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

Heijs, B.P.A.M.

Chapter 5

Multimodal mass spectrometry imaging of *N*-glycans and proteins from the same tissue section

Bram Heijs‡; Stephanie Holst‡; Inge H. Briaire-de Bruijn; Gabi W. van Pelt; Arnoud H. de Ru; Peter A. van Veelen; Richard R. Drake; Anand S. Mehta; Wilma E. Mesker; Rob A. Tollenaar; Judith V.M.G. Bovée; Manfred Wuhrer; Liam A. McDonnell, "Multimodal mass spectrometry imaging of *N*-glycans and proteins from the same tissue section," *Analytical Chemistry* 88(15), pp. 7745-7753, 2015.

On-tissue digestion matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) is able to record spatially correlated molecular information from formalin-fixed, paraffin-embedded (FFPE) tissue sections. In this work we present the in-situ multimodal analysis of N-linked glycans and proteins from the same FFPE tissue section. The robustness and applicability of the method are demonstrated for several tumors, including epithelial and mesenchymal tumor types. Major analytical aspects, such as lateral diffusion of the analyte molecules and differences in measurement sensitivity due to the additional sample preparation methods have been investigated for both N-glycans and proteolytic peptides. By combining the MSI approach with extract analysis we were also able to assess which mass spectral peaks generated by MALDI-MSI could be assigned to unique N-glycan and peptide identities.

5.1 Introduction

Matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) is an analytical technique for the spatially resolved biomolecular investigation of tissue sections (Caprioli et al., 1997; McDonnell and Heeren, 2007). A major advantage of MSI is that it can be combined with histology so that cell-type specific mass spectral profiles can be extracted from the MSI datasets. Such cell-type specific molecular signatures have been used to identify diagnostic and prognostic biomarkers (Balluff et al., 2011, 2015; Dekker et al., 2014; Meding et al., 2012).

MALDI-MSI is also able to analyze different molecular classes, using essentially the same technology but different sample preparation strategies. MALDI-MSI has been used to assess how protein, lipid, metabolite, and glycan profiles change in cancerous tissues. To readily exploit the potential insights and improved diagnostic performance that may be obtained by combining the molecular information from different molecular classes, the multimodal molecular profiles from identical histological features can be compared. Multimodal MSI is a widely-used term and covers different strategies. For example, one could combine MSI with a complementary imaging modality, or apply repeated MSI analyses to the same tissue: (i) utilizing different ionization modalities (Eberlin et al., 2011; Amstalden van Hove et al., 2010), (ii) analyzing multiple polarities using the same ionization source (Ellis et al., 2016), or (iii) analyze multiple analyte classes using the same ionization source. Previous work by Steven et al., in which the separate analysis of lipids and proteins from the same fresh frozen mouse brain section was presented, shows the potential of the latter multimodal MSI approach (Steven and Bunch, 2013). However, to better exploit the tissue banks available in most clinical centers, as well as the improved histological clarity, the multimodal analysis should be applicable to formalin-fixed, paraffin-embedded (FFPE) tissues.

FFPE tissue sections can be analyzed using on-tissue digestion, in which enzymes are directly applied to the tissue and MALDI-MSI used to analyze the released species (Shimma et al., 2006). On-tissue digestion MALDI-MSI of proteins has been performed using a variety of different proteolytic enzymes, including trypsin, pepsin, elastase, Arg-

C, Lys-C, and recombinant Lys-N (Enthaler et al., 2013b; Heijs et al., 2015). A recent extension to the on-tissue digestion approach is the use of peptide *N*-glycosidase F (PNGase F), which releases *N*-linked glycans, or *N*-glycans, from glycoproteins (Powers et al., 2013).

N-glycans are one of the most common post-translational modifications (Figure 5-1A+B), play a fundamental role in many molecular and cellular processes, and are established biomarkers in clinical research (Adamczyk et al., 2012; Pinho and Reis, 2015; Varki et al., 2015). Serum biomarkers, such as the glycoproteins CA19-9 and carcinoembryonic antigen are currently used for the prediction of pancreatic cancer and follow-up in colorectal cancer (CRC) (Drake et al., 2010; Thirunavukarasu et al., 2011). In MALDI-MSI, despite only being recently introduced, PNGase F-based on-tissue digestion has been successfully applied for the analysis of *N*-glycans in normal tissues (Miyoshi and Nakano, 2008; Muinelo-Romay et al., 2008; Pinho and Reis, 2015), in human hepatocellular carcinomas (Powers et al., 2015), and in tissue microarrays (Powers et al., 2014).

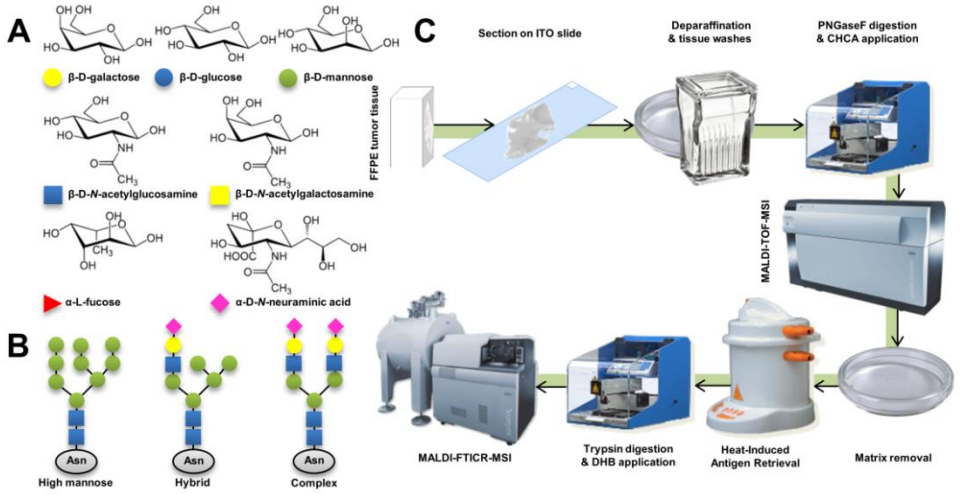


Figure 5-1: (A) An overview showing the structural formulas of the most prevalent monosaccharide building blocks of N-linked glycans, and their symbol in the schematic representation. **(B)** Schematic representations of the three N-glycan classes with common core structure. The asparagine (Asn) residue represents the first amino acid of the consensus sequence "Asn-X-Thr/Ser". **(C)** An overview of the sequential on-tissue digestion MALDI-MSI workflow, showing the order of the various sample preparation steps.

Combining the information on the protein and glycan levels in a given tissue has been demonstrated to increase their potential utility as diagnostic biomarkers: α-1,6-core-fucosylated haptoglobin is explored as a biomarker in various malignancies such as pancreatic and colorectal cancer, and the U.S. Food and Drug Administration (FDA) approved α-1,6-core-fucosylated alpha-fetoprotein was demonstrated as biomarker for

hepatocellular carcinomas. (Miyoshi and Nakano, 2008; Muinelo-Romay et al., 2008; Park et al., 2011)

Here, we report a multimodal MALDI-MSI analysis of both *N*-linked glycans and proteolytic peptides from a single tissue section using sequential on-tissue digestion MALDI-MSI. Importantly, the tissue section remains almost un-damaged by the analysis and so can be histologically stained, annotated and registered to the MSI datasets. In this manner, *N*-glycans and proteins from exactly the same cells or histological features can be extracted, and used for further analysis.

5.2 Methodology

5.2.1 Chemicals and reagents

All chemicals were purchased from Sigma-Aldrich (Steinheim, Germany) except ethanol and xylene (Merck, Darmstadt, Germany). Trypsin was purchased from Promega (Madison, USA). Recombinant Peptide *N*-Glycosidase F (PNGase F) from *Flavobacterium meningosepticum* was expressed and purified as previously described (Powers et al., 2013).

5.2.2 Tissues & sample collection

Leiomyosarcoma (LMS, L1091-3), myxoid liposarcoma (MLS, L2039), and colorectal carcinoma (CRC, C67) tissues were formalin-fixed and paraffin-embedded following established histopathological methods (Figure S5-1, Supplementary information). Following Dutch national ethical guidelines (Code for proper secondary use of human tissue, Dutch Federation of Medical Scientific Societies), all tissue samples were handled in a coded fashion. Using a microtome, 6 µm thick sections were cut from the FFPE tissue blocks. The tissue sections were mounted on poly-L-lysine coated indium-tin-oxide (ITO) glass slides (Bruker Daltonics, Bremen, Germany), dried overnight at 37°C and stored at 4°C protected from light until further treatment. Slides dedicated for MALDI-MSI analysis contained two consecutive tissue sections, one used for the enzymatic digestions, the other used for a negative, no-enzyme control. Single sections for histology staining and molecular extractions were placed on Starfrost adhesive microscope slides (Light Laboratories, Dallas, TX).

5.2.3 Sequential on-tissue digestion MALDI-MSI

A buffer exchange, using a 10 kDa molecular weight cut-off (MWCO) spin filter (Merck Millipore, Billerica, MA), was performed on the recombinant PNGase F enzyme to ensure uniform purity and enzyme concentration over all performed experiments. The buffer exchange was performed according to the manufacturers' protocol. Prior to the sample preparation, the ITO slides containing FFPE tissue sections were placed on a hot plate (60°C) for 60 minutes to increase the adherence of the tissues to the ITO slides. Paraffin was removed by submerging the slides in xylene (wash 1 - 5 min; wash 2 - 10 min). The tissues were washed with ethanol (100%, 2x 2 min) and deionized

water (2x 5 min), dried in a vacuum desiccator (~10 min) before homogeneously applying PNGase F (100 ng/µL in deionized water) using a SunCollect pneumatic sprayer (SunChrom, Friedrichsdorf, Germany; 15 layers at 10 µL/min). Following overnight incubation, MALDI matrix (5 mg/mL α-2,4-hydroxycinammic acid in 50% ACN and 0.1% TFA) was homogeneously deposited on the tissue sections using the SunCollect sprayer (6 layers: (1) 10 µL/min, (2) 20 µL/min, (3) 30 µL/min, and (4+) 40 µL/min). MSI analysis of the *N*-linked glycan species was performed on a Bruker UltrafleXtreme MALDI-TOF/TOF instrument (Bruker Daltonics). Data was recorded in reflectron positive ion mode covering the mass range between *m/z* 900 to *m/z* 4,000, using 500 laser shots per spot and a 100 x 100 µm pixel size. More in-depth information on the MALDI-TOF-MS method is provided in Table S5-1A (Supplementary information).

Following the *N*-glycan data acquisition, excess MALDI matrix was removed from the tissues using ethanol washes (70%, 2x 1 min) and the tissue was conditioned for heat-induced antigen retrieval (HIAR) by washing with deionized water (2x 5 min). HIAR was performed in a 10 mM citric acid buffer adjusted to pH6 with 1 M NaOH. After HIAR, the tissues were carefully washed with deionized water (2x 2 min) and trypsin (20 ng/µL in deionized water) was deposited homogeneously using a SunCollect pneumatic sprayer (15 layers at 10 µL/min). Following overnight incubation at 37°C, in a saturated environment (50% methanol in deionized water), MALDI matrix was applied using the SunCollect system and a solution of 20 mg/mL 2,5-dihydroxybenzoic acid (DHB) in 50% ACN and 0.1% TFA. MSI analysis of the proteolytic peptides was performed on a 9.4 T Bruker SolariX XR MALDI-FTICR instrument equipped with a dynamically harmonized ParaCell™. Data was acquired in positive ion mode in the mass range *m/z* 600 to *m/z* 3,500 with an average resolving power of 54,000 over the analyzed mass range, using 150 laser shots per spot and a 100 x 100 µm pixel size. More in-depth information on the MALDI-FTICR-MS method is provided in Table S5-1B (Supplementary information).

For both *N*-glycan, and proteolytic peptide analysis, control experiments were performed on consecutive tissue sections on the same glass slides. The control samples were treated following the same protocols described above but without application of enzymes, and MALDI-MSI data acquisition was performed at a 250 x 250 µm pixel size. Also, to show repeatability of the method on the same tissue, the sequential on-tissue digestion MALDI-MSI analysis of the LMS (L1091-3) was performed in duplicate. The two analyses are indicated as LMS-1 and LMS-2.

5.2.4 *N*-glycan extraction and identification

For MS/MS identification of the *N*-glycans present in the tissue sections, separate tissue sections on Starfrost adhesive microscope slides were treated following the protocols described in the previous section, although matrix application was omitted. Released *N*-glycans were extracted by carefully distributing 100 µL deionized water on top of the

tissue, 15 min incubation at 37°C, and collected by micropipette. Extracts were dried in a vacuum centrifuge and resuspended in 20 µL deionized water. For purification, 10 µL of the sample was brought to 85% ACN, followed by cotton hydrophilic interaction liquid chromatography (HILIC) micro-solid phase extraction (SPE) using 200 µL pipette tips (Selman et al., 2011). *N*-glycans were eluted in 15 µL deionized water from which 5 µL were spotted in duplicate onto a Bruker AnchorChip plate. Dried samples were co-crystallized with 0.5 µL super DHB matrix (5 mg/mL in 50% ACN) supplemented with 1 mM NaOH. MS1 and MS2 spectra were recorded using an UltrafleXtreme MALDI-TOF/TOF-MS for the identification of the *N*-glycans. Furthermore, overall average spectra from the MALDI-MSI analyses were exported from flexImaging 4.1 (Bruker Daltonics) to CSV format. The CSV-formatted spectra were imported in mMass (v5.5.0, http://www.mmass.org) (Strohalm et al., 2008), where baseline correction, smoothing, peak picking (S/N ≥ 6), de-isotoping and internal re-calibration was performed. The mMass peak lists were exported to both Excel 2010 (Microsoft) and the Glyco-Peakfinder tool (http://www.eurocarbdb.org/ms-tools/) in GlycoWorkbench (2.1 stable build 146, http://www.eurocarbdb.org/), which was used for generating a preliminary glycan composition list. The suggested compositions were then compared to the results of the MS/MS analyses for confirmation.

5.2.5 Proteolytic peptide extraction and identification

Consecutive tissue sections were prepared for extracts using the same sample preparation protocol as for MALDI-MSI. The matrix application and MALDI-MSI analysis were omitted, and instead, the proteolytic peptides were extracted from the matrix using a series of solvents: (i) 50 µL of 0.1% TFA, and (ii) 50 µL of 50% ACN in 0.1% TFA. The extracts were combined, dried, resuspended in 0.1% TFA, and cleaned using Omix C18 tips (Agilent, Santa Clara, California). Peptide extracts were analyzed using an Easy nLC1000 (Thermo, Bremen, Germany) coupled to a Q-Exactive mass spectrometer (Thermo). Fractions were injected onto an in-house prepared 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% 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 as follows: 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 s, and allowed charges were 1−5. The Thermo.RAW files were converted to MGF files using the MSconvert software (64-bit for Windows, http://proteowizard.sourceforge.net). Peptide and protein identifications were extracted from the SwissProt database using the Mascot server. A maximum of two missed cleavages was allowed and methionine oxidation, deamidation (NQ), and N-terminal acetylation were set as variable modifications. Peptide assignments were made with a tolerance of 10 ppm. MS/MS fragment tolerance was

set to 20 mmu. Results were searched against a decoy database. 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 and a Mascot significance score ≥ 38 were included in the final lists of unique peptides.

5.2.6 Peptide identity assignment

Average spectra from the MALDI-FTICR-MSI datasets were exported into CSV format and loaded in mMass. Peak picking was performed on peaks with S/N ≥ 5 and a relative intensity ≥ 2.5%. The peak lists were treated with a deisotoping algorithm set to remove isotope peaks with a maximum charge of 2+ and an isotope mass tolerance of 0.02 Da. MALDI-FTICR-MSI peak lists and Mascot database search results from the LC-ESI-MS/MS analyses were exported to Excel 2010 (Microsoft). The *m/z* features originating from the MSI data were aligned to the *m/z* values of the identified peptides from the Mascot search results. Peptide identity assignments were based on a mass tolerance of ±10 ppm.

5.2.7 Fold change calculations

Regions of interest (ROIs) for (non-)tumor regions were selected in both N-glycan and peptide MSI datasets. The MSI data was loaded in Data Analysis 4.2 (Bruker Daltonics) to obtain average spectra and extract intensities for *N*-glycans and peptides from the ROIs. Intensities from (non-)tumor ROIs were averaged in Excel 2010 to obtain a single intensity value for the (non-)tumor areas. The fold change was calculated as the ratio $I_{tumor} / I_{non\text{-}tumor}$.

5.3 Results and discussion

On-tissue digestion MALDI-MSI, utilizing proteases and endoglycosidases, enables the analysis of proteins and glycans from FFPE tissues (Casadonte and Caprioli, 2011; Gustafsson et al., 2013; Heijs et al., 2015; Powers et al., 2013, 2014). In this work we have merged both on-tissue digestion protocols to create a new multimodal method that enables the analysis of both *N*-glycans and proteolytic peptides from the same FFPE tissue section (Figure 5-1C). By combining the workflows, we increased the comparability between the MSI datasets of two intimately related analyte classes.

5.3.1 Antigen retrieval

When analyzing proteins in FFPE tissues an antigen retrieval step is necessary to remove the cross-links between the proteins established by formalin fixation (Shi et al., 2013, 1991). However, it was found that heat-induced antigen retrieval (HIAR) prior to PNGase F digestion (and subsequently followed by proteolytic digestion) caused de-sialylation of the *N*-glycans and severe delocalization of the proteins (data not shown). These artefacts were circumvented by performing the HIAR-based antigen retrieval after analysis of the *N*-glycans and removal of the excess MALDI-matrix (Figure 5-1C). For *N*-glycan MSI analysis, protocols both including (Gustafsson et al., 2015; Powers et

al., 2014) and excluding (Powers et al., 2013) HIAR have been described. The chemical reactions involved in formalin fixation (Tanca et al., 2012) and the chemical composition of the *N*-linked glycans (Figure 5-1A+B) make it unlikely that the *N*-glycans will become part of the formalin-cross-linked network and so remain accessible to PNGase F. This was confirmed by our *N*-glycan MALDI-MSI analyses prior to HIAR, as a large number of *N*-glycan species were detected (Table 5-1 & Figure 5-2A). The *N*-glycan MSI analyses of the different tumor tissues yielded a total of 59 unique *N*-glycan species, of which we were able to confirm 25 by MALDI-MS/MS on glycan extracts (Table 5-1 & online Supplementary information).

5.3.2 *N*-glycan MALDI-TOF-MSI

Visualizing the distributions of several specific *N*-glycan species revealed that their abundance in the CRC, LMS, and MLS tissues was restricted to specific morphological structures in the tissue (Figure 5-2 & Figures S5-2 and S5-3, Supplementary information). For example, the complex-type *N*-glycan Hex5HexNAc4dHex1 (Hexose = Hex, N-acetylhexosamine = HexNAc, fucose = dHex) detected at *m/z* 1,809.6 [M+Na]+ co-localized exclusively with healthy striated muscle cells and collagen-rich areas in both analyses of the LMS tissue, while other *N*-glycans, such as Hex4HexNAc4dHex1 at *m/z* 1,647.6 [M+Na]+, and Hex4HexNAc4 at *m/z* 1,501.5 [M+Na]+ were almost exclusively localized in the tumor area of the same LMS tissues (Figure 5-2B+C & Figure S5-1A and S5-2, Supplementary information). In the CRC and MLS tissues similar patterns were observed, e.g. the high-mannose type *N*-glycans (Hex6HexNAc2 at *m/z* 1,419.5 [M+Na]+, Hex7HexNAc2 at *m/z* 1,581.5 [M+Na]+, and Hex9HexNAc2 at *m/z* 1,905.6 [M+Na]+) were detected in the tumor areas. These *N*-glycans have previously been associated with tumor cells in pancreatic, prostate, and hepatocellular carcinoma tissues (Powers et al., 2014, 2015). Hex5HexNAc2 at *m/z* 1,257.4 [M+Na]+, was highly abundant in several non-calcified necrotic regions present within the tumor area of the CRC tissue (Figure S5-3B+C, Supplementary information). The high mannose-type glycans are the precursors of the intracellular *N*-glycan biosynthesis pathway (Freeze, 2006; Varki et al., 2015). The high abundance of this specific glycan in the necrotic regions is likely the result of increased cell lysis, a common phenomenon in necrotic tissue, causing an uncontrolled release of intercellular molecules (Proskuryakov et al., 2003).

Table 5-1: Overview of the N-glycans detected in the MALDI-MSI and MS/MS datasets. Structural confirmation by MS/MS is within ± 1 monosaccharide.

m/z	Composition	Neutral Exchange	LMS		CRC		MLS
[M+Na]+			*MSI*	*MS/MS*	*MSI*	*MS/MS*	*MSI*
933.32	H3N2		X		X		X
1079.37	H3N2F1		X	X	X	X	X
1095.37	H4N2		X		X		X
1136.40	H3N3		X		X		X
1257.42	H5N2		X	X	X	X	X
1282.45	H3N3F1		X	X	X		X

1298.45	H4N3		X	X	X	X	X
1339.48	H3N4		X		X		X
1419.48	H6N2		X	X	X		X
1444.51	H4N3F1		X	X	X	X	X
1460.50	H5N3		X		X		X
1485.53	H3N4F1		X	X	X	X	X
1501.53	H4N4		X		X	X	X
1542.56	H3N5		X	X	X		X
1581.53	H7N2		X	X	X		X
1590.57	H4N3F2		X				X
1606.56	H5N3F1						X
1622.55	H6N3		X		X	X	X
1647.59	H4N4F1		X	X	X	X	X
1663.58	H5N4		X	X	X	X	X
1688.61	H3N5F1		X	X	X	X	X
1704.61	H4N5		X		X		X
1743.58	H8N2		X		X	X	X
1793.64	H4N4F2		X	X			X
1809.64	H5N4F1		X	X	X	X	X
1825.63	H6N4		X		X	X	X
1850.67	H4N5F1		X	X	X	X	X
1866.66	H5N5		X		X		X
1891.69	H3N6F1		X	X	X		X
1905.63	H9N2		X		X	X	
1955.70	H5N4F2		X				
1976.66	H5N4S1	Na-H	X		X	X	
1996.72	H4N5F2		X				
2012.72	H5N5F1		X		X	X	
2028.71	H6N5		X		X		
2053.75	H4N6F1		X				
2100.73	H5N4S1F1		X		X		
2122.72	H5N4S1F1	Na-H	X	X	X	X	
2158.78	H5N5F2		X				
2174.77	H6N5F1		X		X		
2199.80	H4N6F2		X				
2215.80	H5N6F1		X		X		
2267.75	H5N4S2	Na-H	X		X		
2289.74	H5N4S2	2Na-2H	X	X	X	X	
2304.83	H5N5F3		X				
2320.83	H6N5F2		X				
2325.80	H5N5S1F1	Na-H	X		X		
2341.79	H6N5S1	Na-H			X		
2361.86	H5N6F2		X				
2377.85	H6N6F1		X		X		
2393.85	H7N6				X		
2435.79	H5N4S2F1	2Na-2H			X		
2466.89	H6N5F3		X				
2487.85	H6N5S1F1	Na-H	X		X		
2507.91	H5N6F3		X		X		
2523.91	H6N6F2		X				
2539.90	H7N6F1		X		X		
2580.93	H6N7F1		X				
2669.97	H6N6F3		X				

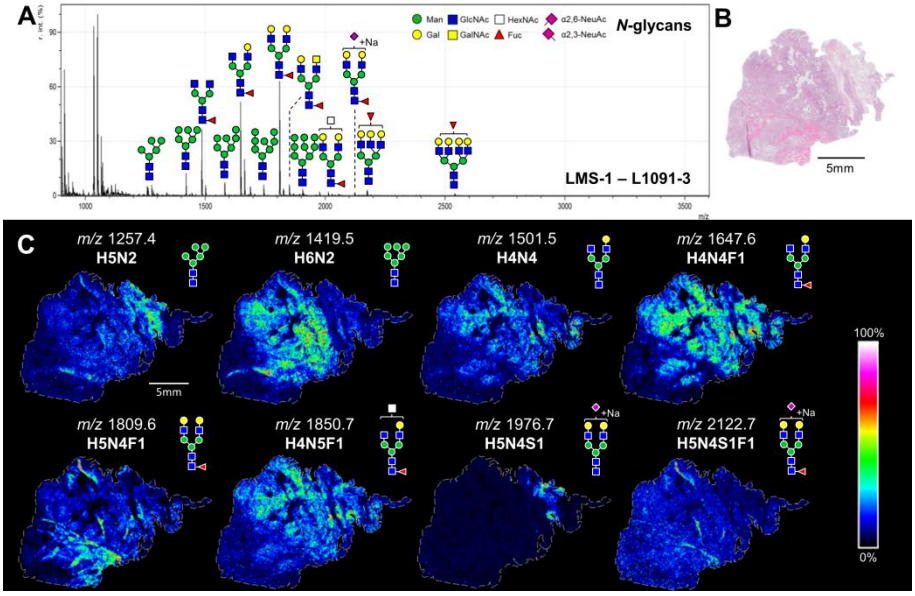


Figure 5-2: (A) The overall average mass spectrum after the *N*-glycan MALDI-MSI analysis of the LMS-1 tissue (L1091-3). The annotated glycan compositions were either confirmed by MS/MS identification or were known from literature. **(B)** Scanned image of the H&E stained LMS-1 (L1091-3) tissue. **(C)** Visualizations of the distribution of various *N*-glycans throughout the tissue. The shown glycan species were confirmed by MS/MS or confirmed by literature. The images were obtained from a TIC normalized dataset.

5.3.3 MALDI-FTICR-MSI of proteolytic peptides

The performance of the trypsin digestion after *N*-glycan release, referred to as sequential digestion, was compared to a regular on-tissue trypsin digestion experiment, referred to as the control digestion. The MALDI-FTICR-MSI data obtained from sequential digestions and control digestions were very similar for all analyzed tumor types (Figure 5-4A & Figures S5-4 and S5-5A+D, Supplementary information). On average, the spectra from the sequential digestions yielded 806(±148) isotope clusters and the spectra from the control digestions contained 844(±207) isotope clusters, with a peak overlap of 79(±14)%. The MSI images of specific peptide ions were also highly comparable between both control and sequential digestions and duplicate analyses (Figure 5-4B & Figures S5-4 and S5-5C+F, Supplementary information). For example, for sequential and control digestions of the LMS and CRC tissues the Histone H2A type 1 peptide AGLQFPVGR (*m/z* 944.5312 [M+H]+) was highly specific to the tumor areas; the proteolytic fragment NSLESYAFNMK (*m/z* 1,303.5986 [M+H]+) originating from heat shock cognate 71 kDa protein (HSP7C_HUMAN) was found to be very abundant in the tumor regions in both control and sequential digestion of the MLS tissue. Heat shock cognate 71 kDa protein is an established marker to distinguish myxoid liposarcoma from myxofibrosarcoma (Wang et al., 2014).

It was then investigated whether the additional sample preparation procedures for the multimodal on-tissue digestion MALDI-MSI, based on sequential analysis of glycans

and proteins, induced lateral diffusion of proteins. Figures 5-3 and 5-4 provide examples of proteolytic peptides highly localized to specific morphological features, indicating that any lateral diffusion was insignificant on the length-scale of the experiments.

5.3.4 Multimodal accessible proteins and glycoproteins

To determine if multimodal on-tissue digestion influenced which proteins are accessible by on-tissue digestion, and to determine if it included glycosylated peptides the matrix proteomes of both control digestion and sequential digestion of the LMS and CRC tissue sections were investigated by LC-MS/MS (Table 5-2 & online Supplementary information). The results show that the number of unique peptides identified from LMS increased from 795 (corresponding to 325 proteins) for the control digestion to 940 (corresponding to 351 proteins) for the sequential digestion. Similarly, for CRC the number of unique peptides were 948 (corresponding to 236 proteins) for the control digestion and 1,334 (corresponding to 509 proteins) for the sequential digestion. The number of proteins unique to the sequential digestion was notably greater compared to the control digestion (Figure S5-6A, Supplementary information). These numbers indicate increased proteome coverage as a result of the sequential digestion. Furthermore, the sequential digestion also increased protein sequence coverage, as the number of identified peptides per protein systematically increased (Figure S5-6B, Supplementary information).

N-glycans are attached via Asn residues present in the consensus sequence "Asn-X-Ser/Thr" ("X" can be any amino acid except proline). *N*-glycan release by PNGase F induces a deamidation of the Asn, but Asn deamidation is also a common protein modification. To identify which peptides may have been deglycosylated by PNGase F digestion the deamidated peptides, with the consensus sequence, obtained after PNGase F treatment were determined. The number of candidate deamidated peptides identified from LMS increased from 2 for the control digestion to 8 for the sequential digestion; for CRC no deamidated peptides were detected from the control digestion and 20 from the sequential digestion (Table 5-2).

Table 5-2: Overview of the peptide identifications obtained after LC-MS/MS analysis of tissue extracts.

Tumor	Control Trypsin			Sequential Trypsin		
Type	*Unique peptides*	*Deglycosylated**	*Assigned to MSI*	*Unique peptides*	*Deglycosylated**	*Assigned to MSI*
LMS	795	2	94	940	8	88
CRC	948	0	132	1334	20	134

*** Deamidation of asparagine occurred in glycosylation consensus sequence (N-*x*-T/S)**

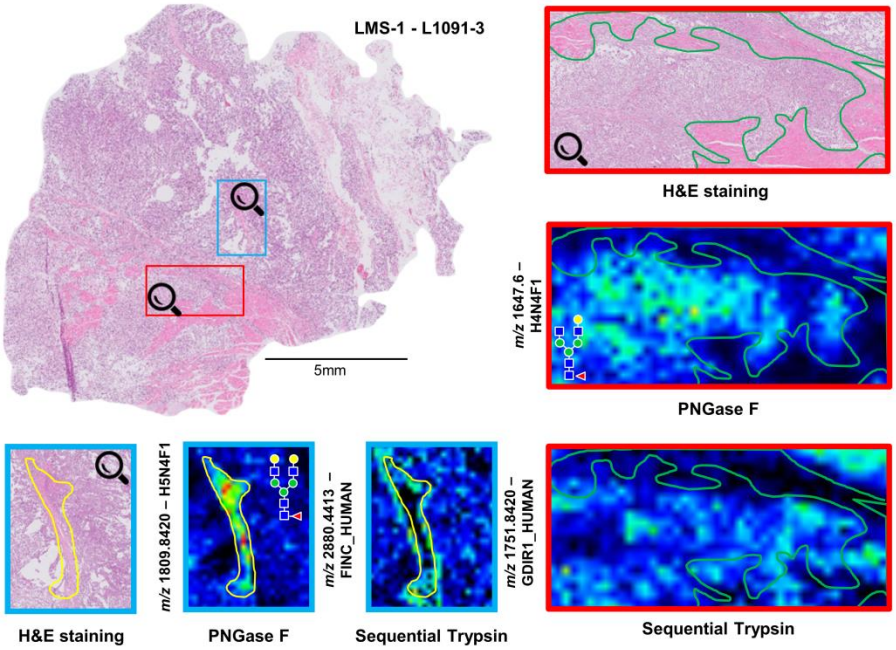


Figure 5-3: Visualizations of the distributions of several N-glycan structures and proteolytic peptides corresponding to specific morphological features (red box: healthy striated muscle tissue and connective tissue (green), blue box: collagen fibrotic connective tissue (yellow)) in the LMS-1 tissue (L1091-3).

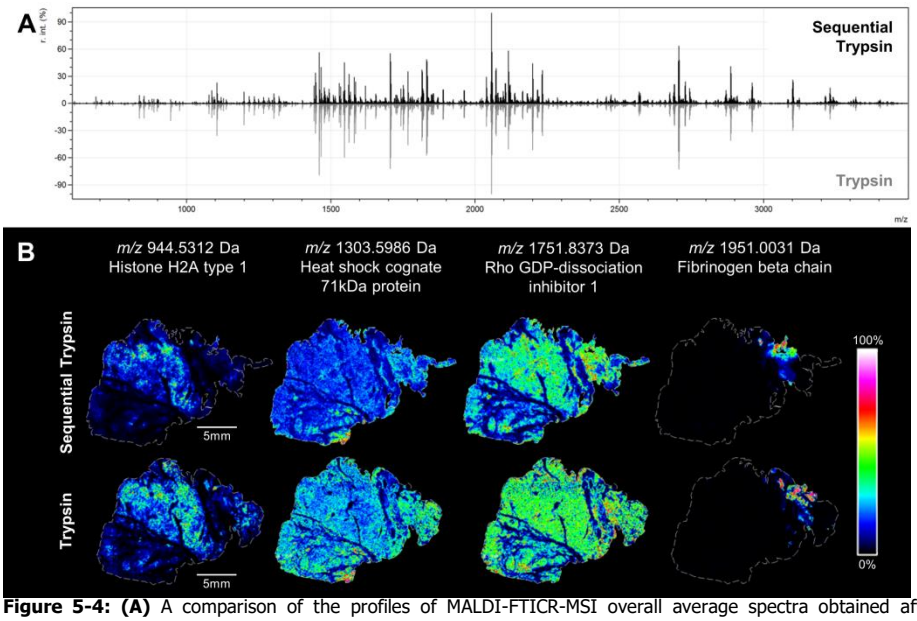


Figure 5-4: (A) A comparison of the profiles of MALDI-FTICR-MSI overall average spectra obtained after both sequential trypsin digestion (top, black) and control trypsin digestion (bottom, grey) of the LMS-1 tissue (L1091-3). **(B)** Comparison of the visualizations of specific peptide distributions. The peptides were confirmed to be present in the MALDI-matrix by LC-MS/MS of extracts and were assigned to the MALDI-FTICR-MSI datasets with an error < 10 ppm. The images were obtained from RMS normalized datasets.

5.3.5 On-tissue deglycosylation increases accessibility of proteins to proteolysis

The increase in the number of peptides identified after *N*-glycan release was significantly greater than the number of deamidated peptides. It was found that the sequential digestion affected the protein sequence coverage of both glycoproteins, and proteins without *N*-linked glycan modification. Three additional tryptic peptides of the glycoprotein galectin-3 binding protein (LG3BH_HUMAN) were detected in the sequential digestion of CRC tissue (online Supplementary information), and all three peptides were deamidated, indicating that the trypsin cleavage sites were hindered by the *N*-glycans in the control digestion. Five additional tryptic peptides of vimentin (VIME_HUMAN) were detected in the sequential digestion of the LMS tissue (online Supplementary information). Vimentin does not contain *N*-glycans (only *O*-glycans, which are not released by PNGase F), but is known to interact with other glycoproteins (Keuschnigg et al., 2009; Wang et al., 2010). This indicates that, within the tissue environment, deglycosylation of the proteome can reduce the steric hindrance caused by the *N*-glycans, making proteins more accessible to proteolytic digestion. The increased detection of non-glycosylated peptides has been previously observed in LC-MS/MS, and is even exploited as a biomarker assay (Lee et al., 2011).

The peptides detected by MALDI-FTICR-MSI were assigned to the peptide identifications obtained from LC-MS/MS extract analysis (Table 5-2 & online Supplementary information). The number of assigned peptides (matching criteria < 10 ppm, S/N > 5, intensity > 2.5% of basepeak) in the LMS control digestion experiment was 69 (66 unique peptides, corresponding to 49 proteins), and was similar in the sequential digestion (65 peptides, 54 of which were unique, corresponding to 47 proteins). Similarly, the number of assigned peptides in the CRC control digestion experiment was 110 (101 unique peptides, corresponding to 72 proteins), and was similar in the sequential digestion (112 peptides, 96 of which were unique, corresponding to 74 proteins). These numbers are greater than those reported in recent MALDI-MSI investigations of FFPE tissues (Casadonte et al., 2014; Meding et al., 2013), and included a deamidated peptide, which was detected only in the sequential digestion of both LMS and CRC and contains a known glycosylation site. The deamidated peptide detected at *m/z* 1,669.8431, could be assigned to TVIRPFYLTN*SSGVD ([M+H]+), and is the C-terminal end of the glycoprotein Galectin-3 binding protein (LG3BP_HUMAN), whose hyperglycosylated form is associated with tumor cell aggregation (Chen et al., 2009b; Lin et al., 2015).

5.3.6 Differentially expressed glycans and proteins

The results described above indicate that multimodal on-tissue digestion MALDI-MSI is robust and able to deliver high quality MALDI-MSI data for *N*-glycans and (an increased number of) proteolytic peptides from the same FFPE tissue section. By way of example Figure 5-5A shows the LMS tissue section reported in Figures 5-2, 5-3, and 5-4, with

annotations for tumor and non-tumor regions. Figure 5-5B shows the detected (area normalized) fold changes of 18 *N*-glycans, confirming the previous observation from visual inspection of the images that most glycans were detected at lower levels in the tumor. Figure 5-5C shows the fold changes from all peptides from the proteins reported in Figure 5-4. It can be seen that there is a good agreement between peptides from the same proteins, and even from the same peptide if it is detected as both protonated (red bars) and sodiated (blue bars) pseudomolecular ions.

These results, together with the results from other tumors presented in the Supporting Information, indicate how such multimodal imaging could prove valuable in future clinical studies, as many of the assigned peptides originate from proteins

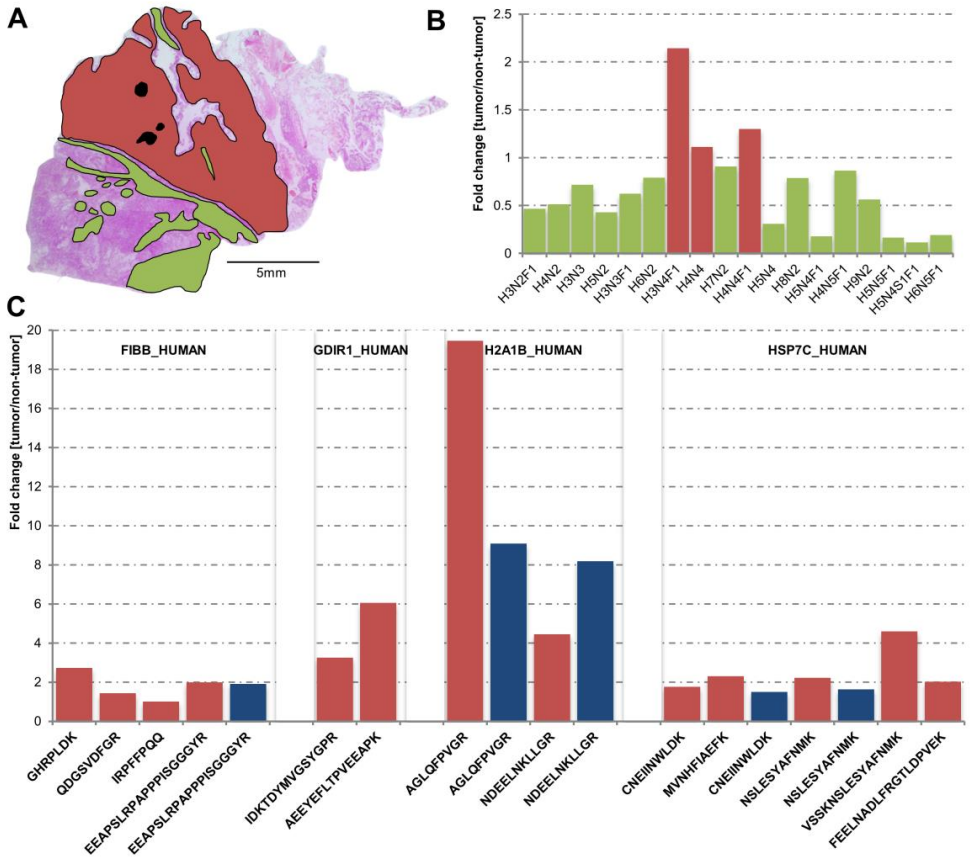


Figure 5-5: (A) Histology image of LMS-1 (L1091-3) with annotated non-tumor (green) and tumor (red) regions. Spectra acquired from holes in the tissue (black) were excluded from data analysis. **(B)** Area normalized fold changes of 18 *N*-glycans detected by MALDI-MSI. The green bars represent higher abundance in non-tumor regions. The red bars indicate a higher abundance in tumor regions. **(C)** Area normalized fold changes of protonated (red) and sodiated (blue) peptide ions originating from FIBB_HUMAN, GDIR1_HUMAN, H2A1B_HUMAN and HSP7C_HUMAN as detected by MALDI-MSI.

previously associated with disease. For example, multiple peptides from different histone variants (H1.2, H2A, and H4) were detected and it is well established that histones are heavily involved in tumor biology (online Supplementary information) (Singh et al., 2015; Wang et al., 2015a). Amongst the assigned peptides were fragments originating from vimentin, and glucose regulated protein 78kDa, proteins that have been associated with ovarian cancer and pancreatic adenocarcinoma in previous MALDI-MSI based studies (online Supplementary information) (Djidja et al., 2009b; El Ayed et al., 2010). In the CRC tissue we were able to detect proteolytic fragments from Carcinoembryonic antigen-related cell adhesion molecule 5 (CEAM5_HUMAN), a member of the carcinoembryonic antigen (CEA) family of proteins. The CEA proteins have been in use as diagnostic markers for CRC (online Supplementary information) (Zhang et al., 2015).

5.4 Conclusions

In this work we have presented multimodal analysis of both *N*-glycans and proteolytic peptides from the same tissue section. We have been able to: (i) image 59 unique *N*-glycan species from the *N*-glycan MALDI-TOF-MSI datasets, and (ii) assign a total of 89 unique peptide identifications to high-mass resolution MALDI-FTICR-MSI datasets from LMS tissue sections and 154 from CRC tissue sections, including a deglycosylated (*N*-linked glycan) peptide. Considering these results, the sequential on-tissue digestion MALDI-MSI method shows great promise for application to any FFPE tissues of clinical relevance.

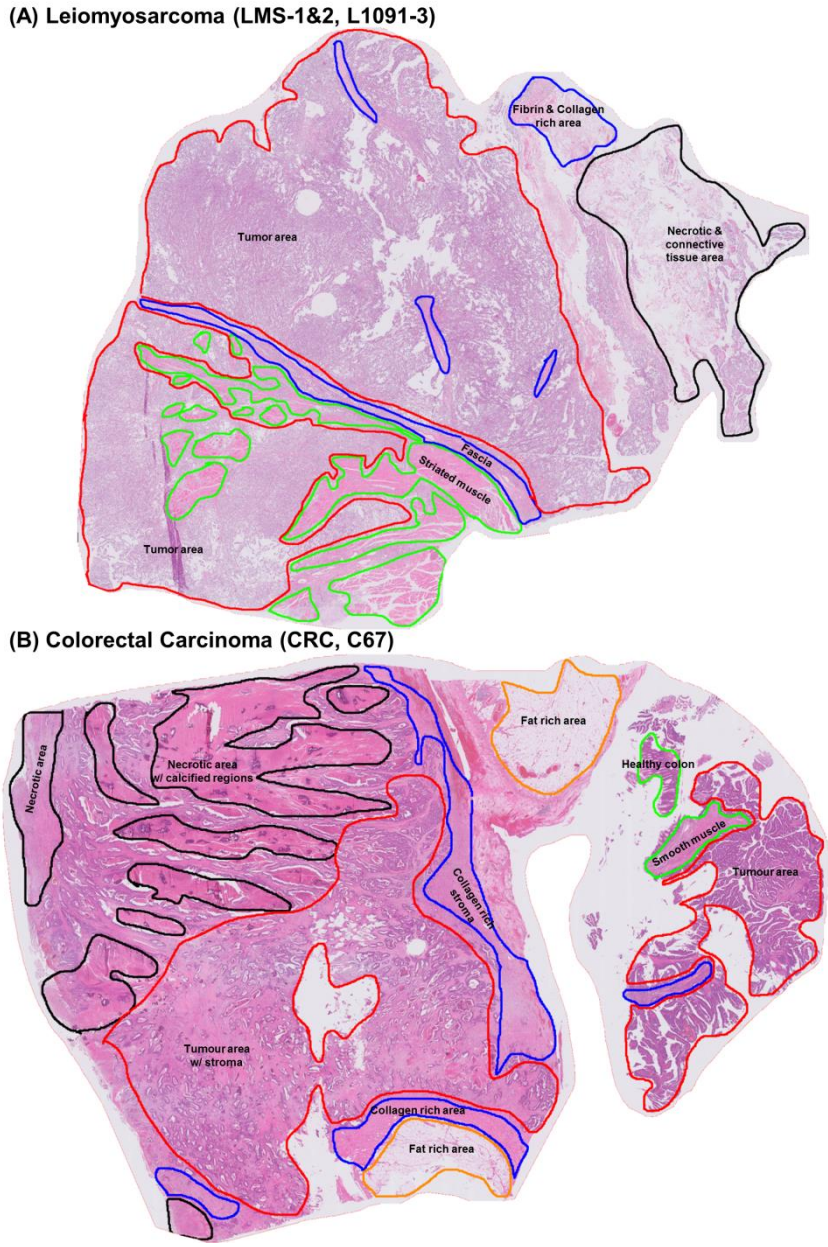
5.5 Acknowledgements

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; B.H.), the European Union Seventh Framework Programmes HighGlycan project (No. 278535; S.H.), the National Institutes of Health/National Cancer Institute (No. R21CA186799; R.R.D.), the state of South Carolina SmartState Endowed Research program to R.R.D, and by grants R01 CA120206 and U01 CA168856 from the National Cancer Institute (NCI), the Hepatitis B Foundation, and an appropriation from The Commonwealth of Pennsylvania to A.M. (R.R.D.). The authors thank Dr. P. Angel for her valuable input.

5.6 Supplementary information

The full supplementary information, files and tables are available on the internet via http://pubs.acs.org/doi/abs/10.1021/acs.analchem.6b01739.

5.6.1 Supplementary figures



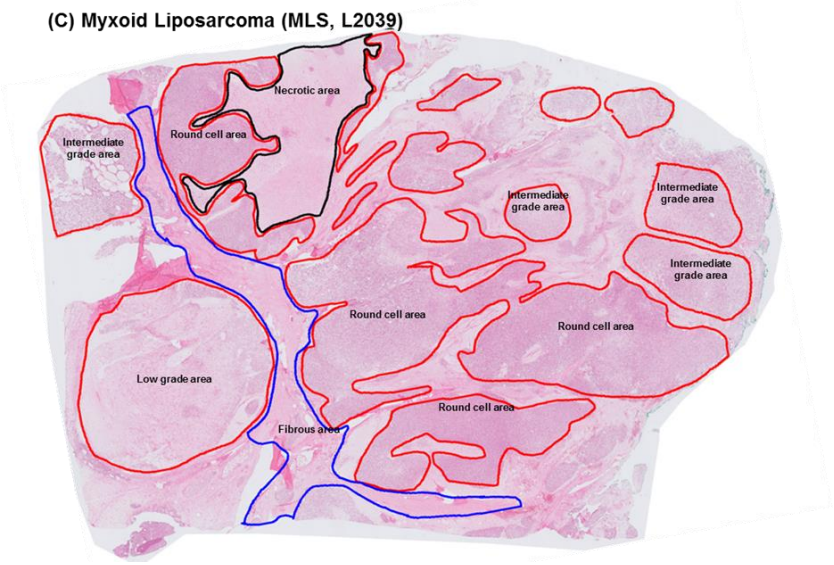


Figure S5-1: Annotated histology images of **(A)** Leiomyosarcoma (LMS-1&2, L1091-3), **(B)** Colorectal carcinoma (CRC, C67), **(C)** Myxoid liposarcoma (MLS, L2039).

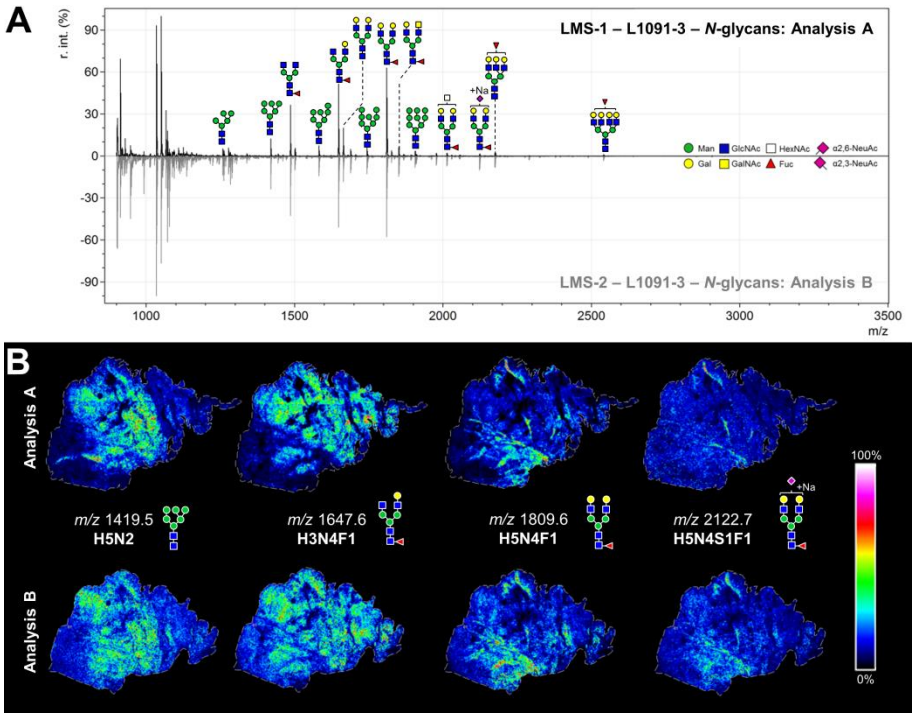


Figure S5-2: (A) A comparison of the overall average spectra obtained from the LMS-1 (L1091-3, black) and, the duplicate, LMS-2 (L1091-3, grey) tissue sections during *N*-glycan MALDI-MSI. **(B)** Visualizations of the distribution of various *N*-glycans throughout the LMS-1 and LMS-2 tissues. The annotated glycan structures represent compositions which were either confirmed by MS/MS identification or were known from literature, but structural isomers cannot be excluded. The images were obtained from TIC normalized datasets.

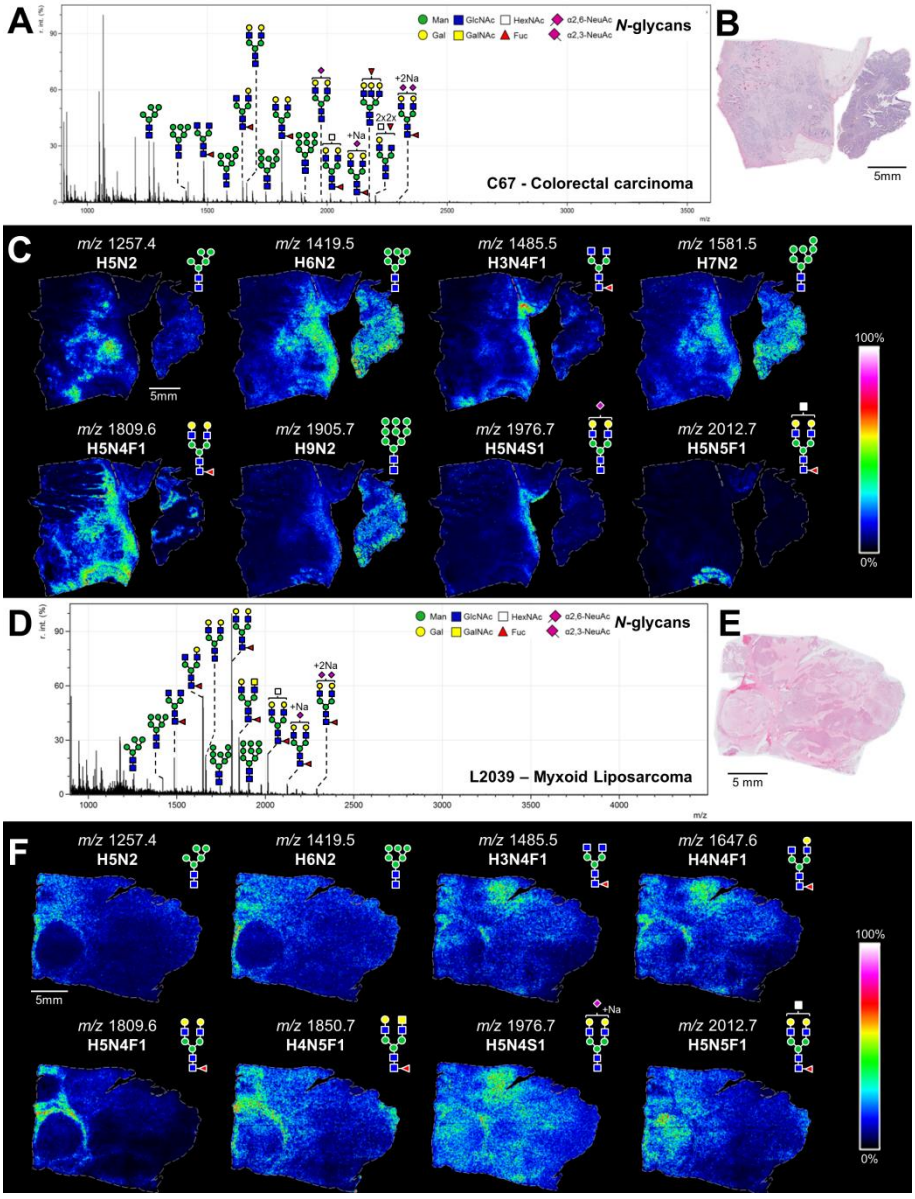


Figure S5-3: (A) The overall average mass spectrum after the *N*-glycan MALDI-MSI analysis of the CRC (C67) tissue. **(B)** Scanned image of the H&E stained CRC (C67) tissue. **(C)** Visualizations of the distribution of various *N*-glycans throughout the tissue. The images were obtained from a TIC normalized dataset. **(D)** The overall average mass spectrum after the *N*-glycan MALDI-MSI analysis of the MLS tissue (L2039). **(E)** Scanned image of the H&E stained MLS (L2039) tissue. **(F)** Visualizations of the distribution of various *N*-glycans throughout the tissue. The annotated glycan structures represent compositions which were either confirmed by MS/MS identification or were known from literature, but structural isomers cannot be excluded. The images were obtained from a TIC normalized dataset.

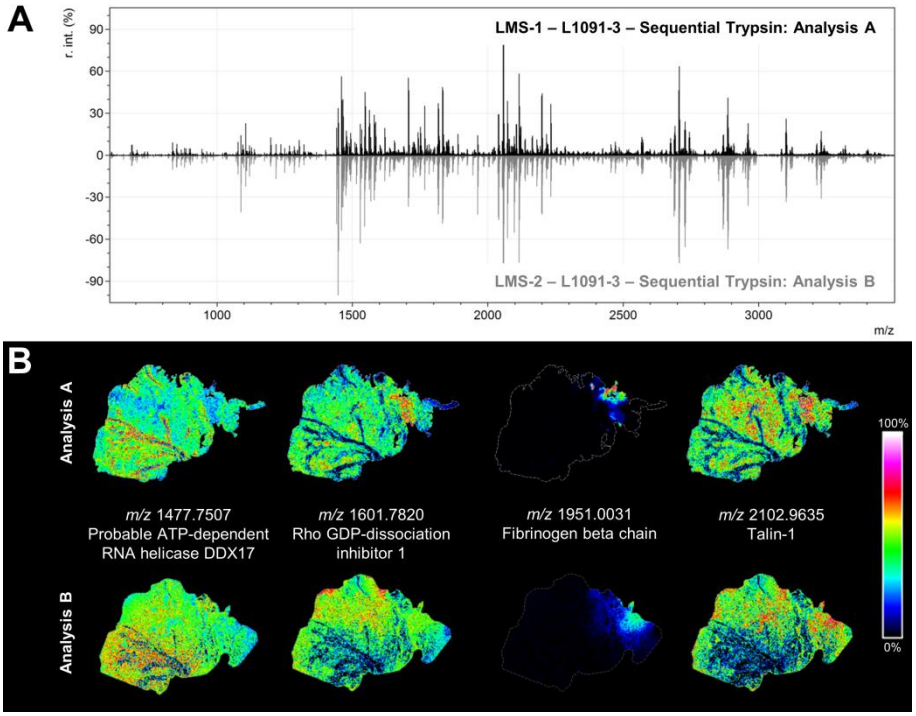


Figure S5-4: (A) A comparison of the overall average spectra obtained from the LMS-1 (L1091-3, black) and, the duplicate, LMS-2 (L1091-3, grey) tissue sections during MALDI-MSI analysis of proteolytic peptides. **(B)** Comparison of the visualizations of specific peptide distributions in both duplicate LMS analyses. The peptides were confirmed to be present in the MALDI-matrix by LC-MS/MS of extracts and were assigned to the MALDI-FTICR-MSI datasets with an error < 10 ppm. The images were obtained from a RMS normalized datasets.

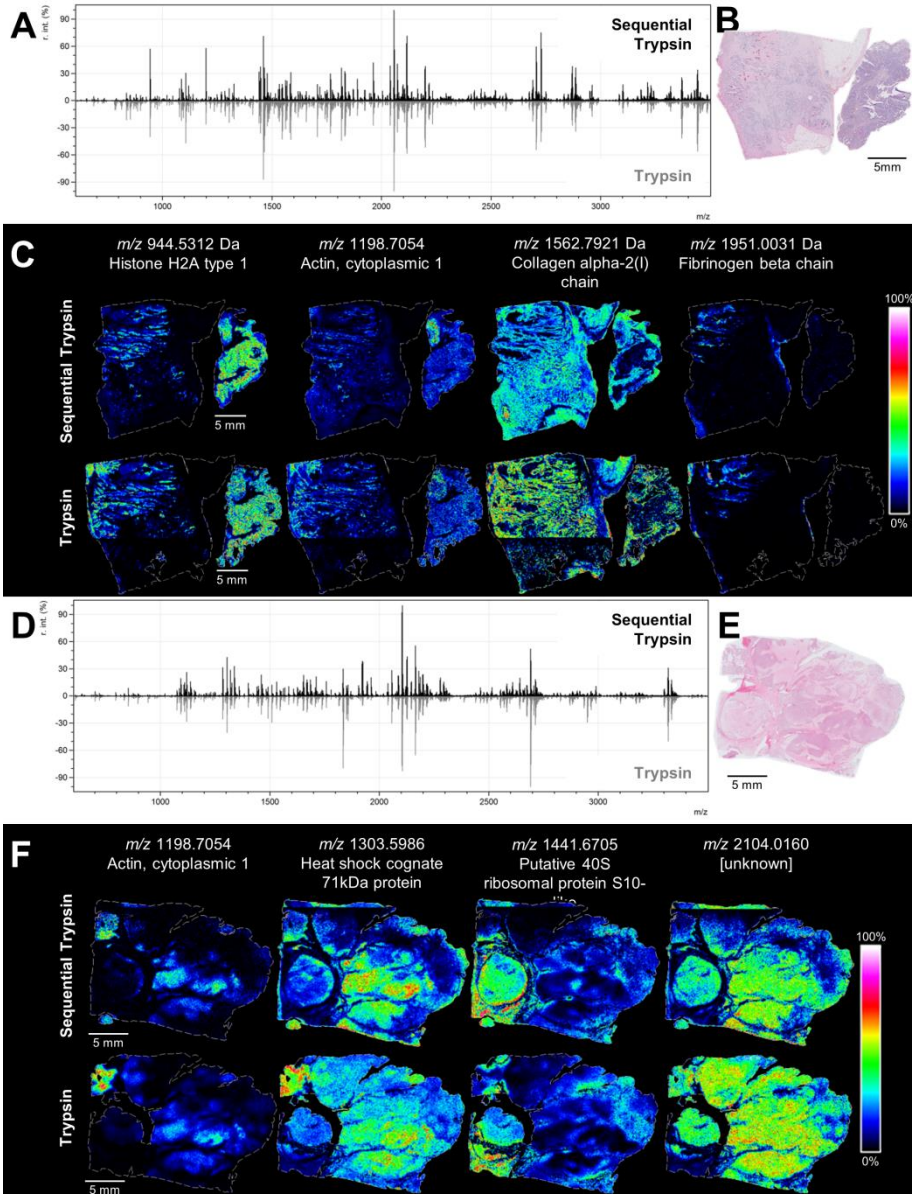


Figure S5-5: (A) A comparison of the profiles of MALDI-FTICR-MSI overall average spectra obtained after both sequential trypsin digestion (top, black) and control trypsin digestion (bottom, grey) of the CRC (C67) tissue. **(B)** Scanned image of the H&E stained CRC (C67) tissue. **(C)** Comparison of the visualizations of specific peptide distributions. **(D)** A comparison of the profiles of MALDI-FTICR-MSI overall average spectra obtained after both sequential trypsin digestion (top, black) and control trypsin digestion (bottom, grey) of the MLS (L2039) tissue. **(E)** Scanned image of the H&E stained MLS (L2039) tissue. **(F)** Comparison of the visualizations of specific peptide distributions. Unless indicated otherwise, the peptides were confirmed to be present in the MALDI-matrix by LC-MS/MS of extracts and were assigned to the MALDI-FTICR-MSI datasets with an error < 10 ppm. The images were obtained from a RMS normalized datasets.

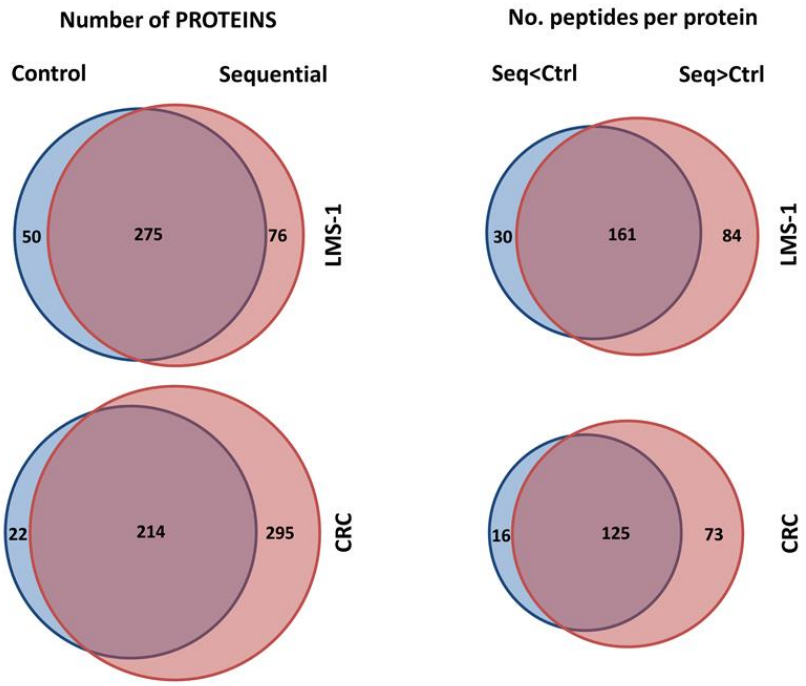


Figure S5-6: Venn diagrams showing the differences and overlap between **(A)** the number of protein identifications in both control and sequential digestions, and **(B)** the differences in the number of identified peptides per protein. The numbers in (B) represent the number of proteins. In blue, the number of proteins for which the number of peptides in the control digestion was larger than in the sequential digestion. In red, the number of proteins for which the number of peptides was higher in the sequential digestion than in the control digestion. The overlapping part represents the proteins where the number of peptides was equal between control and sequential digestions.

5.6.2 Supplementary tables

Table S5-1: (A) An overview of important parameters in the MALDI-TOF-MS method used to obtain *N*-glycan MALDI-MSI data. **(B)** An overview of important parameters in the MALDI-FTICR-MS method used to obtain MALDI-MSI data on proteolytic peptides.

A - Bruker UltrafleXtreme	
Parameter	**Value**
Laser power	70%
Laser shots per pixel	500
Laser focus setting	3_medium
Laser frequency	2000 Hz
Ion Source 1 voltage	25.0 kV
Ion Source 2 voltage	22.5 kV
Lens voltage	8.5 kV
Reflector 1 voltage	26.46 kV
Reflector 2 voltage	13.45 kV (50.8% refl)
Matrix suppression	Deflection up to 890 Da

B - Bruker SolariX XR (9.4T)	
Parameter	**Value**
Laser power	35%
Laser shots per pixel	150
Laser focus setting	Minimum
Laser frequency	2000 Hz
Source optics – plate offset voltage	100.0 V
Source optics – deflector plate voltage	200.0 V
Source optics – capillary exit voltage	150.0 V
Source optics – funnel 1 voltage	150.0 V
Source optics – skimmer 1 voltage	18.0 V
Source optics – funnel RF amplitude	150.0 Vpp
Octopole – frequency	5 MHz
Octopole – RF amplitude	350.0 Vpp
Quadrupole – Q1 mass	1000 *m/z*
Collision cell – collision voltage	-2.0 V
Collision cell – DC Extract bias	0.5 V
Collision cell – RF frequency	2 MHz
Collision cell – collision RF amplitude	1500.0 Vpp
Transfer optics – time of flight	1.7 ms
Transfer optics – frequency	2 MHz
Transfer optics – RF amplitude	300.0 Vpp
Para Cell – transfer exit lens voltage	-20.0 V
Para Cell – analyzer entrance voltage	-10.0 V
Para Cell – side kick voltage	0.0 V
Para Cell – side kick offset voltage	0.0 V
Para Cell – front trap plate voltage	1.5 V
Para Cell – back trap plate voltage	1.5 V
Para Cell – back trap plate quench voltage	-30.0 V
Para Cell – sweep excitation power percentage	45.0%