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

Heijs, B.P.A.M.

Chapter 7

Summary and discussion

The work presented in this thesis describes the improvement and application of on-tissue chemistry for in-situ biomolecular analysis using matrix assisted-laser desorption/ionization mass spectrometry imaging (MALDI-MSI). We have proposed new methodologies, applying on-tissue (enzymatic) chemistry, to increase the molecular information obtained in a MALDI-MSI analysis. We have also developed an automated histology-guided MSI platform, based on state-of-the-art image processing tools, to facilitate high mass and spatial resolution MALDI-MSI while maintaining reasonable data loads and acquisition times. We have shown the importance of these methods in a clinical biomarker discovery study on myxoid liposarcoma tissues. The following section will summarize, discuss, and provide future perspectives on the reported developments.

7.1 Technical developments

7.1.1 Improving and innovating on-tissue chemistry for MALDI-MSI analysis of proteins and post-translational protein modifications

Since its introduction in 2006 by Shimma *et al.* (Shimma et al., 2006) on-tissue digestion has become increasingly popular in the MSI community, owing to its ability to increase proteome coverage and aid the (*in-situ*) identification of peptides/proteins. Moreover, it enables the proteomic analysis of formalin-fixed, paraffin-embedded (FFPE) tissues by MALDI-MSI, thereby greatly increasing the number of patient samples available for clinical MSI studies.

The use of proteolytic enzymes in bottom-up proteomics has been frequently investigated, especially with regard to the repeatability and quantification of the analysis (Brownridge and Beynon, 2011). Only a small number of studies have been conducted into the fundamentals and reproducibility of on-tissue enzymatic digestion, even though a workshop organized by the European MSI network COST Action BM1104 specifically focusing on comparing different on-tissue digestion methods found very little consensus. Diehl *et al.* have previously published a study on the effect of different parameters during on-tissue digestion, including the protease, the incubation time, and the MALDI matrix. However, this study was mostly focused on improving the spatial resolution, and the quality and repeatability of the MSI images (Diehl et al., 2015). Two more recent studies by Erich *et al.* and De Sio *et al.* have compared multiple previously published on-tissue digestion sample preparation methods and tested the relative repeatability of those methods (Erich et al., 2016; Sio et al., 2015). Erich *et al.* proposed an automated, user-independent, computational approach to provide a more objective evaluation of the repeatability of MALDI-MSI (Erich et al., 2016). In **Chapter 2**, we investigated the spatiotemporal dynamics of on-tissue enzymatic digestion. The investigation was based on the analysis of three sample groups with varying incubation times (1,5 h, 3 h, and 18 h), resulted in the finding that longer digestion times improved the repeatability of the on-tissue digestion. Additionally, it was found that differences in the chemical background of distinct morphological regions, specifically

the white and grey matter in mouse brain, resulted in differences in digestion efficiency. These sample preparation artifacts disappeared with the use of longer incubation times, supporting the hypothesis that longer incubation times improve the repeatability of on-tissue digestion experiments by ensuring a more complete digestion for all morphological regions.

The on-tissue digestion dynamics study focused on the global differences between the three incubation times by comparing the number of peaks in the total average spectra. Additional insight into the digestion dynamics was obtained by tracing the digestion of Myelin Basic Protein (MBP, MBP_MOUSE) throughout the tissue and the duration of the digestion. In order to reduce ionization bias and increase the comparability of the different datasets (n = 27, nine replicates per incubation time), three proteolytic fragments of MBP were selected and isotopically-labeled reference standard (ILRS) peptides were synthesized and sprayed homogeneously onto the tissue. The MBP fragments included two 'limit peptides' (end product of the proteolytic digestion), and one 'missed-cleavage peptide'. The fragments were chosen in such a manner that the full digestion of the missed-cleavage peptide resulted in one of the limit peptides, so that relative intensity differences between the two peptides could be used to represent the efficiency of the digestion (with the proviso that other proteins might have different digestion efficiencies and other tissue types might affect the digestion dynamics). Also, it should be taken into account that the results were obtained for analysis from fresh frozen tissues; the dynamics of on-tissue digestion for FFPE tissues may well differ. While this study provides useful insight into the dynamics of on-tissue digestion, additional investigations spanning more proteins, more tissue types and tissue formats would be needed for a complete understanding of the factors affecting the dynamics of on-tissue digestion. Nevertheless the methodology reported here, to obtain quantitative data of digestion rate (for one cleavage site) from different tissue regions, presents the methodology that should be expanded for all future investigations.

The extensive work performed in the field of proteomics over the last two decades is an invaluable resource for the improvement and adaptation of MALDI-MSI methods. As described above, adaptation of the bottom-up proteomics approach led to the introduction of on-tissue digestion MALDI-MSI. As is the case in LC-MS/MS-based bottom-up proteomics, on-tissue proteolytic digestion is predominantly performed using trypsin. Due to its cleavage specificity, C-terminal of the basic amino acid residues arginine and lysine, trypsin produces proteolytic peptides with intrinsically positively charged C-termini that are detected with high sensitivity by mass spectrometry (Brownridge and Beynon, 2011; Vandermarliere et al., 2013).However, previous work in the proteomics field has also concluded that the use of different enzymes (even with a shared cleavage specificity) can lead to increased proteome coverage (Choudhary et al., 2003; Hohmann et al., 2009). In **Chapter 3** and **Chapter 5** we describe two different multi-enzyme approaches; **Chapter 3** describes the combination of several

proteolytic enzymes with lysine and arginine cleavage specificity, and how the use of different enzymes could increase proteome coverage as well as protein sequence coverage; **Chapter 5** describes a multimodal MSI approach, in which sequential on-digestions were performed with the glycosidase PNGase F and trypsin in order to analyze both *N*-linked glycans and proteolytic peptides in two sequential MALDI-MSI analyses, but from the same FFPE tissue section. While the methodology proposed in **Chapter 3** is dependent on the analysis of multiple tissue sections (although the analysis with the combination of trypsin/Lys-C is based on the same principle), the methodology in **Chapter 5** requires only a single tissue section for the acquisition of two datasets on two closely related molecular classes. The benefit is that the information obtained in both datasets originates from exactly the same morphological and histological features. It should be noted that the MALDI-MS analysis of *N*-glycans can be improved further by performing a linkage-specific sialic acid derivatization. This derivatization stabilizes the otherwise labile sialic acid residues, enables the differentiation between the isobaric α2,3- and α2,6-linked sialic acids, and reduces the adverse effects on positive-mode ionization of the negative charge of native sialic acids (Holst et al., 2016). A potential complication is the dimethylamidation derivatization reaction modifies all free carboxylic acid (COOH) groups, and therefore also affects protein C-termini and acidic amino acid residues. The exact derivatization-induced changes of the proteins were not characterized in detail, but which could complicate the identification of proteolytic fragments from the same FFPE tissue section. Nevertheless, the sequential on-tissue digestion approach has proven itself a useful tool for *N*-glycan MALDI-MSI as is described in **Chapter 5**.

7.1.2 Histology-guided MALDI-MSI

Recent technological advances, mainly the development of an ultrafast MALDI-TOF-MS platform, and resolution for MALDI-FTICR-MS analysis, have made high spatial resolution and high mass resolution analysis readily available for MSI. However, the increased cellular and chemical specificity achieved by the high-resolution analysis come at the cost of data acquisition time and data load. Especially in the field of clinical MSI where large patient tissue series need to be analyzed, this imposes challenges on both the MSI infrastructure and data management systems. To overcome these challenges, we have developed a histology-guided MSI (HG-MSI) platform based on an automated image registration pipeline in **Chapter 4**. The image registration pipeline registers a high-resolution annotated histological image of a tissue section to a lower resolution image of an adjacent tissue section, prepared for MALDI-MSI analysis. Following the image registration, the annotation borders are reproduced on the low-resolution image, enabling the exclusive MSI analysis of the *a priori* selected regions of interest. In a study on a small tissue cohort the HG-MSI approach resulted in a reduction of data acquisition time and data load of 80%, and thereby improving the applicability of high spatial and mass resolution MSI for clinical research.

The prior histological specification of the HG-MSI approach is well suited for applications in which an *a priori* histological specification is possible, meaningful, and needed, e.g. the analysis of tumor progression based on cellular dedifferentiation or other specific morphological hallmarks. The tool is especially useful for those applications where high spatial resolution, or high mass resolution is required, but at the same time impractical or limited by extremely large pixel counts and/or slow data acquisition speeds. **Chapter 4** describes the application of the HG-MSI platform for the analysis of proteolytic peptides from FFPE tissues using an ultrahigh mass resolution MALDI-FTICR-MS instrument manufactured by Bruker Daltonics. To ensure this tool can become a valuable asset for the MSI community it is vendor neutral, applicable to all MSI modalities (e.g. SIMS, DESI, LESA, etc.), and freely available online.

The analysis of tissue microarrays (TMA), as described in **Chapter 6**, offers a solution for the analysis of a large patient series. A TMA is an assembly of needle core biopsies (diameter 0.5 – 2 mm), obtained from histologically predefined tissue regions from many patients. The major advantages of analyzing TMAs are that a large number of patient samples can be rapidly compared; the experimental variance is reduced substantially as all needle cores are prepared during the same sample preparation and analyzed together; and a substantial histological pre-selection of tissue regions has already taken place.

7.2 Clinical application

7.2.1 The molecular background of tumor progression in myxoid liposarcoma

The methods described in this thesis, and especially those in **Chapters 4** and **5**, were ultimately designed for their application in clinical research. **Chapter 6** describes the application of the multimodal MSI method described in **Chapter 5** in a study focusing on the molecular characterization of tumor progression in myxoid liposarcoma (MLS). We have demonstrated an association between a higher abundance of high-mannose type glycans, and highly branched complex-type glycans with MLS tumor progression. While this phenomenon was not previously described for MLS or any other mesenchymal tumor, it was described for several other tumor types, including colorectal cancer, breast cancer, and ovarian cancer. High-mannose type glycans are found both intra- and extracellular. However, the MALDI-MSI experiments in **Chapter 6** were not performed at a spatial resolution needed to discriminate between intra- or extracellular molecules. Therefore, the results presented in **Chapter 6** should be validated with independent techniques such as immunohistochemistry to provide additional information on the exact location of the *N*-glycans in the cellular environment. Further insights into the changes associated with MLS progression could be garnered by also including protein and lipid MSI data as well as an in-depth analysis of the proteome, to further unravel the relevant pathways and molecular alterations leading to the advanced stages of MLS.

7.3 Future work

This thesis describes the optimization and application of on-tissue digestion MALDI-MSI to improve its applicability in clinical research. Since the introduction of MSI, the identification of detected molecules has been one of the major challenges. Especially in clinical studies and/or biomarker discovery studies, the identity of the detected molecules is paramount to obtain valuable biological insights about the nature of the potential MSI-based biomarkers. One of the advantages of applying the on-tissue digestion method to *in situ* protein analysis is that it aids protein identification. The work described in the thesis shows that the number of protein identities that could be assigned to *m/z* features increased substantially with high mass resolution MSI. However, it should be noted that approximately only 10-15% of all peaks in the MSI average spectra were assigned with a peptide identity, and which did not include many of the high intensity peaks. The solution to improve the peptide identification rate in MSI is multifaceted.

First, in FFPE tissues, formalin is used to cross-link proteins and does so by connecting the basic amino acids lysine, arginine & histidine within and between proteins. For the analysis of proteins in FFPE tissue, formalin-fixation is reversed by a process called antigen retrieval. However, the reversal of the formalin fixation is often not complete and results in irregular, and predominantly uncharacterized remnants of the formalin fixation attached to the peptides/proteins. This results in both a large number of unmatched MS/MS spectra during the database search following LC-MS/MS analysis, and a large number of unassigned peaks in the MSI average spectrum. In order to utilize the full potential of proteomic MSI from FFPE tissue a thorough characterization of the formalin-induced modifications should be performed. However, previous attempts, and the fact that the majority of the bottom-up proteomics field remains distant from FFPE tissue, indicate the enormous challenges it encompasses. Alternatively, targeted offline LC-MS/MS approaches, based on MSI peak lists, could aid in additional identity assignments.

Second, the digested proteome with its large number of (near-)isobaric proteolytic fragments, is an extremely complex sample and results in a very complex mass spectrum containing many overlapping isotope envelopes. The spectra are further complicated by the presence of adducts. Tissues are a salt-rich environment, and although the tissues are washed during sample preparation, various salt-adducts (Na^+, $2Na\text{-}H^+$, K^+, $2K\text{-}H^+$) are commonly detected in MSI analyses, as is described in **Chapter 5**. While the removal of adduct peaks from the mass spectra is a possible solution to simplify the spectra, the preferred strategy would be to reduce adduct formation during the analysis by applying additional washes, or using a more salt resistant MALDI matrix that reduces the adduct formation. Wang *et al.* recently reported that the matrix (E)-propyl α-cyano-4-hydroxyl cinnamylate, formed by the reaction of the regular peptide MALDI matrix α-cyano-4-hydroxycinnamic acid with

propyl alcohol, displayed high salt tolerance for intact protein analysis (Wang et al., 2015b). We synthesized this matrix but preliminary MALDI-MSI experiments using on-tissue digestion revealed the matrix exhibited an increased propensity to form salt adducts.

Third, the peptide identity assignments described in **Chapter 3** and **Chapter 5** of this thesis were all performed on the basis of accurate mass. While the use of this method provides a lot of valuable information on the potential identities of the detected molecules, the preferred method would involve the acquisition of their tandem MS spectra directly from the tissue section. While this approach is commonly described in literature using low mass resolution MALDI-TOF-MS platforms, its availability for high mass resolution MS systems is very limited. To improve the true peptide identification in MSI, mass spectrometers should be equipped with improved quadrupole technology, allowing a more specific selection of *m/z* features for fragmentation, leading to a more accurate identification directly from tissue.

The additional work described above would improve our basic understanding of the samples we work with, which will ultimately result in more specific and more efficient methods better suited for the MSI analysis of FFPE tissues. Nevertheless, the impact of MSI on clinical research rapidly increased over the last years, improved methodology and technology will ascertain further establishment of MSI in clinical research, and might result in an implementation of MSI in molecular diagnostics in the future.