Data analysis for mass spectrometry imaging : methods and applications
Abdelmoula Walid, Mohamed

Cover Page

The handle http://hdl.handle.net/1887/45501 holds various files of this Leiden University dissertation

Author: Abdelmoula, Walid M.
Title: Data analysis for mass spectrometry imaging : methods and applications
Issue Date: 2017-01-18

AUTOMATIC REGISTRATION OF MASS SPECTROMETRY IMAGING DATA SETS TO THE ALLEN BRAIN ATLAS

2

Mass spectrometry imaging holds great potential for understanding the molecular basis of neurological disease. Several key studies have demonstrated its ability to uncover disease-related biomolecular changes in rodent models of disease, even if highly localized or invisible to established histological methods. The high analytical reproducibility necessary for the biomedical application of mass spectrometry imaging means it is widely developed in mass spectrometry laboratories. However many lack the expertise to correctly annotate the complex anatomy of brain tissue, or have the capacity to analyze the number of animals required in pre-clinical studies, especially considering the significant variability in sizes of brain regions. To address this issue we have developed a pipeline to automatically map mass spectrometry imaging datasets of mouse brains to the Allen Brain Reference Atlas, which contains publically available data combining gene expression with brain anatomical locations. Our pipeline enables facile and rapid inter-animal comparisons by first testing if each animal's tissue section was sampled at a similar location, and enabling the extraction of the biomolecular signatures from specific brain regions.

2.1. INTRODUCTION

MASS spectrometry imaging (MSI) uses spatially resolved proteomics and metabolomics mass spectrometry methods to simultaneously record the distributions of hundreds of endogenous molecules directly from tissue samples [1]. Using essentially the same technology, peptides, proteins, pharmaceuticals and metabolites can be analyzed, without the need for labeling and without prior knowledge of their presence or location within the tissue. One of the driving forces behind the increasing popularity of MSI is its ability to determine relevant molecular biomarkers for disease, in clinical tissue samples and preclinical animal models of disease.

An essential component of biomarker discovery experiments is the need for histological specification because of the variety of cell types present in diseases such as cancer, or the distinct histoarchitecture in tissues such as the brain. The ability to apply histological stains to tissues after the MSI experiment has facilitated the integration of MSI with histology [72]. Nevertheless accurate histological/anatomical annotation of the tissue represents one of the principal challenges for many mass spectrometry laboratories, as it requires detailed knowledge of the histology/anatomy of the tissue. For example a preclinical study focused on the effect of human pathogenic mutations on a neurological disease that is modeled by a specific surgical intervention can involve 6 groups of animals (wild type control; wild type test animal; wild type sham; transgenic control; transgenic test animal; transgenic sham). Just 6 animals in each group leads to the comparison of 36 animals. A high level of expertise in brain anatomy is necessary to ensure accurate anatomical annotation and sufficient throughput to perform such studies. It is for this reason that the clinical application of MSI often involves close collaborations with pathologists and basic (neuro)scientists that are experts in the disease or in the anatomy of the tissue.

Whereas the often complex histology of, for example, cancer tissues requires manual expert histopathological assessment of each patient tissue, reference anatomical atlases are available for rodent brains. The Allen Brain Atlas (ABA)[69] is a publicly accessible collection of extensive gene expression and neuroanatomical reference data for the

mouse brain, developing mouse brain and human brain. The mouse brain reference atlas includes high-resolution histological images accompanied by a hierarchically organized taxonomy of brain structures, as well as online tools for its exploration. George Paxinos' Rat Brain in Stereotaxic Coordinates [73] provides reference data for the large rat brains. Such reference atlases have enabled non-experts to relate MSI results to the histo-architecture of rodent brain tissue sections. For example, Alexandrov *et al.* [31] have used a rat brain tissue atlas to optimize edge-preserving smoothing and cluster analysis of MSI data so that the results reproduced as closely-as-possible the tissue's histo-architecture.

Rodent brain tissue sections remain the most commonly analyzed tissue types in MSI because of their wide availability and widespread use in neurological and neurodegeneration research [74, 75]. Amongst others, MSI has been used to investigate rodent models of Parkinson's disease [7, 76–78], Alzheimer's disease [79, 80], brain tumors [14, 81], stroke [82], migraine [83], epilepsy [84], and obesity [85]. In addition, the molecular information provided by MSI has been used to improve the anatomical definition of the claustrum, a prominent but histologically ill-defined forebrain structure [86]. Until now, the anatomical annotation used in MSI has been performed by visual inspection of the stained tissues. Automatically aligning MSI datasets to reference anatomical tissue atlases, however, would have several key advantages, namely:

  i. Automatic determination of tissue sampling location.

 ii. Automatic annotation of the histological images and MSI datasets.

iii. Facile comparison of animal cohorts, whether by extracting data from specific regions in the mouse brain or comparing distributions in different animals.

 iv. Quantification of morphological changes incurred as the result of disease, or sample preparation protocols.

Automatic alignment of mouse brain tissues and associated MSI datasets has been used for constructing 3D-MSI datasets [87–89]. In this instance a rigid transform is applied to the histological images from sequential tissues sections from a single animal. These rigid transformations were then applied to the sequential 2D-MSI datasets to create the 3D-MSI dataset, which have subsequently been aligned to additional imaging modalities [81, 90]. Such an approach cannot be used for aligning brain tissues from different animals because of the high degree of individual variability. A recent report that compared brain size and the sizes of seven distinct brain regions of more than 10,000 mice found little to low correlation between the sizes of different brain regions, and only mediocre correlation with overall brain size [91]. Accordingly, the ability to place datasets from different animals onto a common coordinate space is essential for comparing the distributions from different animals, and will require elastic deformations to account for differences in brain region size and for deformations that can occur during tissue processing (e.g. folds, tears).

In order to compare the MSI datasets obtained from multiple mice and to automatically annotate each tissue section, we have developed a pipeline for the automatic registration of histological images and their aligned MSI datasets to their best match in

the Allen Brain Atlas. The key elements of the pipeline are: i) image pre-processing; ii) selection of the closest mouse brain tissue section from the Allen Brain Atlas; iii) affine registration to capture global deformations (e.g. scale, translation, and rotation); and iv) non-linear (elastic) image registration to capture local deformations (e.g. local compression or tears) and account for inter-individual variability.

2.2. EXPERIMENTAL SECTION

2.2.1. CHEMICALS AND REAGENTS

All chemicals and reagents were obtained from Sigma-Aldrich (Steinheim, Germany) except ethanol (Merck, Darmstadt, Germany). 9-AA matrix solution (saturated) was prepared in 70% methanol.

TISSUE SAMPLES

Three-month-old, male, C57BL/6J mice bred within the LUMC (Leiden, Netherlands) were sacrificed by cervical dislocation. Brains were excised, flash-frozen in dry ice, and stored at -80 °C for less than one month until analysis. 12 µm thick tissue sections were cut (Bregma level – 1.8 mm) at -12 °C in a cryostat microtome (Leica Microsystems, Wetzlar, Germany), thaw-mounted onto poly-L-lysine coated indium-tin-oxide (ITO) glass slides (Bruker Daltonics, Bremen, Germany) and stored at -80 °C. All experiments were approved by the Animal Experiment Ethics Committee of Leiden University Medical Center.

2.2.2. MASS SPECTROMETRY IMAGING

Tissue sections were equilibrated to room temperature for 30 min. in a vacuum desiccator and then sprayed with 9-AA matrix solution using the SunCollect sprayer (SunChrom, Friedrichsdorf, Germany). MALDI-MSI experiments were performed using an UltrafleXtreme MALDI-TOF/TOF (Bruker Daltonics) in the negative ion mode with 100 μm raster width, 500 laser shots per pixel. Data acquisition, pre-processing and visualization were performed with the flex software suite from Bruker Daltonics: flexImaging 3.0 was used for experiment definition; flexControl 3.4 was used for data acquisition; and flexAnalysis 3.4 was used for mass spectral processing (smoothing and baseline subtraction). After MSI data acquisition the matrix was washed off with 70 % ethanol and the tissue samples stained with cresyl violet (Nissl stain). High-resolution histological images were recorded using a Pannoramic MIDI digital slide scanner (3D Histech, Budapest, Hungary) and then registered to the MSI datasets using flexImaging 3.0.

2.2.3. MS DATA REDUCTION AND COREGISTRATION WITH HIGH-RESOLUTION OPTICAL IMAGES

Co-registration of the MSI data with the high-resolution optical images of the Nissl-stained sections was performed with flexImaging using fiducial registration points. The MSI datasets were then processed and reduced as previously described [28] using custom scripts written in Matlab (MathWorks, Natick, MA). The program is an automated feature detection and feature extraction routine that distils the original raw MSI data

into an image cube containing the distributions of every detected peak. These reduced datasets and their aligned histological images (cropped to include only those parts measured during the MSI experiment) were used in all subsequent steps.

2.2.4. Image Processing and Alignment

The proposed pipeline for image registration was implemented using Matlab and `elastix` [46]. The registration parameters (e.g. cost function, transformation model, optimizer) were carefully selected and passed to `elastix` to efficiently align the high-resolution Nissl-stained images.

2.3. Implementation and Results

In order to automatically compare the MSI datasets obtained from mouse brain tissue sections from different animals we have developed a pipeline for the automatic registration of histological images, and their aligned MSI datasets, to the reference histological images from the Allen Brain Atlas. The pipeline consists of three main steps: i) preprocessing of histological images; ii) automatic selection of target image from the Allen Brain Atlas; and iii) image registration.

2.3.1. Preprocessing of Histological Images

Histological images need to be preprocessed before alignment with images in the Allen Brain Atlas in order to reduce background noise and correct for image acquisition artifacts (e.g. inhomogeneous lighting and exposure, noise due to the presence of dust on the slides). The preprocessing pipeline is depicted in Figure 2.1. First, the optical images of the Nissl-stained tissues were down-sampled from the original 30 M pixel images (approximately 2 μm pixel size) by 0.66X to speed up image processing. The red, green, and blue channels were then separated. Preliminary experiments revealed that the green and blue channels contained the highest and lowest contrast between tissue and background. The background brightness is high in both channels and subtracting the blue channel from the green proved to be highly effective for reducing the image background, thereby enhancing tissue contrast and compensating for non-uniform illumination The image was then smoothed with a 20 x 20 pixel filter to reduce salt-and-pepper noise and clustered using the k-means algorithm (k = 2). A set of morphological operations [92] (opening and closing followed by region-filling) were applied to close any gaps in the k = 2 cluster image. The resulting image is a binary mask that is used to separate the tissue from the background.

2.3.2. Automatic Selection of a Target Image from the Allen Brain Atlas

The MSI experiments focused on the cortex, hippocampus and thalamus (as these brain regions are important for key functions of the brain and of interest to many neuroscientists), corresponding to a region around bregma -1.8 mm of the mouse brain. The experimental design limited the search space of the Allen Brain Atlas to 14 coronal tissue sections between bregma positions -1.22 mm and -2.80 mm. Close examination of the histo-architecture of the 14 candidate atlas tissue sections revealed that the
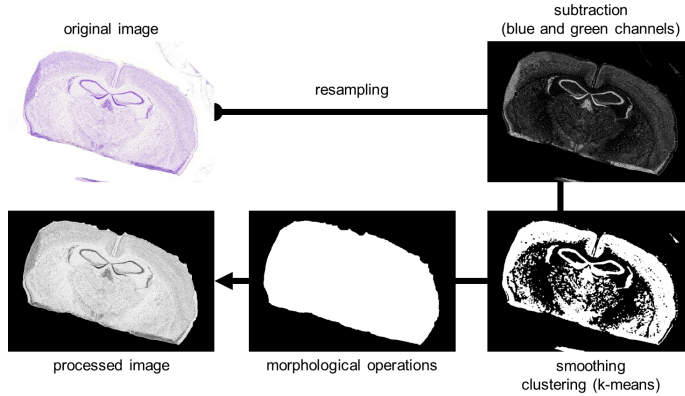
Figure 2.1: Preprocessing pipeline. The high-resolution original image was down-sampled by 0.66X and the blue and green channels subtracted from each other. The subtracted image was then smoothed and clustered using the k-means algorithm (k = 2). A set of morphological operations (opening, closing, region-filling) were then applied to obtain a binary mask that was used to separate the tissue image from the background, generating the processed image.

hippocampus, one of the most prominent features present in the images, could be used to determine the best match for the experimental tissue section. The parameters of the hippocampus, i.e. its maximum chord length, strongly correlate with the position of the atlas tissue section within the region defined by the experimental design (Figure 2.2).

The hippocampus was segmented by setting a global intensity threshold using Otsu's method, followed by applying erosion and dilation using a disk shape structuring element [92], to remove sporadic noise in the thresholded images. The lengths of the major and minor axis of all remaining structures were calculated and only the two with the largest areas were retained. These correspond to the left and right hippocampal structures, Figures 2.2.a) and 2.2.b). The chord length was then calculated for both lateral hemispheres of the brain tissue section and the longer selected as the discriminant feature. A quartile statistical analysis was used to exclude outliers, after which the hippocampal chord length exhibited a near linear relationship, $R^2 = 0.9869$, with the Allen Brain Atlas coronal tissue sections within the selected range, as shown in Figure 2.2.c). The initial location provided by the hippocampus chord length was used to reduce the search space of the Allen Brain Atlas images to three frames for the subsequent image registration step. The rationale was that three sections should provide a reasonable search space (200 μm coverage) following the initial estimate.

2.3.3. IMAGE REGISTRATION

Image registration was divided into two subsequent steps: first a global, affine registration to account for differences in position, orientation and scale (i.e. total brain size) followed by a non-rigid (elastic) transformation to account for local anatomical differences in brain regions and local deformations due to the manual nature of the sectioning process.
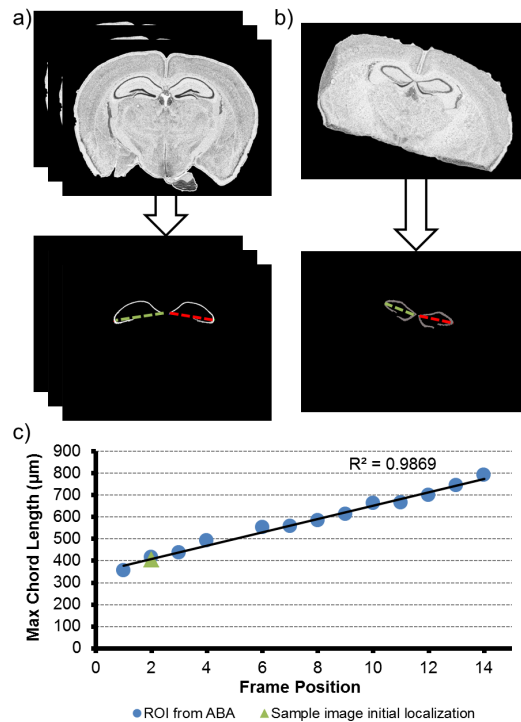
2

Figure 2.2: Automatic selection of a target image from the Allen Brain Atlas. a) Pre-processed mouse brain image from the Allen Brain Atlas (ABA) that was randomly selected from a region of interest (ROI) containing 14 sections (between Bregma -1.22 mm and -2.80 mm); below, the extracted hippocampus structures with the chord length (dotted green and red lines) in both hemispheres. b) Pre-processed sample image and the extracted hippocampus structures with the chord length (dotted red and green lines) in both hemispheres. c) Linear fitting of the extracted chord length versus the frame position in the extracted ROI from the ABA after excluding the outliers ($R^2 = 0.9869$). The green triangle shows the initial position of the example sample image in the ROI from the ABA.

AFFINE REGISTRATION

The binary images of the hippocampus structures were used to estimate the global deformation. Restricting the evaluation to the hippocampus instead of the full histological images was found to increase the reliability of the registration process. The clear geometrical structures of the hippocampus allowed establishing clear spatial correspondences for global registration. The Euclidian distance transform [93] of the binary images was calculated and the adaptive stochastic gradient descent optimizer [94] used to find the set of parameters (i.e. translation, rotation, scaling and shear) that minimized the similarity metric, the sum of squared difference (SSD). This registration process was applied using a multi-resolution approach [58] with the open-source image registration software `elastix` [46]. The Allen Brain Atlas target image with the lowest SSD was considered to be the Allen Brain Atlas image that best matched the sample image.

NON-LINEAR REGISTRATION

Differences in the sizes of brain regions and local deformation were corrected using a B-Spline transform [57] with the mutual information (MI)[52, 53, 62] as the similarity metric to assess the alignment quality. MI is highly effective for multi-modal image registration or even in mono-modal cases when high differences in the intensity distribution between the two images are present, as is the case in the present study. The adaptive stochastic gradient descent optimizer [94] was used to achieve the best similarity measure through optimizing the B-spline transform parameters. Instead of using all image pixels in the registration, as it is computationally expensive for such large high-resolution images, a random subset of pixels was chosen [56, 68], to limit the computational cost. We used a random coordinate sampler [95] to select a subset of 5,000 pixels and then applied a multi-resolution approach to enhance registration accuracy.

A crucial step in non-linear registration using the B-Spline algorithm is defining the spacing between the B-Spline control points. The control points are used to capture the local deformation in the image, and thus the spacing between these points determines how accurately the algorithm can correct for local deformations. A mouse brain histological image is a structure-rich environment with high variations in both the size of distinct brain regions [91] and local deformations. It was found that just one control-point-spacing was not sufficient to model the non-linear deformations. Instead, we used a multi-resolution scheme with eight different control-point-spacings; control points were spaced 15 pixels on the finer grid and linearly increased (multiples of two) to model the larger deformations.

Figure 2.3 shows the result of the registration process. The processed image (background removed, smoothed, and cropped) and the best match from the Allen Brain Atlas are shown in a) and b), respectively. It can be seen that prior to the registration the sample image contains significant global deformations (translation, rotation and scale) as well as local distortions such as folding. The affine transform allows the sample image to be oriented and scaled to the Allen Brain Atlas target image while the non-linear B-spline transform compensates for local distortions (Figure 2.3.c)). A checkerboard image (Figure 2.3.d)), in which alternate tiles show the two sections, allows
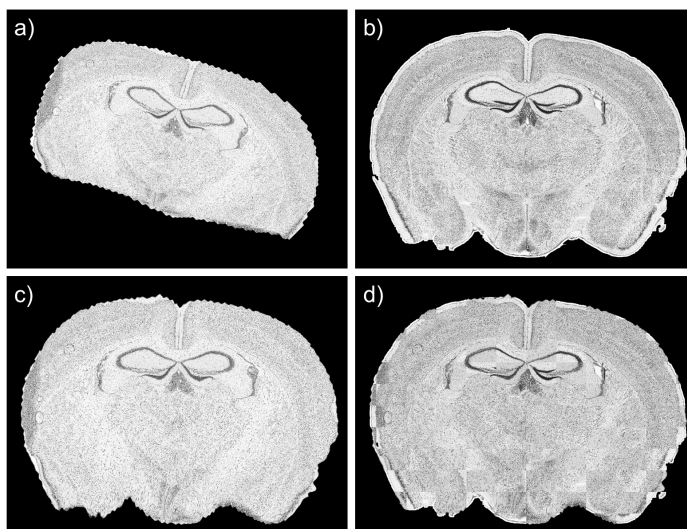
Figure 2.3: Image registration. a) The original sample image after pre-processing. b) The automatically selected target image from the ABA that represents the best match to the sample image shown in a). c) The original image after affine and non-linear registration. d) The ABA target image in b) and the registered sample image in c) are combined to form the checkerboard image to assess the registration result visually.

the comparison of the registered sample image with the automatically identified target image, demonstrating the absence of discontinuities and thus the high quality of the registration procedure.

To assess whether the initial selection of three Allen Brain Atlas reference images, based only on the hippocampal formation, biased the registration we compared the results with those obtained by registering the processed image to all of the Allen Brain Atlas reference images within the ROI. As the entire search space is considered in the latter approach, the selected best reference image is considered as the ground truth. Eight coronal sections from eight mice were used for the test: in all cases both registration approaches selected the same reference image but the method involving the prior selection of three reference images was 4X faster.

2.3.4. REGISTRATION OF MS-IMAGING DATA WITH THE ALLEN BRAIN ATLAS

It is now an established practice in the MSI field to co-register the MSI data with the histological image. Figure 2.4.a–c show the distribution of mass m/z = 863 overlain on the histological images of mice M1, M2, and M3, respectively. As can be seen, each histological image shows different degrees of deformation and different artifacts, such as folding and tearing. Once the registration of the histological image to the Allen Brain Atlas is complete the MSI data is aligned by applying the same transformation matrix, but corrected for the lower spatial resolution of the MSI images, Figure 2.4.g–l. Finally, the three different MS-Imaging datasets were registered to the same coordinate space. In this case the best matching image from the Allen Brain Atlas was found to be the same

for all the datasets. Figure 2.4.m–o show the spatial distribution of mass m/z = 863 from the three different mouse brains mapped onto the Allen Brain Atlas. The distribution of this mass in the three mouse brains shows a high degree of correlation as can be seen by the Pearson correlation values: [M1:M2] = 0.82; [M1:M3] = 0.78; [M3:M2] = 0.81. These correlation coefficients are similar to those previously reported for MSI datasets from adjacent tissue sections (in this case from rat brain)[96].
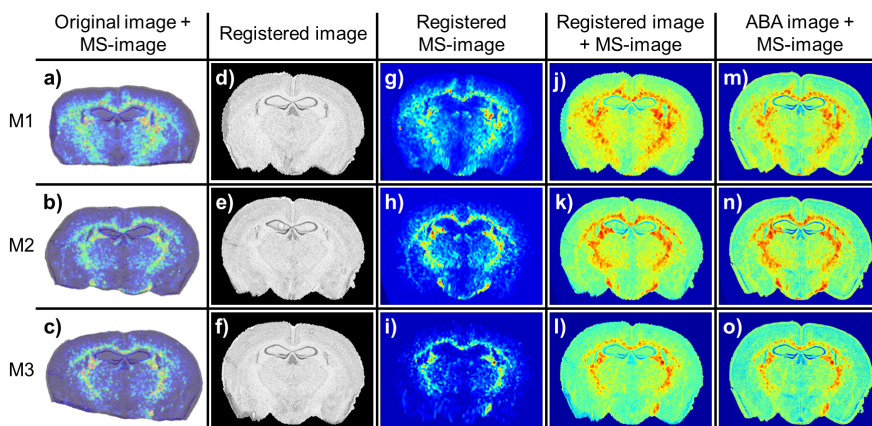

Figure 2.4: Registration of multiple MS-imaging data with the Allen Brain Atlas: three coronal sections from three different mouse brains (M1, M2 and M3), and the respective MS images, were processed and co-registered to the ABA. a), b) and c): distribution of m/z = 863 in the original sample images from mouse M1, M2 and M3 before pre-processing and registration. d), e) and f): the original sample images from mouse M1, M2 and M3 after pre-processing and registration. g), h) and i): the distribution of m/z = 863 from mouse M1, M2 and M3 after registration of the MS-imaging data. j), k) and l): the registered images and the registered MS-imaging datasets from mouse M1, M2 and M3 were superimposed to visualize the distribution of ion m/z = 863. m), n) and o): the registered MS-imaging datasets from mouse M1, M2 and M3 were aligned with the best match from the ABA to visualize the distribution of ion m/z = 863 in a common reference system.

The average total computational time was 8.7 min. on a PC (2.66 GHz Intel Xeon and 6 GB memory). The process for automatic selection of the best reference section among those 14 candidate coronal sections from the ABA took approximately 1.74 min. (~0.3 min for the initial localization followed by ~1.44 min. for applying an affine registration on three sections). The non-linear registration took in total approximately 6.83 min. using the multi-resolution image registration approach with eight resolution levels.

2.3.5. APPLICATION TO ANIMAL COHORTS

Experiments involving animal models typically compare a number of animals per test group (either involving specific genotypes, in case of disease animal models, or specific interventions) in order to assess the statistical significance of an experimental outcome. When applying MSI to such animal cohorts it is important that the tissue sections from each animal are from similar regions of the brain. Figure 2.5.a) shows the distribution of the location of the coronal tissue sections in the mouse brain for a cohort of approximately 32 animals (multiple sections per animal: total 94 sections). The majority

of the coronal tissue sections, cut by expert neurologists, were located within 0.25 mm of each other. Nevertheless the tissues still exhibited variation in animal brain size, brain region size, and tissue processing artifacts. To quantitatively assess the performance of the registration the distance between a series of control points was determined at each step of the registration process. Figure 2.5.b) shows that before registration there were large differences as well as significant variability in the positions of the tissues and their internal structures. However, after the affine (rigid) registration the differences and variability were reduced significantly, although still present. Notably, after the B-spline registration the median error was 10.6 pixels, corresponding to approximately 30 μm, which is significantly less than the spatial resolution of the MALDI MSI experiments reported here.
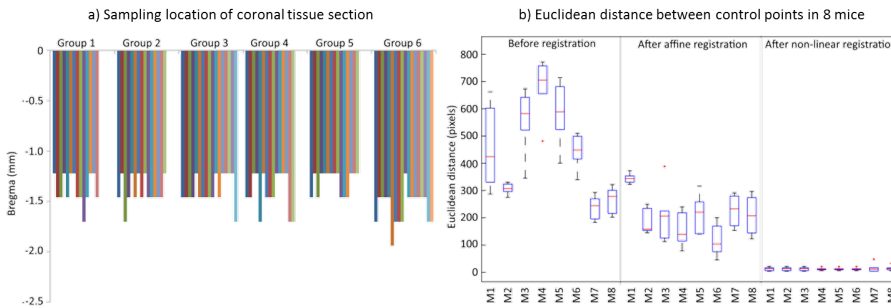

Figure 2.5: Application of Allen Brain Atlas registration algorithm to large animal series, quantification of alignment. A) Position of coronal tissue sections sampled from an animal series, demonstrating the majority were sampled within 0.25 mm of each other. B) Euclidian distance to eight animals before, after the rigid affine registration and after the non-rigid registration. Note: The tissue sections selected for comparison previously determined to have been sampled at the same location.

2.3.6. COMPARISON OF MSI DATA WITH BRAIN ANATOMY
Mapping the MALDI MSI data of different animals to the Allen Brain Atlas makes it possible to use the anatomical reference atlas to query the data. A custom Matlab script was created that takes as an input the string name of any anatomical region defined in the Allen Brain Atlas and then generates the corresponding anatomical binary mask. As an example the four histo-anatomical regions hippocampus, thalamus, isocortex, and striatum are shown in Figure 2.6.a. These image masks may be used to extract region-specific mass spectra or used to identify molecular ions that exhibit a distribution localized (correlated) to the selected region, Figures 2.6.b and 2.6.c respectively.

2.4. DISCUSSION
The mouse brain is arguably the most-used model of human neurological diseases [74, 75]. The comparison of animal cohorts is extensively used in pre-clinical investigations aimed at identifying relevant biomarkers through the comparison of MS profiles in different conditions [79]. Yet, as recently reported by Hager et al., there is very little correlation between the overall size of the brain and the different brain regions [91].
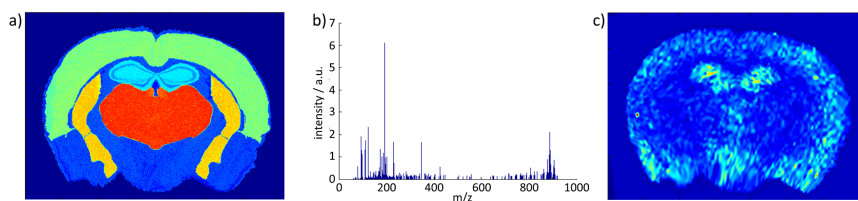
Figure 2.6: Comparison of MSI data with the brain anatomy. a) Registered histological image from mouse M1 brain section overlain with the anatomical binary masks defined in the Allen Brain Atlas for the hippocampus (light blue), isocortex (green), thalamus (red) and striatum (orange). b) The extracted average mass spectrum for the hippocampus. c) The distribution of molecular ion m/z = 175 showing high correlation with the selected anatomical region (hippocampus).

In addition to the intrinsic biological variation, the shape of a mouse brain section and brain structures is also influenced by experimental conditions, such as the type of experiment, the method used for brain excision and its immediate processing (i.e. dry ice, liquid nitrogen, cold isopentane, or heat-stabilization), tissue sectioning and mounting the sections on glass slides [97]. In order to establish a reliable comparison of animal cohorts the data from different animals should be placed onto the same coordinate space. Previous studies have reported alignment tools for the generation of 3D-MSI datasets [87–89]. However, these tools only contemplate the alignment of MSI data from consecutive tissue sections from a single animal and do not account for the high degree of inter animal variability. To overcome this limitation we propose a novel approach that enables the comparison of MSI data sets from multiple animals by automatically co-registering them to a single coordinate system, the Allen Brain Atlas, a publicly available web platform (http://www.brain-map.org/).

We used the anatomical features present in the high-resolution histological images from the MS-analyzed brain sections to register the MSI datasets. Thus, the alignment is not biased by the quality of the MSI data and is applicable to MSI datasets obtained using any ionization method and mass spectrometer. The only prerequisite is that the tissue section's MSI data is aligned to its histological image (standard practice in modern MSI). The proposed automatic registration pipeline consists of three key steps: i) image preprocessing of histological images; i) automatic selection of target image from the Allen Brain Atlas; iii) image registration.

Preprocessing of the histological images is necessary to correct for image artifacts that may derive from the staining or the image acquisition process and ensures that images acquired with different microscope scanners have sufficient brightness and contrast for optimal comparison. The high-resolution histological images were also down-sampled to speed up the registration pipeline. Reducing the image size decreased the processing time, which, especially in large scale studies, will have a significant impact not only on the overall processing time and but also on data size.

The Allen Brain Atlas contains 132 coronal sections uniformly spaced at 100 μm covering the entire mouse brain. Using the entire range would significantly increase the processing times and complicate the registration pipeline. Instead we propose to use the experimental design to define a region of interest, in this example starting at bregma -1.22 mm and finishing at -2.80 mm, which contains only 14 coronal sections. Within this

region of interest the hippocampus is highly prominent and provided a readily calculable metric to estimate the location of the experimental tissue sections within the Allen Brain Atlas. Alternative brain regions that show clear and distinctive shapes, such as the lateral ventricles or the medial habenula, could have also been used to find the best ABA match to the experimental section. The hippocampus was chosen owing to its ease of isolation and the clear linear relationship with the frame number of the ABA sections.

The final non-rigid registration step used the full histological image information of the experimental and ABA images. The pre-selection of ABA images using the hippocampus greatly simplified and sped up the registration process by limiting the full calculation to just three reference images. The similar anatomies of the ABA images and the experimental images implicitly lead to only small image-alignment transformations. The non-rigid registration is also able to compensate for minor tissue processing artifacts such as tears. As an example, we used MSI datasets from different mice, each presenting different artifacts and a different degree of deformation (Figure 2.4.a-c). After registration an increase in similarity between the different animal sections is readily observed, Figure 2.4.d–f) and Figure 2.5. Nevertheless it remains important to ensure the tissue sections are of high quality, as tissue processing artifacts such as tears and folds may affect the mass spectral data.

APPLICATION

Manual annotation by visual inspection of stained tissues remains the standard procedure when histology/anatomy is used to guide MSI data analysis. MSI is usually implemented within mass spectrometry laboratories and so such annotation is the main bottleneck for its biomedical application, particularly so when the pre-clinical study requires the analysis of a large number of samples. The development of automatic alignment and annotation tools capable of processing many samples will not only increase the throughput of data analysis, but also facilitate the comparison of MSI data from animal cohorts; i) by extracting data from specific brain regions; ii) by comparing the MSI distributions in different animals using the same coordinate space; and iii) and by enabling the correlation of large MSI datasets with, for instance, the anatomical or gene expression information available in the Allen Brain Atlas.

A number of multivariate and cluster analysis methods have been applied to MSI data to identify regions that have correlated mass spectral profiles. Alexandrov and workers [31] have demonstrated that with suitable parameterization the resulting segmentation map of the MSI data can resemble the anatomy of the analyzed tissue sections. Such segmentation of MSI data from pathological tissues has been used to highlight areas with distinct mass spectral profiles (so called molecular histology) but the results must be compared with the tissue's histology/anatomy to assess how the segmentation maps (and associated mass spectral profiles) relate to the disease.

Many brain disorders are known to lead to pathological changes in specific regions of the brain. For example amyloid plaques in the hippocampus of Alzheimer's disease patients, Parkinsonian Lewy bodies in the substantia nigra, and cortical spreading depression in migraine with aura. For each of these examples MALDI MSI has revealed biomolecular changes in the associated regions in rodent models of the disease [7, 76–80, 83]. The ability to register MALDI MSI datasets from different animals onto the Allen

Brain Atlas reference images enables the signatures from specific brain regions to be readily extracted (Figure 2.6), and thus to target those regions known to be affected by disease. Such a region-specific analysis avoids the significant variation in MS signatures associated with the different regions present in the brain (the segmentation maps that match the brain anatomy reflect the different MS signatures) and so may be more effective at identifying molecular changes associated with disease.

The present work is focused on the alignment of coronal tissue sections from well-matched animals. The male, 3-month old C57BL/6J mice used here are near identical to the male, 56-day old C57BL/6J mice used in the Allen Brain Atlas [69]. If using animals of different strain, different age, or different sex potential differences in their anatomy must first be considered before they can be registered to the Allen Brain Atlas. In case of significant differences in anatomy then a new reference atlas will be necessary (i.e. register all animals to a single reference animal). The current work also does not correct for significant differences in the angles of the coronal plane; such differences would undermine the ability to draw biological conclusions from the MSI experiments and so the standard operating procedure used here goes to great lengths to produce coronal sections in the coronal plane and at the same location (see Figure 2.5 for the high reproducibility).

Finally while the current work has focused on MSI the same registration process could be used to register LC-MS-based voxelization experiments [98, 99] to the Allen Brain Atlas. In voxelization experiments the brain is dissected into small cuboids or voxels (e.g. 1 mm^3), each of which is subject to label-free protein quantitation using LC-MS analysis. Using this approach Petyuk *et al.* have made low spatial resolution maps of protein expression for approximately one thousand proteins [98, 99] and compared the results with the gene expression data contained in the Allen Brain Atlas. The registration tool reported here could make this comparison explicit. The registration tool could also be developed to identify identical voxels in different animals, thereby enabling them to be pooled for deeper proteome coverage.

2.5. CONCLUDING REMARKS

We propose an image-processing pipeline to automatically map the biomolecular structures provided by MSI from different mouse brains to the Allen Brain Atlas. This enables facile and rapid inter-animal comparisons, quantification of regional biomolecular content, and in the future correlation of the MSI data to the gene-expression image data provided by the ABA.