.66), and 5 and 9 (correlated) that all three processing steps of mass spectral processing, image processing and mask application are necessary to obtain correlation coefficients that resemble the visual distributions of each image pair. For the remaining image pairs, the processing steps do not significantly impact the calculated correlations. The dif-

ferential signals at the tissue edge have a physical cause: different matrix crystals and thus MALDI signals, and therefore, there is good justification to omit this area from the correlation coefficients. Indeed, it is our finding that it is recommended.

The justification for arbitrarily setting an intensity threshold to reduce the effects of residual background in the calculations is less clear, because one cannot distinguish between residual background and real signal. Furthermore, it was found that an intensity offset of 10% of the maximum image intensity gave a mean reduction of correlation coefficients of just 0.02. The lack of justification, especially for the larger offsets needed to impact the distributions, argues against the application of an arbitrary intensity offset.

Figure 3a shows the correlation matrices of the RAW data and the fully processed data (mass spectral smoothing, baseline subtraction, and normalization as well as image processing and mask application) for 28 lipid (dimer), peptide and protein peaks. The complete array has 756 entries (28 of which are self-correlation entries along the diagonal, which have been colored white in order to be less distracting). Such an array is impractical in tabular form; here, the data is shown as a correlation matrix, in which the correlation is color-coded, thus, enabling the analyst to quickly identify those peaks that are highly correlated/anticorrelated. Prior to the mass spectrometry, image and mask processing steps the correlation map indicates that a large number of the images are correlated. It was shown in Figures 1 and 2 that this positive correlation is due to the nonspecific background, the noisy signal and the enhanced intensities associated with the edge of the tissue. To obtain more realistic correlations, the influence of these artifacts needs

**a) correlation matrix of 28 peptide, protein & lipid dimer signals****b) variance and loadings following PCA of processed data**

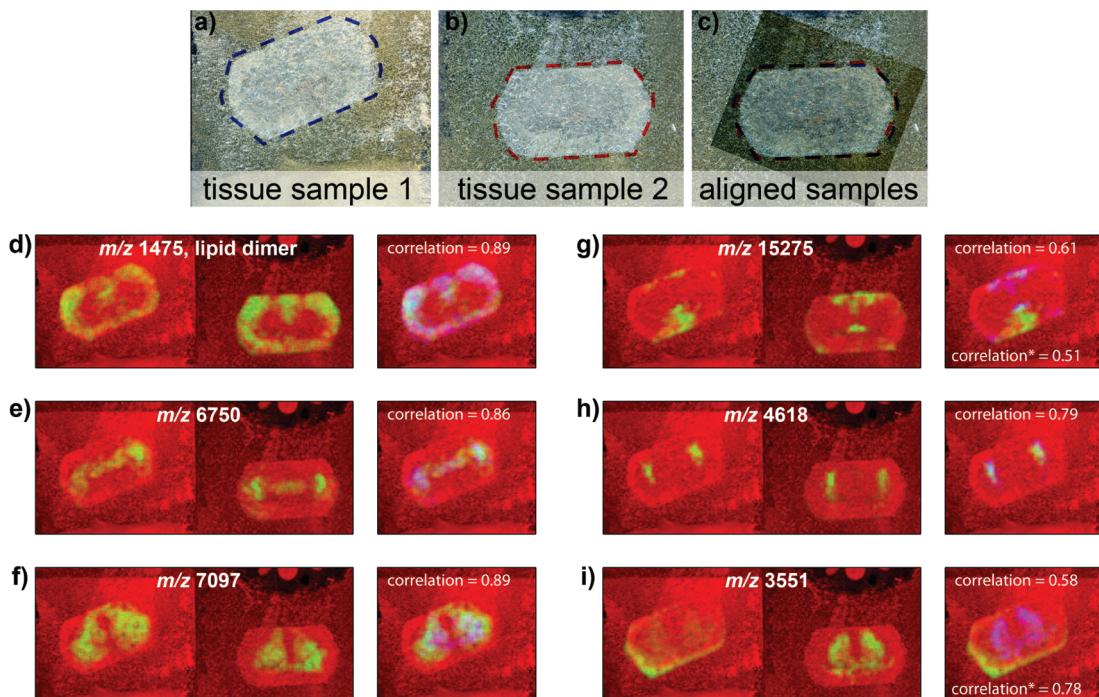
**Figure 3.** (a) Correlation matrix describing the similarities of the distributions of 28 biomolecular ions in the rat brain tissue sections for the RAW data, left, and after mass spectral and image processing, right. (b) Variance of the first 10 principal components and the cumulative variance (blue line), left-hand image, and the loadings of the 28 images on the first 3 principal components, following PCA of the 28 image data cube after mass spectral and image processing. Nominal *m/z* values of the peaks are a, 1475, lipid dimer; b, 1499, lipid dimer; c, 4970; d, 6750; e, 7097; f, 8956; g, 10 010; h, 14 207; i, 15 275; j, 15 954; k, 18 540; l, 4618; m, 1528, lipid dimer; n, 3551; o, 2612; p, 1300; q, 1234; r, 1354; s, 1710; t, 2038; u, 1870; v, 3080; w, 2447; x, 3475; y, 5095; z, 5510; A, 5191; B, 10 709.

to be reduced. The correlation matrix following all data processing steps, also shown in Figure 3a, clearly reveals which ions are correlated, anticorrelated and largely independent (see figure caption for mass spectrometry details about the 28 images). Figure 3b shows the results of principal component analysis of the processed 28 image data cube. These plots display the variance described by the first 10 principal components (and the cumulative variance, blue line), and the loadings of the 28 images in the first 3 principal components. PCA results are commonly interpreted as peaks with same loading-sign (positive/negative) in a given principal component are correlated. However, PCA does not quantify correlation, it maximizes the variance in the entire data set (the first PC describes the maximum variance, the second PC the maximum variance orthogonal to the first PC, and so on). The 'correlations' determined by PC are thus dependent on which PC is investigated and include contributions from the *entire* data set. Considering that the variance of the first 5 PCs are significant, no single principal component can be used to quantify correlations between images.

Correlating images within a data set establishes the similarities between molecular distributions. This is useful to test hypotheses in which two ions are thought to be related, for example, through biological, pathological or pharmacological action. To adequately test any such hypothesis, it is important that the experiments are repeatable. In all biomarker discovery investigations that target biochemical changes associated with a biological, pathological or pharmacological state, it is important to establish the significance of an observation:<sup>19</sup> in

imaging mass spectrometry, this can be formulated by investigating the repeatability of correlations within data sets and investigating the repeatability of a given molecule's distribution in different samples. This second aspect is crucially important when the focus of a research question is a change in distribution associated with a particular state.

Figure 4 shows an example of the alignment and correlation of two imaging mass spectrometry data sets. Figures 4a,b show optical images of two rat brain tissue sections after MALDI sample preparation, and include the areas selected for the imaging experiments (the data shown in Figures 1–3 was obtained from tissue sample 1). As can be seen, the tissue sections are located in different positions and in different orientations. The areas selected for imaging experiments are typically manually defined to encompass the tissue sections and the areas immediately surrounding the tissue sections. Consequently, any alignment must be performed using the optical images. Figure 4c shows an example of aligned tissue sections, in which the outline of the tissue sections was used to align the samples (the areas defined for the imaging mass spectrometry experiment are defined manually and thus do not match). In this case, the tissue sections were taken from the same rat brain and separated by just 72 µm, and careful placement of the tissue sections led to little spatial deformation. Consequently, alignment of the tissue section images required only translation and rotation. For tissue sections from different specimens that differ more substantially, automated alignment routines can be used that include translation, rotation, scaling and warping.



**Figure 4.** Alignment and correlation of two imaging mass spectrometry data sets.

The alignment of any two images that involves rotation, scaling or warping implicitly involves a projection of one data set onto the coordinate system of the other data set. To ensure that the IMS images were aligned on a common coordinate system, a composite RGB image cube was created in which the red channel corresponded to the (resolution adjusted) optical image, the green channel to the data sets prior to alignment, and the blue channel to the results of the alignment procedure for data set 2. Figures 4d–i show the original distributions from the two samples (blue channel set to zero) and the results of alignment and the calculated correlation parameters for six biomolecular ions. These images were selected to include examples showing significant correlation,  $m/z\ 1475$ , 6750 and 7097 (correlation coefficients  $>0.85$ ), and examples in which the correlation was not as successful,  $m/z\ 15\ 275$  and 3551. In the final example,  $m/z\ 3551$ , tissue sample 1 contains a faint distribution that matches the distribution shown in tissue sample 2; however, this faint distribution is dominated by strong signals associated with the edge of the tissue. Using an image mask (re. Scheme 1) to omit the strong signals at the edge of the tissue increases the correlation coefficient from 0.58 to 0.78, indicating their similarity. A similar edge effect is present in the  $m/z\ 15\ 275$  images, Figure 4g, but in this case, the strong signals associated with the tissue edge are present in both images. The correlation between these edge artifacts increases the apparent correlation coefficient between the images from the two samples. Using the image mask to omit the edge artifact reduces the correlation parameter from 0.61 to 0.51.

The higher signals associated with the edge of the tissue sections (in some images) and the visible differences between the images from the two samples (even for those with high correlation coefficients) currently limit the repeatability of an experiment. These factors arise from the sample preparation steps of a MALDI IMS experiment. Sample preparation is a key element of all IMS experiments, and several thorough investigations have been reported.<sup>24–28</sup> Commercial, automated

sample-preparation devices (Bruker's ImagePrep, Labcyte's Portrait<sup>TM+</sup> 630, Leap Technologies' TM-Sprayer) now enable more consistent sample preparation, but sample-to-sample variation still occurs and, thus, limits the repeatability of an IMS experiment (and thus influences what changes can be reliably detected). The correlation determination reported here will enable the repeatability of an experiment to be included during the development of sample preparation strategies.

The data processing strategy described above consists of processing each pixel's mass spectrum followed by image filtering and finally image correlation. The results shown here reaffirm the necessity of mass spectral processing in order to extract reliable protein distributions from the data sets<sup>29</sup> and demonstrate that the reduction of noise in the images provided by the image processing steps embolden the molecular distributions. Both mass spectral processing and image processing are necessary for obtaining reliable correlation parameters.

The correlation parameters were calculated by representing each image as a vector and calculating the correlation between the vectors. This method was chosen to ensure that the methodology was applicable to region-of-interest analysis, that is, selecting an area using an image mask, and to the arbitrarily shaped areas defined in FlexImaging (an IMS experiment is often limited to the tissue-section and the immediate surrounding area in order to minimize experiment duration). In this manner, the correlation calculation is limited to areas defined in the experiment or by the analyst during data analysis, for example, in Figures 2 and 4 in which it was demonstrated that the high signals associated with the tissue edge needed to be removed using an image-mask in order to obtain more reliable correlation parameters. Additional experiments that analyzed rectangular areas containing the tissue of interest, instead of limiting the experiment to the tissue and the immediate surrounding area, exacerbated the artifactual positive-correlation due to the differential signals on and off

the tissue section. These results clearly indicated the need to focus the calculation on the areas of interest, namely, the tissue section.

Image correlation routines are extensively used in many fields ranging from biological sciences to satellite image analysis to fingerprint identification. In confocal microscopy, localized image correlation techniques, based on pixel neighborhood analysis, have been developed to identify colocalized structures defined by virtue of shapes, position, and contrast.<sup>30,31</sup> Such techniques have been developed further to analyze time-lapse images and, thus, determine the velocity fields of tagged proteins.<sup>32</sup> Another common biomolecular application of image correlation is automated alignment of 2D-gels. In all of these examples, image correlation involves a small number of channels (for example, RGB). The correlations within an imaging mass spectrometry data set involve correlating each image with every other image. Figure 3 shows a correlation map for just 28 images selected from more than 100. For such a large number of channels, examining local correlations, as performed with fluorescence microscopy, will demand additional tools to sort through the large number of correlation coefficients for each position.

The image processing used here to smooth the images consists of a simple moving average in which each pixel of the original image was converted to the average of  $3 \times 3$  pixels (centered on the pixel of interest) all with equal weighting. Different center-biased weightings (including linear, Gaussian and Laplacian) and circular moving averages were investigated. In all cases, it was possible to generate visibly smoothed images. Provided the images were visibly smoothed, the resulting correlation parameters were not significantly changed. There are many image processing tools and correlation functions that could be applied to imaging mass spectrometry data sets. For example, PCA could be used to reduce the data set to those signals that provide the most variance (for example, reconstituting the data set using the first 20 principal components) and wavelet filtering could be used to remove the higher frequency noise. It is beyond the scope of this paper to compare all of the different data processing strategies for correlating imaging mass spectrometry data sets. Rather, it is hoped that the research presented here will prompt a wider effort to establish which of the myriad of data processing strategies is best for quantifying the similarity of imaging mass spectrometry data sets.

## Conclusion

The results presented here demonstrate that the similarity of the distributions of specific molecules can be quantified within and between imaging mass spectrometry data sets. Extensive investigations into the mass spectral processing, image processing and mask applications have established the effects of each step on the correlation coefficients and has established that all steps are necessary to obtain reliable correlation coefficients. The resulting correlation matrix allows the user to rapidly identify and quantify correlations, and to determine the repeatability of imaging mass spectrometry results. This latter aspect is crucial for determining the reliability of changes in distributions associated with a pathology or application of a pharmaceutical.

**Acknowledgment.** L.A.M. gratefully acknowledges Roger Adan and Linda Verhagen for the tissue sections and Jean Davoust (Généthon, France) for helpful discussions.

## References

- Cornett, D. S.; Reyzer, M. L.; Chaurand, P.; Caprioli, R. M. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat. Methods* **2007**, *4*, 828–833.
- McDonnell, L. A.; Heeren, R. M. A. Imaging mass spectrometry. *Mass Spectrom. Rev.* **2007**, *26*, 606–643.
- Touboul, D.; Kollmer, F.; Niehuis, E.; Brunelle, A.; Laprévote, O. Improvement of biological time-of-flight secondary ion mass spectrometry imaging with bismuth cluster ion source. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1608–1618.
- Altelaar, A. F. M.; Taban, I. M.; McDonnell, L. A.; Verhaert, P. D. E. M.; de Lange, R. P. J.; Adan, R. A. H.; Mooi, W. J.; Heeren, R. M. A.; Piersma, S. R. High-resolution MALDI imaging mass spectrometry allows localization of peptide distributions at cellular length scales in pituitary tissue sections. *Int. J. Mass Spectrom.* **2007**, *260*, 203–211.
- Chaurand, P.; Rahman, M. A.; Hunt, T.; Mobley, J. A.; Gu, G.; Latham, J. C.; Caprioli, R. M.; Kasper, S. Monitoring mouse prostate development by profiling and imaging mass spectrometry. *Mol. Cell. Proteomics* **2008**, *7*, 411–423.
- Hsieh, Y.; Casale, R.; Fukuda, E.; Chen, J.; Knemeyer, I.; Wingate, J.; Morrison, R.; Korfmacher, W. Matrix-assisted laser desorption/ionization imaging mass spectrometry for direct measurement of clozapine in rat brain tissue. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 965–972.
- Khatib-Shahidi, S.; Andersson, M.; Herman, J. L.; Gillespie, T. A.; Caprioli, R. M. Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry. *Anal. Chem.* **2006**, *78*, 6448–6456.
- Stoeckli, M.; Staab, D.; Staufenbiel, M.; Wiederhold, K. H.; Signor, L. Molecular imaging of amyloid beta peptides in mouse brain sections using mass spectrometry. *Anal. Biochem.* **2002**, *311*, 33–39.
- Svensson, M.; Sköld, K.; Nilsson, A.; Fälth, M.; Svensson, P.; Andren, P. E. Neuropeptidomics: expanding proteomics downwards. *Biochem. Soc. Trans.* **2007**, *35* (3), 588–593.
- Stauber, J.; Lemaire, R.; Franck, J.; Bonnel, D.; Croix, D.; Day, R.; Wiszterski, M.; Fournier, I.; Salzet, M. MALDI imaging of formalin-fixed paraffin-embedded tissues: application to model animals of parkinson disease for biomarker hunting. *J. Proteome Res.* **2008**, *7*, 969–978.
- Stoeckli, M.; Chaurand, P.; Hallahan, D. E.; Caprioli, R. M. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nat. Med.* **2001**, *7*, 493–496.
- Chaurand, P.; Schwartz, S. A.; Caprioli, R. M. Profiling and imaging proteins in tissue sections by MS. *Anal. Chem.* **2004**, *76* (5), 86A–93A.
- Touboul, D.; Brunelle, A.; Halgand, F.; De La Porte, S.; Laprévote, O. Lipid imaging by gold cluster time-of-flight secondary ion mass spectrometry: application to Duchenne muscular dystrophy. *J. Lipid Res.* **2005**, *46*, 1388–1395.
- Chaurand, P.; Schwartz, S. A.; Billheimer, D.; Xu, B. J.; Crecelius, A.; Caprioli, R. M. Integrating histology and imaging mass spectrometry. *Anal. Chem.* **2004**, *76*, 1145–1155.
- Cornett, D. S.; Mobley, J. A.; Dias, E. C.; Andersson, M.; Arteaga, C. L.; Sanders, M. E.; Caprioli, R. M. A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer. *Mol. Cell. Proteomics* **2006**, *5*, 1975–1983.
- McCombie, G.; Staab, D.; Stoeckli, M.; Knochenmuss, R. Spatial and spectral correlations in MALDI mass spectrometry images by clustering and multivariate analysis. *Anal. Chem.* **2005**, *77*, 6118–6124.
- Schwamborn, K.; Krieg, R. C.; Reska, M.; Jakse, G.; Knuechel, R.; Wellmann, A. Identifying prostate carcinoma by MALDI-imaging. *Int. J. Mol. Med.* **2007**, *20*, 155–159.
- Chaurand, P.; Schwartz, S. A.; Caprioli, R. M. Assessing protein patterns in disease using imaging mass spectrometry. *J. Proteome Res.* **2004**, *3*, 245–252.
- Hilaro, M.; Kalousis, A.; Pellegrini, C.; Müller, M. Processing and classification of protein mass spectra. *Mass Spectrom. Rev.* **2006**, *25*, 409–449.
- Taguchi, F.; Solomon, B.; Gregorc, V.; Roder, H.; Gray, R.; Kasahara, K.; Nishio, M.; Brahmer, J.; Spreafico, A.; Ludovini, V.; Mission, P. P.; Dziadziuszko, R.; Schiller, J.; Grigorieva, J.; Tsypin, M.; Hunsucker, S. W.; Caprioli, R.; Duncan, M. W.; Hirsch, F. R.; Bunn Jr, P. A.; Carbone, D. P. Mass spectrometry to classify non-small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study. *J. Natl. Cancer Inst.* **2007**, *99* (11), 838–846.

- (21) Stauber, J.; Lemaire, R.; Franck, J.; Bonnel, D.; Croix, D.; Day, R.; Wisztorski, M.; Fournier, I.; Salzet, M. MALDI imaging of formalin-fixed paraffin-embedded tissues: application to model animals of parkinson disease for biomarker hunting. *J. Proteome Res.* **2008**, *7* (3), 969–978.
- (22) Gavin, A.-C.; Aloy, P.; Grandi, P.; Krause, R.; Boesche, M.; Marzioch, M.; Rau, C.; Jensen, L. J.; Bastuck, S.; Dümpelfeld, B.; Edelmann, A.; Heurtier, M.-A.; Hoffman, V.; Hoeffert, C.; Klein, K.; Hudak, M.; Michon, A.-M.; Schelder, M.; Schirle, M.; Remor, M.; Rudi, T.; Hooper, S.; Bauer, A.; Bouwmeester, T.; Casari, G.; Drewes, G.; Neubauer, G.; Rick, J. M.; Kuster, B.; Bork, P.; Russell, R. B.; Superti-Furga, G. Proteome survey reveals modularity of the yeast cell machinery. *Nature* **2006**, *440* (30), 631–636.
- (23) Savitski, M. M.; Nielsen, M. L.; Adams, C. M.; Bergström, S.; Hedberg, L.; Pettersson, U.; Zubarev, R. A. In Study of Fragmentation Patterns of Modified Peptides and Their Native Counterparts, 55th ASMS Conference on Mass Spectrometry and Allied Topics, Indianapolis, IN, 2007.
- (24) Aerni, H.-R.; Cornett, D. S.; Caprioli, R. M. Automated acoustic matrix deposition for MALDI sample preparation. *Anal. Chem.* **2006**, *78*, 827–834.
- (25) Kruse, R.; Sweedler, J. V. Spatial profiling invertebrate ganglia using MALDI MS. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 752–759.
- (26) Lemaire, R.; Tabet, J. C.; Ducoroy, P.; Hendra, J. B.; Salzet, M.; Fournier, I. Solid ionic matrixes for direct tissue analysis and MALDI imaging. *Anal. Chem.* **2006**, *78*, 809–819.
- (27) Lemaire, R.; Wisztorski, M.; Desmons, A.; Tabet, J. C.; Day, R.; Salzet, M.; Fournier, I. MALDI-MS direct tissue analysis of proteins: improving signal sensitivity using organic treatments. *Anal. Chem.* **2006**, *78* (20), 7145–7153.
- (28) Seeley, E. H.; Oppenheimer, S. R.; Mi, D.; Chaurand, P.; Caprioli, R. M. Enhancement of protein sensitivity for MALDI imaging mass spectrometry after chemical treatment of tissue sections. *J. Am. Soc. Mass Spectrom.* **2008**, in press.
- (29) Norris, J. L.; Cornett, D. S.; Mobley, J. A.; Andersson, M.; Seeley, E. H.; Chaurand, P.; Caprioli, R. M. Processing MALDI mass spectra to improve mass spectral direct tissue analysis. *Int. J. Mass Spectrom.* **2007**, *260*, 212–221.
- (30) Wiseman, P. W.; Petersen, N. O. Image correlation spectroscopy. ii. optimization for ultrasensitive detection of preexisting platelet-derived growth factor-b receptor oligomers on intact cells. *Biophys. J.* **1999**, *76*, 963–977.
- (31) Costes, S. V.; Daemans, D.; Cho, E. H.; Dobbin, Z.; George Pavlakis, Z.; Lockett, S. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys. J.* **2004**, *86*, 3993–4003.
- (32) Hebert, B.; Costantino, S.; Wiseman, P. W. Spatiotemporal Image correlation spectroscopy (STICS) theory, verification, and application to protein velocity mapping in living CHO cells. *Biophys. J.* **2005**, *88* (5), 3601–3614.

PR800214D