Lab-on-a-tissue : optimization of on-tissue chemistry for improved mass spectrometry imaging

Heijs, B.P.A.M.

Chapter 3

Comprehensive analysis of the mouse brain proteome sampled in mass spectrometry imaging

Bram Heijs; Ricardo J. Carreira; Else A. Tolner; Arnoud H. de Ru; Arn M.J.M. van den Maagdenberg; Peter A. van Veelen; Liam A. McDonnell, "Comprehensive analysis of the mouse brain proteome sampled in mass spectrometry imaging," *Analytical Chemistry* 87(3), pp. 1867-1875, 2015.

On-tissue enzymatic digestion is performed in MSI experiments to access larger proteins and to assign protein identities. Most on-tissue digestion MSI studies have focused on method development rather than identifying the molecular features observed. Herein, we report a comprehensive study of the mouse brain proteome sampled by MSI. Using complementary proteases, we were able to identify 5,337 peptides in the MALDI matrix, corresponding to 1,198 proteins. 630 of these peptides, corresponding to 280 proteins, could be assigned to peaks in MSI data sets. Gene ontology and pathway analyses revealed that many of the proteins are involved in neurodegenerative disorders, such as Alzheimer's, Parkinson's and Huntington's disease.

3.1 Introduction

Matrix-assisted laser desorption/ionization (MALDI) mass spectrometry imaging (MSI) is an analytical technique in which the distributions of hundreds of biomolecular ions can be recorded directly from tissue sections (Caprioli et al., 1997). It allows the simultaneous and untargeted investigation of many molecular classes, including pharmaceuticals, metabolites, lipids, peptides, and proteins (McDonnell and Heeren, 2007). One of the principal application areas of MALDI-MSI is the investigation of the molecular content of pathological tissue samples in order to find biomarkers or provide insights into the molecular mechanisms underlying a disorder. The discovery of putative biomarkers has often focused on proteins since the results can then be independently validated using immunohistochemistry (Balluff et al., 2010; Cazares et al., 2011; Meding et al., 2012).

To increase proteome coverage, aid protein identification, and enable the analysis of formalin-fixed paraffin embedded (FFPE) tissues, the enzymatic protein digestion performed in classic bottom-up proteomics has been adapted for MALDI-MSI (Groseclose et al., 2008; Lemaire et al., 2006; Shimma et al., 2006). Protein identification directly from the tissue section can be performed by MS/MS analysis of the proteolytic peptides but remains challenging due to the very high complexity of the peptide mixture generated after digestion (Houel et al., 2010). Instead the peptide identities are commonly assigned to those previously identified via extraction of proteolytic peptides followed by LC-MS/MS analysis (Maier et al., 2013; Schober et al., 2011).

As in bottom-up LC-MS/MS based proteomics, on-tissue digestion is most often performed using trypsin; its cleavage specificity, C-terminal of arginine and lysine, results in proteolytic peptides that have an intrinsic positive charge on the C-terminus, thus enhancing their detection by positive-ion MS (Brownridge and Beynon, 2011; Vandermarliere et al., 2013). A crucial difference is that MALDI-MS is biased towards the detection of Arg-terminated peptides (Krause et al.,1999). The suppression of Lys C-terminated tryptic peptides in MALDI-MS results in an undesired loss of sequence

information and proteome coverage. This is likely to be exacerbated in MALDI-MSI because of the absence of any explicit peptide separation.

The use of multiple proteases has been shown to have a beneficial effect on proteome and sequence coverage in both LC-ESI-MS and LC-MALDI-MS based studies (Choudhary et al., 2003; Gatlin et al., 2000; Hohmann et al., 2009; Wa et al., 2006). A recent study indicated that similar complementarities might be obtained in on-tissue digestion MALDI-MSI experiments: a combination of trypsin, pepsin and elastase was shown to have a positive effect on protein sequence coverage (Enthaler et al., 2013b). However since pepsin and elastase do not cleave basic amino acids many of the proteolytic peptides will not be detected with high sensitivity. To maintain high detection sensitivity it is essential that the proteases cleave the most basic amino acids, i.e. Lys and Arg. While these are the same residues cleaved by trypsin, previous LC-MS/MS results have already demonstrated that the different sequence and conformation specificities of Lys-C, Arg-C and Lys-N can improve proteome and sequence coverage by reducing the number of missed cleavages (Brownridge and Beynon, 2011).

Here we have investigated the degree of increased proteome coverage that may be obtained in MALDI-MSI by using a similar complementary cohort of enzymes, specifically (i) trypsin, (ii) Lys-C, (iii) recombinant Lys-N (r-Lys-N), (iv) Arg-C, and (v) a mixture of trypsin and Lys-C (trypsin/Lys-C).

3.2 Methodology

3.2.1 Chemicals and reagents

All chemicals were purchased from Sigma-Aldrich (Steinheim, Germany) except ethanol (Merck, Darmstadt, Germany). All proteases were purchased from Promega (Madison, USA) except r-Lys-N (U-Protein Express BV, Utrecht, The Netherlands).

3.2.2 Sample collection

Three-month-old, male, C57BL/6J mice were sacrificed by cervical dislocation. The brains were excised, flash-frozen on dry-ice and stored at -80°C until analysis. Twelve-µm thick coronal tissue sections were obtained using a cryostat microtome (Leica Microsystems, Wetzlar, Germany) at -12°C. The sections were 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.

3.2.3 Tissue preparation

The tissue sections mounted on indium tin oxide (ITO) coated slides were collected from -80°C storage and equilibrated to room temperature in a freeze dryer for 30 minutes. Thereafter, all tissues were washed as follows: (i) submerge in 70% ethanol for 30 s; (ii) submerge in 96% ethanol for 30 s; (iii) 10 short dips in deionized water;

(iv) submerge in 70% ethanol for 30 s; (v) submerge in 96% ethanol for 30 s (Enthaler et al., 2013a). Finally, the sections were dried in a vacuum desiccator for 15 min.

3.2.4 On-tissue disulphide bond reduction

Tissue sections were covered with 4 layers of 2 mM tris(2-carboxyethyl)phosphine (TCEP) (in deionized water) using the SunCollect automatic sprayer (SunChrom, Friedrichsdorf, Germany). The flow rate for the TCEP application was set to 10 µL/min, which resulted in the application of 360 nmol TCEP/cm^2. Detailed information on the SunCollect settings can be found in Table S3-1 (Supplementary information). The total incubation time at 23°C (room temperature) was 30 min (including ca. 10 min spraying time).

3.2.5 On-tissue enzymatic digestion

All proteases were dissolved in deionized water to a final concentration of 0.02 µg/µL. 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) containing enzyme buffers for Lys-C and r-Lys-N were exchanged with deionized water using a 10 kDa molecular weight cut off (MWCO) spin filter (Merck Millipore, Billerica, MA, USA). Arg-C was dissolved in 2 mM dithiothreitol (DTT) to activate the enzyme. Five layers of proteolytic enzymes were applied using the SunCollect automatic sprayer at a flow rate of 5 µL/min. This resulted in the application of 7.5 µg enzyme/cm^2 (Table S3-1, Supplementary information). The tissues were incubated for 18 hours at 37°C in a saturated air chamber (50% methanol in deionized water). Finally, MALDI matrix was applied using the SunCollect sprayer. For the MALDI-TOF-MSI analyses 5 mg/mL α-cyano-4-hydroxycinnamic acid (CHCA) in 50% acetonitrile (ACN) and 0.3% trifluoroacetic acid (TFA) was applied. For MALDI-MSI experiments performed with the MALDI-FTICR mass spectrometer it was found that the CHCA matrix produced excessive matrix clusters. Accordingly for MALDI-FTICR-MSI analyses, 50 mg/mL 2,5-dihydroxybenzoic acid (DHB) in 50% ACN and 0.1% TFA was used as matrix.

3.2.6 MALDI-TOF imaging

MALDI-TOF-MSI was performed on an UltrafleXtreme MALDI-TOF/TOF (Bruker Daltonics) in positive-ion reflectron mode, using 500 laser shots per spot and 100 x 100 µm pixel size. Data was acquired in a *m/z* range from 800-3,000 Da. Data acquisition, pre-processing and visualization were performed using the flex software package by Bruker Daltonics (flexControl 3.4; flexAnalysis 3.4; flexImaging 3.0).

3.2.7 MALDI-FTICR imaging

MALDI-FTICR-MSI was performed on a 9.4 T SolariX XR mass spectrometer (Bruker Daltonics) in positive-ion mode, using 250 laser shots per spot and 150 x 150 µm pixel size. Data was acquired in a *m/z* range from 600-3,500 Da with a 512k datapoint transient (1.1 s duration) and an estimated resolution of 200,000 at *m/z* 400 Da. Data acquisition was performed using ftmsControl (Bruker Daltonics) and visualizations were obtained from flexImaging 4.0 (Bruker Daltonics).

3.2.8 Data analysis MALDI-TOF-MSI

Regions of interest (ROIs) containing the full area of the coronal section were selected in flexImaging and 500 random spectra from within each ROI extracted into ClinProTools 3.0 (build 22, Bruker Daltonics). The spectra underwent smoothing and baseline subtraction (Top Hat algorithm), total-ion-count normalization, and peak picking at different signal-to-noise ratios. The peak lists were exported to Excel 2010 for further calculations.

3.2.9 Peptide extraction

Consecutive tissue sections were prepared for MALDI-MSI using the same sample preparation protocol. Instead of MALDI-MSI analysis, the proteolytic peptides were extracted from the matrix coating using a series of solvents with increasing organic content: (i) 10 µL of 0.1% TFA (repeat 4 times); (ii) 10 µL of 50% ACN / 0.1%TFA (repeat 4 times); (iii) 90% ACN / 0.1% TFA (repeat 4 times) (Enthaler et al., 2013a). Extracts were combined, dried and resuspended in 0.1% TFA, and then cleaned with Omix C18 tips (Agilent). The purified extracts were dried and stored at -20°C until LC-MS/MS analysis.

3.2.10 LC-MS/MS analysis

Peptide extracts were analyzed using an Easy nLC1000 (Thermo, Bremen, Germany) coupled to a Q-Exactive mass spectrometer (Thermo). Fractions were injected onto a homemade pre-column (100 µm × 15 mm; Reprosil-Pur C18-AQ 3 µm, Dr. Maisch, Ammerbuch, Germany) and eluted via a homemade analytical column (15 cm × 50 µm; Reprosil-Pur C18-AQ 3 µm). The gradient was 0% to 30% solvent B (90% ACN / 0.1% formic acid (FA)) in 120 min. The analytical column was drawn to a tip of ~5 µm and acted as the electrospray needle of the MS source. The Q-Exactive mass spectrometer was operated in top10-mode. Parameters were: Full scan - resolution 17,500; AGC target 3,000,000; max fill time 20 ms; MS/MS - resolution 35,000; AGC target 1,000,000; max fill time 60 ms; intensity threshold 17,400; Apex trigger was set to 1-5 seconds, and allowed charges were 1-5.

3.2.11 LC-ESI-MS/MS database search

Peptide and protein identifications were extracted from the SwissProt database using the Mascot server. Up to two missed cleavages were allowed and methionine oxidation was set as a variable modification. Peptide assignments were made with a tolerance of 10 ppm. MS/MS fragment tolerance was 20 mmu. Protein identifications were assigned based on a minimum of one confident peptide at 1% false discovery rate (FDR). Only peptides and proteins with red bold notification were included in the final lists of non-redundant peptides. For trypsin and trypsin/Lys-C, peptides and proteins with a Mascot significance score ≥ 30 were included, for Lys-C, r-Lys-N and Arg-C the threshold score was ≥ 25.

3.2.12 Gene ontology and pathway analysis

A list containing the Uniprot accession numbers of the identified proteins was uploaded into the STRAP software tool (v1.1.0.0, Boston University School of Medicine) (Bhatia et al., 2009). The KEGG pathway analysis (PA) was performed by uploading the list in the online STRING 9.1 tool (Jensen et al., 2009). As p-value correction, the "FDR correction" option was applied.

3.2.13 Data analysis MALDI-FTICR-MSI

Average spectra from the MALDI-FTICR-MSI datasets were exported into CSV format and loaded into mMass (http://www.mmass.org) (Strohalm et al., 2008). Peak picking was performed on peaks with S/N ≥ 5 and intensity ≥ 1,500 A.U.. The peak lists were then deisotoped with a maximum charge of 3+ and an isotope mass tolerance of 0.05 Da.

3.2.14 MALDI-FTICR-MSI peptide identity assignment

MALDI-FTICR-MSI peak lists and Mascot database search results from the LC-MS/MS analyses were exported to Excel 2010. The MSI peaks were assigned to identified peptides based on a mass tolerance of 20 ppm.

3.3 Results

On-tissue digestion is used to increase proteome coverage and to assign identities to the peaks detected by MALDI-MSI. In this study we investigated whether additional arginine and lysine proteolytic enzymes, with different cleavage specificities, could further improve on-tissue digestion MALDI-MSI by increasing sequence and proteome coverage. MALDI-MSI was combined with LC-MS/MS of on-tissue digestion matrix extracts to compare the accessible proteolytic peptides produced by the enzymes trypsin, Lys-C, trypsin/Lys-C, r-Lys-N and Arg-C, and to determine which proteolytic peptides were detected by MALDI-MSI.

3.3.1 MALDI-TOF-MSI: image comparison

To assess the effects of sample preparation (online Supplementary information) for the different enzymes, ion distributions for proteolytic peptides from the same protein were compared with both each other and the gene expression data from the Allen Mouse Brain Atlas (http://www.brain-map.org). Figure 3-1 shows example mass spectra and MSI images obtained from coronal sections of a C57BL/6J mouse brain subject to on-tissue digestion with the different proteases. The images are of proteolytic peptides originating from myelin basic protein (MBP), assigned to confidently identified peptides from the LC-MS/MS analysis of matrix extracts. The mass resolution of the UltrafleXtreme MALDI-TOF/TOF platform used for these MSI experiments is insufficient to resolve the isobaric ions produced by on-tissue digestion, and so the images may include contributions from unresolved isobaric ions. The distributions of the MBP

proteolytic peptides were consistent with the MBP gene expression images contained in the Allen Brain Atlas, Figure 3-1F, indicating that any isobaric ions were minor contributors and that no artifacts resulted from the on-tissue digestion procedure.
It is important to note that the quality of the MSI images was poor when the enzymes were applied using the buffer system recommended by the manufacturers (e.g. ammonium bicarbonate; HEPES; Tris-HCl), with no visualization of clear anatomical structures for different *m/z* values (Figure S3-1, Supplementary information). For this reason, all enzymes, except Arg-C, were dissolved in milliQ water (pH 6.5 – 7).

3.3.2 MALDI-TOF-MSI: comparison of average spectra

The average mass spectra obtained by MALDI-MSI analysis for both trypsin (Figure 3-1A) and trypsin/Lys-C (Figure 3-1B) digestions presented a curved mass spectral profile that has already been described in several publications (Groseclose et al., 2008; Lemaire et al., 2006; Shimma et al., 2006). The profile reflects the very high complexity of the resulting proteolytic peptide mixtures as well as the inability of the reflectron-TOF mass analyzer to fully resolve all peaks. Proteolytic digestions using proteases with cleavage specificity for single amino acids, such as Lys-C, r-Lys-N or Arg-C, yielded a smaller number of peptides. As can be seen in Figure 3-1C-E, when Lys-C, r-Lys-N or Arg-C were used, the spectra exhibited a flatter baseline and peaks with higher absolute intensity. Figure 3-1G shows that digestion with trypsin yielded a higher total number of *m/z* features than digestions using proteases with single-site cleavage specificity. However, on-tissue digestion with Arg-C presented the largest number of high intensity *m/z* features, S/N > 20, consistent with the previously reported bias of MALDI towards Arg-terminated tryptic peptides. Note: the number of *m/z* features reported in Figure 3-1G should be considered a low estimate because of the inability of the reflectron-TOF mass analyzer to resolve isobaric ions (Schober et al., 2012).

3.3.3 Peptide and protein identification

To further investigate the dependence of the on-tissue digestion MALDI-MSI experiment on the protease, the proteolytic peptides extracted from the tissue section, following deposition and crystallization of the matrix, were analyzed by LC-MS/MS (using consecutive tissue sections that had undergone identical sample preparation). The analysis of the matrix extracts of the five protease preparations resulted in a total of 5,050 non-redundant peptides (online Supplementary information), over 2.8 times the number obtained from the matrix extract of the trypsin digested tissue. When including redundant peptides, the total number increased to 5,339 (online Supplementary information), which was also 2.8 times the number of peptides identified with trypsin alone (Figure 3-2A). The relatively low number of peptides identified after digestion with Lys-C and r-Lys-N was unexpected, but confirmed by repeat experiments. This observation is due to the enzyme buffer exchange performed

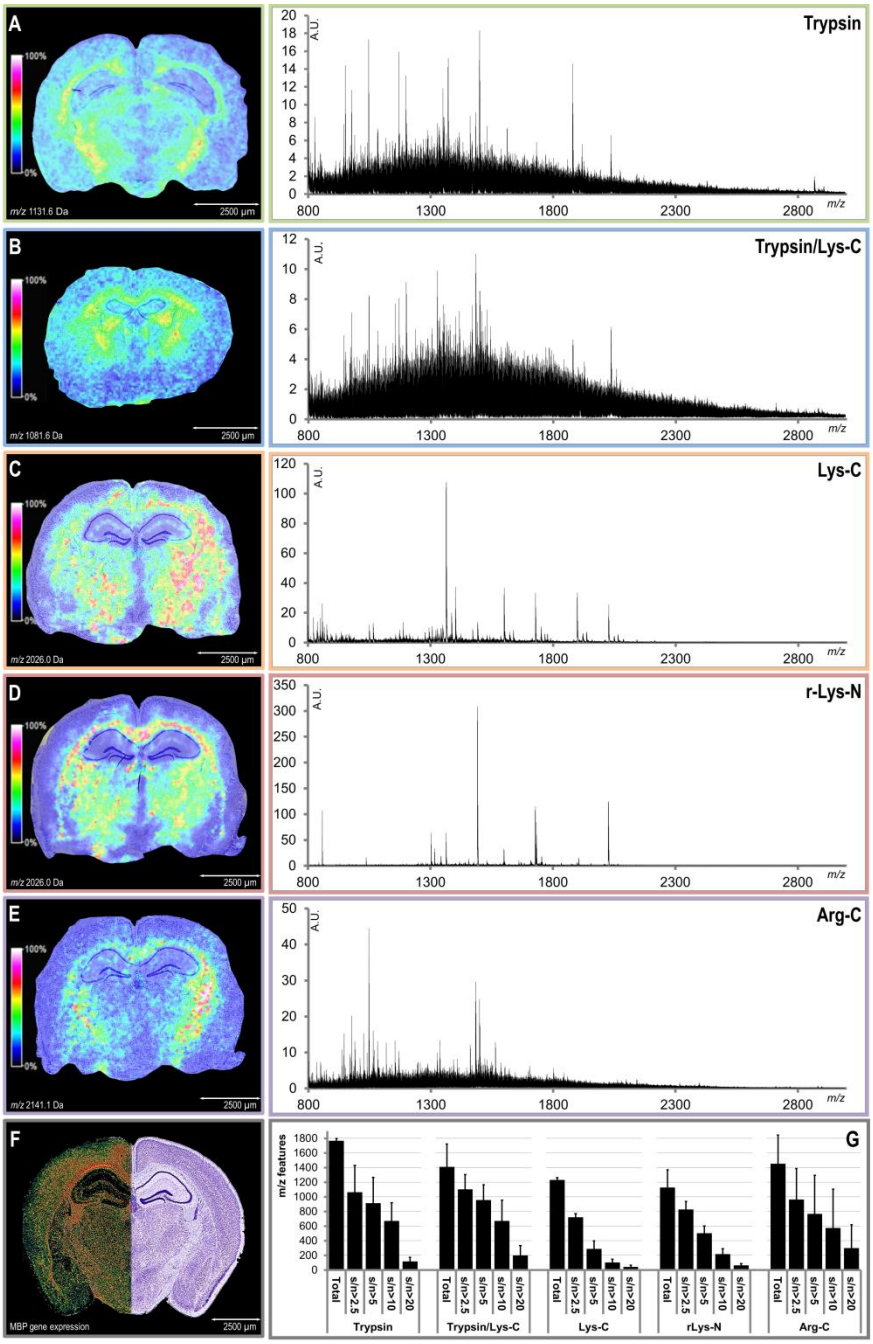


Figure 3-1: Overview of the on-tissue-digestion MALDI-MSI data. Myelin basic protein peptide ion distributions and average mass spectra for the digested tissue sections; **(A)** *m/z* = 1,131.6 (TTHYGSLPQK); **(B)** *m/z* = 1,081.6 (FFSGDRGAPK); **(C)** *m/z* = 2,026.0 (SQHGRTQDENPVVHFFK); **(D)** *m/z* = 2,026.0 (KSQHGRTQDENPVVHFF); **(E)** *m/z* = 2,141.1 (TQDENPVVHFFKNIVTPR); **(F)** mouse brain Nissl histologic staining and MBP gene expression data from the Allen Brain Atlas; **(G)** average number (n = 4) of extracted *m/z* features with different signal to noise ratios.

for both enzymes. To achieve optimal digestion conditions proteases are usually dissolved in specific buffer systems containing additives, such as chelating agents, salts and sugars. The buffers for both Lys-C and r-Lys-N contained HEPES, a compound known to cause signal suppression in MALDI-MS above a certain concentration threshold (Amini et al., 2000; Signor and Erba, 2013). This effect is demonstrated in Figure S3-1 (Supplementary information), which compares the average spectra obtained after on-tissue digestion MALDI-MSI with Lys-C dissolved in the manufacturer's recommended buffer (50 mM HEPES (pH 8.0), 10 mM ethylenediaminetetraacetic acid (EDTA) and 5 mg/mL raffinose), and the average spectra obtained after digestion with the same enzyme dissolved in MQ-water after buffer exchange. The latter spectrum shows an increase of more than 2-fold in absolute intensity but the number of peptides is significantly less than would be expected on the basis of previous LC-MS/MS experiments.

Similar results were achieved at the protein level (Figure 3-2B). The combined data of all on-tissue digestion experiments resulted in a total of 1,194 confidently identified non-redundant proteins. When redundant proteins where included, a total of 1,198 proteins was identified of which 52.2% had not been previously identified in the matrix extract of the trypsin-digested tissue section. Thirty-seven proteins, including redundant proteins, were identified exclusively in the tryptic digests, corresponding to 3.1% of the total number of identified proteins. The largest gain in proteome coverage was achieved using Arg-C: 22.7% of the total number of identified proteins originated exclusively from this extract. A subset of only 96 proteins was identified in all matrix extracts, which corresponds to 8.0% of the total number of identified proteins.

Besides the increase in proteome coverage, individual protein sequence coverage also gained from combining the data obtained with multiple proteases (Figure 3-2C). For instance, analysis of spectrin-1 (SPTN1; 2,472 amino acid residues; MW = 284 kDa; Uniprot entry: P16546), a highly abundant protein in mouse brain that is involved cytoskeletal structure (Zagon et al., 1986), resulted in a sequence coverage of 26.3% from the matrix extract of the trypsin digested tissue section. By combining the data from all enzymes the sequence coverage for SPTN1 increased to 44%. Similar trends were observed for other proteins as shown in Figure 3-2C.

To test whether the proteome coverage could be improved by reducing protein disulfide bonds before enzymatic digestion, a series of experiments were performed in which on-tissue enzymatic digestion was preceded by on-tissue disulfide bond reduction using 2 mM TCEP (in milliQ water), a compound that was previously shown to be compatible with MALDI-MS (Fischer et al., 1993; Kulak et al., 2014). The ion distributions from the reduced tissues were of similar quality as the ones obtained from non-reduced tissues, indicating that the TCEP reduction did not affect the localization of the peptides in the tissue (data not shown). The gain in proteome coverage for each

A

Significantly identified peptides		
Protease	*Non-redundant peptides (#)*	*Incl. redundant peptides (#)*
Trypsin	1756	1913
Trypsin/Lys-C	2284	2467
Lys-C	383	386
rLys-N	1046	1068
Arg-C	1596	1646
Combined	5050	5337

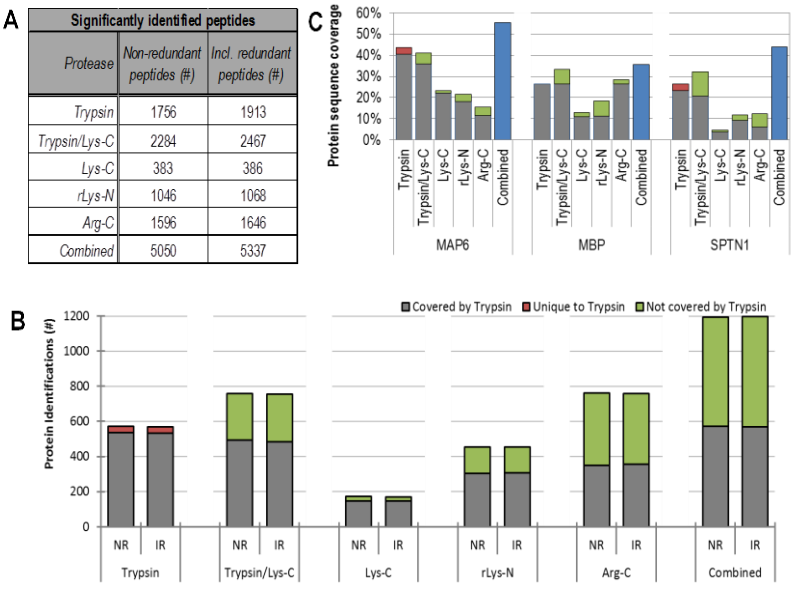
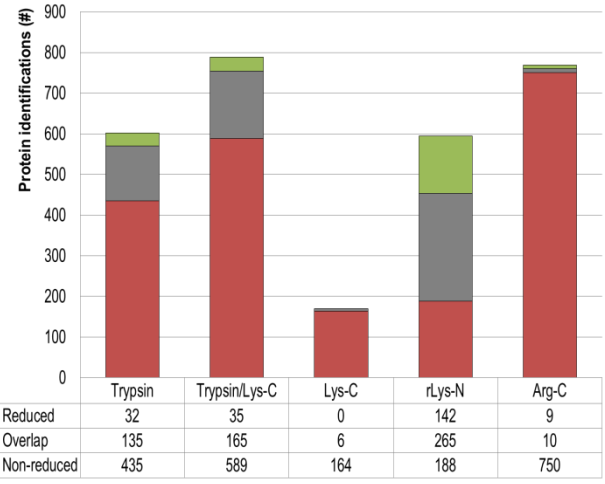


Figure 3-2: Analysis of the on-tissue digested matrix proteome. **(A)** Number of peptides identified for each protease above the significant threshold. Confident non-redundant peptides required a red bold notation in Mascot and a 1% FDR. Mascot score thresholds were ≥30 for trypsin and trypsin/Lys-C, and ≥25 for Lys-C, r-Lys-N and Arg-C. Addition of redundant peptides was achieved by removing the red bold requirement. **(B)** Number of proteins identified for each protease. Confident non-redundant (NR) protein identifications were assigned based on 1 or more significant peptides, a red bold notation in Mascot and 1% FDR. Mascot score thresholds were ≥30 for trypsin and trypsin/Lys-C, and ≥25 for Lys-C, r-Lys-N and Arg-C. Addition of redundant proteins (IR) was achieved by removing the red bold requirement. **(C)** A comparison of sequence coverage for microtubule associated protein 6 (MAP6), myelin basic protein (MBP) and spectrin-1 (SPTN1). The blue bars represent the total sequence coverage after combining the separate measurements.



	Trypsin	Trypsin/Lys-C	Lys-C	rLys-N	Arg-C
Reduced	32	35	0	142	9
Overlap	135	165	6	265	10
Non-reduced	435	589	164	188	750

Figure 3-3: Comparison of confident unique protein identifications obtained from matrix extracts taken from tissue sections treated with on-tissue digestion (red) and tissue sections treated with on-tissue disulfide bond reduction followed by on-tissue digestion (green). Proteins occurring in both experiments were designated as overlap (grey).

protease was determined by comparing protein identifications from both non-reduced and reduced datasets (Figure 3-3). For r-Lys-N the TCEP-reduced protein extract resulted in the identification of an additional 142 proteins, which were not previously identified in the non-reduced sample. A small gain in proteome coverage was also observed in the extract of trypsin (32 additional identified proteins) and trypsin/Lys-C (35 additional identified proteins).

3.3.4 Gene ontology analysis of the matrix proteome

The combined data from the digested matrix proteomes was submitted to a gene ontology analysis. The results show that the matrix proteome contained mostly cytoplasmic proteins (48.7%), followed by nuclear proteins (14.9%) and membrane proteins (16.3%) (online Supplementary information). These results are in accordance with previous data by Maier *et al.* (Maier et al., 2013), which focused on the intact proteins sampled by MALDI-MSI.

A KEGG pathway analysis (online Supplementary information), revealed that many of the proteins in the digested matrix proteome play a role in metabolism: amongst the most significant pathways are oxidative phosphorylation, the Krebs cycle, glycolysis and pyruvate metabolism. Protein pathways involved in neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease and Huntington's disease) as well as in long-term depression were identified amongst the most significant hits.

3.3.5 MALDI-FTICR-MSI: linking MSI to LC-MS/MS

To resolve isobaric peptide ions and assign identities the on-tissue digestion experiments were repeated with a 9.4T MALDI-FTICR-MS. For trypsin digestion 156 peptides, corresponding to 100 proteins (Table 3.1) and 8.2% of the peptides identified by LC-MS/MS, could be assigned to peaks in the MSI dataset. For the other proteases the numbers were: (i) trypsin/Lys-C 9.8%; (ii) Lys-C 17.4%; (iii) r-Lys-N 12.6%; (iv) Arg-C 9.2%. The combined list of assigned peptides for the five proteases corresponded to 280 proteins (an increase of 180 proteins if compared to just trypsin alone).

Table 3.1: Results of the peptide identity assignment on the MALDI-FTICR-MSI data.

Protease	Assigned peptides incl. redundant peptides total (#)	Corresponding proteins total (#)
Trypsin	156	100
Trypsin/Lys-C	242	133
Lys-C	67	40
r-Lys-N	135	86
Arg-C	152	115

For several proteins, more than one peptide was assigned, which allowed for comparison of the peptide distribution within one tissue section or between tissue sections that were digested with different proteases. Figure 3-4 shows several examples of MBP proteolytic fragments that were assigned based on the described methodology. The peptide distributions were verified using the gene expression profile for the *Mbp* gene obtained from the Allen Brain Atlas (Figure 3-1F). More examples are presented in the online Supplementary information.

3.4 Discussion

One of the most comprehensive studies regarding the proteome sampled by MALDI-MSI identified 1,400 proteins from diverse human tissues (Maier et al., 2013). It was based on the extraction of intact proteins from tissues coated with MALDI matrix, the matrix proteome, followed by separation by gel electrophoresis and in-gel trypsin digestion. Despite the valuable information collected, in terms of the proteins ultimately accessible by MSI, the published data does not allow a direct comparison with MALDI-MSI because most identified proteins are substantially larger than those usually detected by MALDI-MSI (McDonnell et al., 2014; Schober et al., 2012). Instead, the identification of the proteins analyzed by MALDI-MSI is often performed using on-tissue enzymatic digestion, in which the MALDI-MSI data is assigned, on the basis of accurate mass, to LC-MS/MS data from tissue extracts (Maier et al., 2013; Groseclose et al., 2007; Gustafsson et al., 2013). Digestion with different enzymes has been shown to increase proteome coverage in bottom-up proteomics (Choudhary et al., 2003; Gatlin et al., 2000; Hohmann et al., 2009; Wa et al., 2006). Similarly, Enthaler *et al.* (Enthaler et al., 2013b) demonstrated that the combination of trypsin, pepsin and elastase increased protein sequence coverage in MALDI-MSI. The increased sequence coverage follows from the high number of cleavage sites – elastase cleaves at the C-terminus of small hydrophobic amino acids, G, S, I, L, A, V (Rietschel et al., 2008) and pepsin at the C-terminus of amino acids, P, Y, W, L (Mótyán et al., 2013) – but at the expense of very complex MSI datasets containing many isobaric ions.

Instead we investigated whether the different activities of additional enzymes that cleave at the basic amino acids (to maintain MS detection sensitivity) can also aid MALDI-MSI. The enzymes trypsin, Lys-C, trypsin/Lys-C, Lys-N and Arg-C were investigated. A standard digestion protocol was first established using trypsin. Once similar results were reproducibly obtained, including from different animals, the method was adapted for the other proteases by changing the solution conditions (minimal changes to the spraying and incubation conditions). The success of the different proteases further demonstrates the robustness of the method (see Table S3-1 (Supplementary information) for spray method). In this paper we have focused on the improvement in proteome coverage that may be obtained in on-tissue digestion MALDI-MSI by combining the results from different proteases; to ensure comparability the data presented are obtained from a single animal.

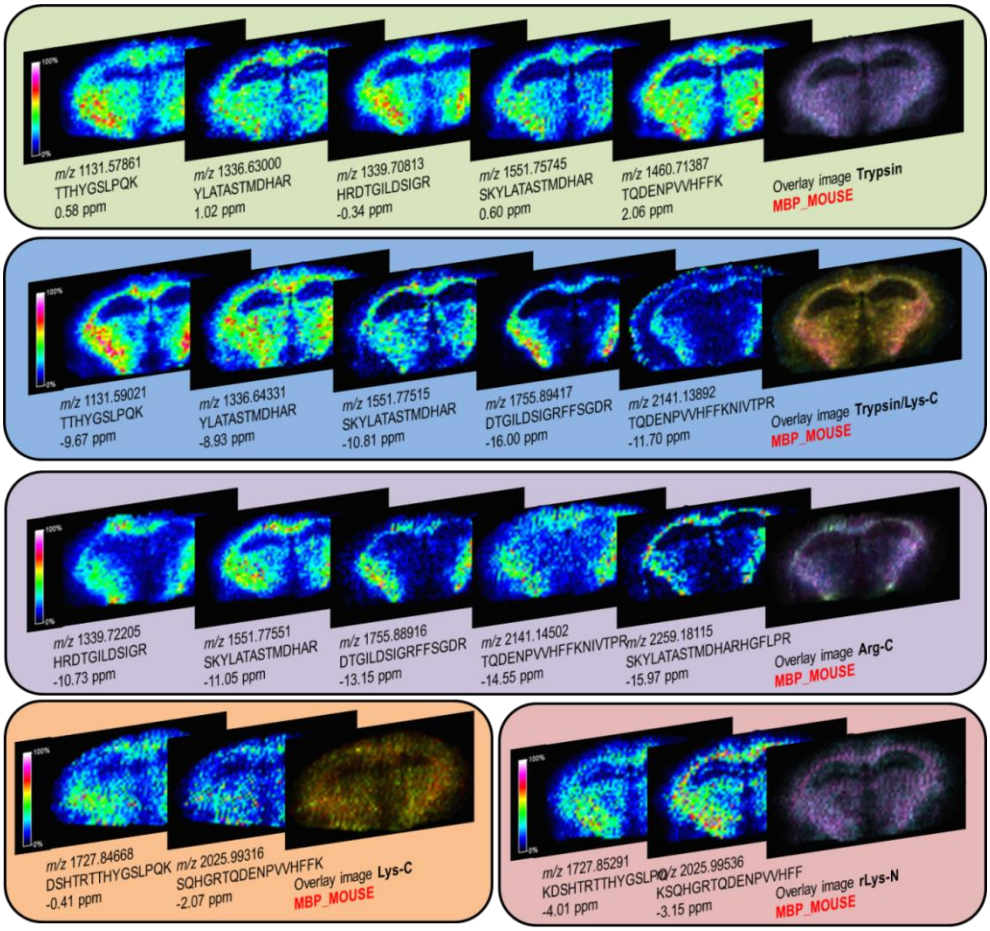


Figure 3-4: Ion distributions representing MBP peptides obtained from MALDI-FT-ICR-MSI were assigned to peptides identified by LC-ESI-MS/MS. The images were obtained from the MALDI-FT-ICR-MSI datasets of (i) trypsin (green box), (ii) trypsin/Lys-C (blue box), (iii) Arg-C (purple box), (iv) Lys-C (orange box), (v) r-Lys-N (red box). For each of the single ion images (represented by the jet intensity scale) the corresponding sequence and measurement error are presented

All enzymes were diluted to the same concentration and sprayed over the tissue with the same method. The presence of salts and buffers in the protease solutions of Lys-C and r-Lys-N led to poor matrix crystallization, low signal intensity and noisy MSI images (Figure S3-1, Supplementary information). A buffer exchange with milliQ water (pH 6.5 – 7) improved the quality of the MS images but the low number of peptide and protein identifications indicated that the lack of salts and buffers adversely affected enzymatic activity (Figure 3-2). Still, the combination of proteases greatly increased the number of detected peptides and proteins; the 5,337 peptides and 1,198 proteins represent an increase of 179% and 110%, respectively, as compared to trypsin alone (1,913 peptides; 570 proteins).

Unlike LC-MS/MS based protein identification, MSI of proteolytic peptides does not differentiate unique peptides from non-unique peptides and all peptides are detected in the same mass spectrum. Consequently, the datasets will contain non-unique peptides as well as many isobaric ions. Both of these characteristics frustrate efforts to identify proteolytic peptides directly from tissue using MS/MS (Houel et al., 2010). The use of multiple enzymes for on-tissue digestion has the advantage of producing different peptides from the same protein that can be used to confirm its distribution in the tissue sample (Figures 3-1 and 3-4).

To connect the confidently identified peptides and proteins from the matrix extracts to the on-tissue digestion MALDI-MSI experiments high mass resolution MALDI-MSI data was acquired on a 9.4T MALDI-FTICR instrument. Between 8.2 - 17.4% of the peptides identified by LC-MS/MS could be assigned to peaks in the high mass resolution MALDI-MSI datasets (Table 3.1). The combined total number of 633 peptides originated from 280 proteins (online Supplementary information). A previous publication by Schober *et al.* (Schober et al., 2012) reports a similar approach where over 1,100 peptides were first identified from a mouse brain tissue homogenate, of which 13% were assigned to *m/z* features in a high mass resolution MALDI-MSI dataset, and corresponded to 101 proteins. The Schober paper used extensive peptide fractionation to increase protein identification rates and analyzed the entire tissue's proteome. In contrast the results reported here utilized no fractionation and only analyzed the matrix proteome. Without the fractionation step 140 peptides were identified in the LC-MS/MS analysis of which 60 could be assigned to *m/z* features in the MALDI-MSI dataset, corresponding to 38 proteins (Schober et al., 2012). Here, using complementary enzymes for on-tissue digestion and LC-MS/MS of the matrix peptides (without any fractionation step) we could assign more than 600 peptides, corresponding to 280 proteins.

The peptide assignments were based on a mass error tolerance of ±20 ppm. The cyclotron frequency of an ion inside an ICR cell, and therefore the accurate calibration of an FTICR instrument, is dependent on the magnetic field strength as well as the local electric field. Accordingly, the calibration is dependent on the number of ions in the cell, which in MALDI-MS varies from shot-to-shot, and in MALDI-MSI from pixel to pixel. Consequently, there are slight mass deviations in all pixels, resulting in broader peaks in the average mass spectrum. Peak picking and peptide identity assignment were based on the average, non-aligned mass spectrum and therefore a higher mass tolerance was applied. However, the consensus between the images obtained for different peptides, utilizing different proteases provides additional corroboration.

The LC-MS/MS data reported here of MALDI matrix extracts also acts as a guide to the proteins/peptides that may be analyzed via MALDI-MSI. It has previously been shown that MSI detected only a fraction of those detected by LC-MS/MS analysis of tissue extracts (Schober et al., 2011). A single 100 x 100 µm pixel analyses the

equivalent of ≈12 cells (assuming average cell size of 20 μm and tissue thickness of 10 μm), and which is analyzed without any explicit purification and separation step. To be detected by MALDI-MSI the peptides are required to be present in the MALDI matrix and present at sufficiently high levels. While we have demonstrated an increase in proteome coverage the number remains a fraction of the tissue's total proteome, because of the sample-volume-limited nature of the analysis and the need for matrix incorporation of the peptides. The FTICR-MS used here, and the Orbitrap used previously (Schober et al., 2012), are characteristically very high dynamic range mass analyzers. To significantly increase the number of peptides detected by MSI will require significant increases in the charge capacity of these ion trap mass analyzers (while maintaining performance) in order for the lower level peptides to be above the detection threshold.

Amongst the 280 assigned proteins detected here were several histones. Histones are highly implicated in cancer progression through their role in determining chromatin structure and gene accessibility. Proteolytic fragments of histones have been previously reported in cancerous tissues analyzed by on-tissue-digestion MALDI-MSI (Djidja et al., 2009a, 2009b). Several peptides were assigned to subunits of mitochondrial ATPases (AT1A1, -2, -3, AT2B1, -2, ATPA, ATP5H & ATP5J). These trans-membrane proteins are known to be involved in a multitude of neurological disorders, e.g. familial hemiplegic migraine (Morth et al., 2009; Poulsen et al., 2010) and Alzheimer's disease (Markesbery, 1997; Zlokovic, 2011). These results were confirmed by the GO analysis and KEGG pathway analysis which show that proteins involved in several metabolic pathways and neurological disorders are well represented in the digested matrix proteome (online Supplementary information).

3.5 Conclusions

This work reports the peptides identified from the on-tissue-digestion matrix proteome as well as the peptides detected by ultra-high mass resolution MALDI-MSI, for multiple enzymes; both valuable sources of information for researchers applying on-tissue digestion MSI. The results demonstrate the ability to increase proteome coverage by using different enzymes and indicate that on-tissue digestion MALDI-MSI analysis of the mouse brain can be applied in a wide range of disease related, biomarker discovery or other neuro-scientific research applications.

It should be noted that we have been deliberately conservative with the thresholds used in the assignments (>0.6% base peak). It is known that FTICR and Orbitrap mass spectrometry can span a very wide dynamic range, and more peptides could have been assigned if a lower threshold was used. However, at lower thresholds many peaks were associated with noisy images and the greatly increased number of peaks is accompanied by an increased risk of mis-assignment. It is expected that with additional improvements in dynamic range, sensitivity and on-tissue MS/MS based verification of

peptide identities that even more peptides will be assignable and the application potential of MALDI-MSI increased further.

3.6 Acknowledgements

This work was supported by the ZonMW Zenith project Imaging Mass Spectrometry-Based Molecular Histology: Differentiation and Characterization of Clinically Challenging Soft Tissue Sarcomas (No. 93512002; BH), Centre for Medical Systems Biology (CMSB) in the framework of the Netherlands Genomics Initiative (NGI) (AvdM), EU-funded Marie Curie Action ENIGMAS FP7-PEOPLE-2011-IEF (No. 303344; RC), Marie Curie Career Integration Grant (No. 294233; EAT), FP7 EUROHEADPAIN (No. 602633;AvdM) and Marie Curie IAPP Program BRAINPATH (No. 612360; AvdM & EAT). The authors thank Nathalie Rieff for technical assistance.

3.7 Supplementary information

The full supplementary information, files and tables are available on the internet via http://pubs.acs.org/doi/suppl/10.1021/ac503952q.

3.7.1 Supplementary tables

Table S3-1: Overview of the settings used on the SunChrom SunCollect sprayer during TCEP-, protease and matrix application.

Setting	TCEP	Protease	Matrix
Vial X *(mm)*	0.5	0.3	0.5
Vial Y *(mm)*	2.0	0.3	2.0
Z *(mm)*	25.0	45.5	25.0
Z-offset *(mm)*	0.0	0.0	0.0
Layers *(#)*	4	5	3
Flowrate Layer 1 *(µL/min)*	10	5	10
Flowrate Layer 2 *(µL/min)*	10	5	35
Flowrate Layer 3 *(µL/min)*	10	5	35
Flowrate Layer 4> *(µL/min)*	10	5	-
Speed X	Low (3)	Low (4)	Low (3)
Speed Y	Medium (1)	Medium (1)	Medium (1)

3.7.2 Supplementary figures

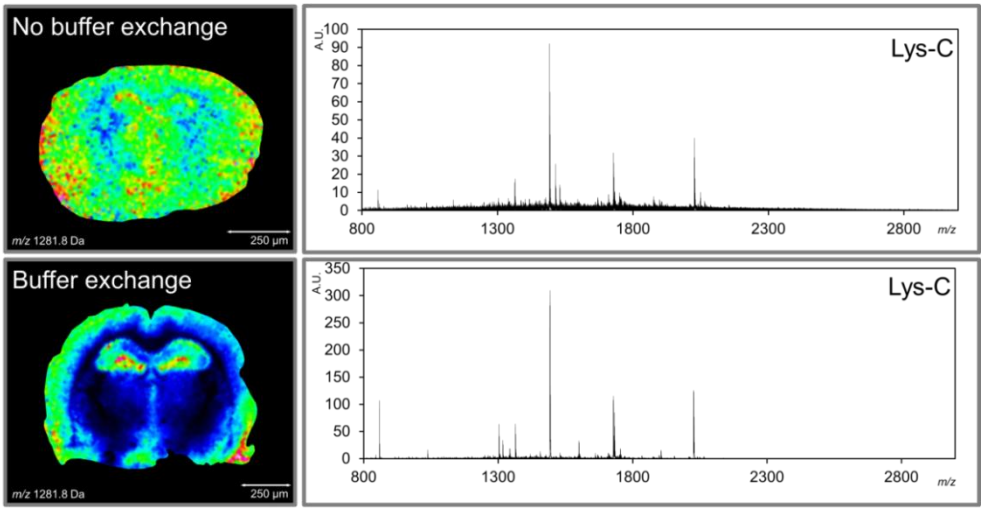


Figure S3-1: The effect of the enzyme buffer exchange procedure required to remove ion suppressing compounds (e.g. HEPES) from the enzyme buffer. The **top** panel contain an image and an average spectrum taken from a lys-C on-tissue digestion MALDI-MSI analysis, without performing enzyme buffer exchange. The **bottom** image and spectrum were taken from a lys-C on-tissue digestion MALDI-MSI analysis, subsequent to the enzyme buffer exchange procedure. Ion intensities in the MALDI-MSI data are represented by the jet intensity color scale.