

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Trends in Sample Preparation for Proteome Analysis

Jakub Faktor, David R. Goodlett and Irena Dapic

Abstract

Sample preparation is a key step in proteomics, however there is no consensus in the community about the standard method for preparation of proteins from clinical samples like tissues or biofluids. In this chapter, we will discuss some important steps in sample preparation used for bottom-up proteome profiling with mass spectrometry (MS). Specifically, tissues, which are an important source of biological information, are of interest because of their availability. Tissues are most often stored as fresh frozen (FF) or formalin-fixed paraffin-embedded (FFPE). While FF tissues are more readily available, paraffin embedding has historically been routinely used for tissue preservation. However, formaldehyde induced crosslinks during FFPE tissue preservation present a challenge to the protocols used for protein retrieval. Moreover, in our view, an important aspect to consider is also the amount of material available at the start of a protocol since this is directly related to the choice of protocol in order to minimize sample loss and maximize detection of peptides by MS. This “MS sensitivity” is of special importance when working with patient samples that are unique and often available in limited amounts making optimization of methods to analyze the proteins therein important given that their molecular information can be used in a patients’ diagnosis and treatment.

Keywords: sample preparation, tissue, digestion, mass spectrometry, proteomics

1. Introduction

Proteomics is an important tool in the study of human biological material with the aim to extract knowledge that can improve a patients’ treatment outcomes. Molecular information obtained from patient samples can be complementary to pathological observations all with the goal of faster and more accurate diagnosis, and subsequent treatment. Molecular analysis of tissue by proteomics can lead to disease classification and reveal underlying disease pathways that can further serve as a target for medical treatment.

Sample size and origin is an important aspect in sample preparation. Today, there are numerous sample preparation procedures existing which aim to improve sensitivity of detection or protein recovery from a sample. Release of proteins from native or artificial material is a crucial step in sample preparation and to improve protein recovery there are different additives such as detergents, chaotropes, buffers and salts added during the sample preparation that must be considered. Moreover, targeting special groups of proteins (e.g. membrane proteins), which are involved in key cellular functions and may be a target of pharmaceutical treatment, often represent a

challenge in their isolation and analysis. Their amphipathic nature may require use of appropriate enrichment procedures all with the goal to achieve better detection.

Further, sample loss during most standard preparation procedures is inevitable, and it is even more accentuated when minute amounts of material are being processed. To minimize sample loss and thus increase sensitivity of the analysis at the MS step there have recently been several technologies developed. Specifically, improvement in technologies that allow detection of proteins down to a single cell have become available. Some of these technologies, such as nanoPOTS and microPOTS, have already been applied to human tissues. These new possibilities to analyse small regions of tissue samples with sufficient sensitivity is opening the door to many applications such as profiling of selected regions of a tumorous zone or detection of proteins from subcellular populations. These new applications aimed at working with 1 to 100s or 1000s of cells will likely have increasing importance in clinics, but only if they can be developed into routine and robust methods.

2. Tissue preservation

Human tissue samples are a valuable source of information for diagnostics, therefore a lot of effort has gone into best preservation methods that minimize changes that can occur over time in storage. For example, following clinical surgery tissues need to be stored according to the protocols that minimize chemical, enzymatic, mechanical or thermal degradation and protect their molecular content. Today, tissues are most often preserved as fresh frozen (FF) or formalin-fixed paraffin-embedded (FFPE) tissues.

2.1 Fresh frozen tissues

FF tissues are obtained usually with snap freezing of tissues where the temperature achieved is below -70°C , most often in dry ice or liquid nitrogen (**Figure 1B**). To minimize variability between the samples storage and thus to minimize potential effect on molecular structure and integrity of the tissue, the European Human Frozen Tumour tissue bank (TuBa-Frost) has standardized tissue preservation by freezing in 2006 [1, 2]. An important aspect in preservation of tissues by the FF method is prevention of formation of artefacts that might result in changes to the tissue structure and morphology. For example, ice crystals that can disrupt structures within the tissue may form as a consequence of the freezing procedure due to moisture present within the tissue [3]. An alternative to snap freezing is the optimal cutting temperature (OCT) compound, which is used for tissue embedding and contains polyvinyl alcohol, polyethylene glycol (PEG) and benzalonium chloride. The OCT substance preserves tissue and enables optimal microdissection of the tissue. However, where samples will later be analyzed by mass spectrometry (MS), OCT compounds must be removed prior to analysis. This is usually achieved by washing the tissue with a special grade of alcohol or Carnoy's fluid [4] or with the use of other protocols for sample purification.

2.2 Formalin-fixed paraffin-embedded tissues

An alternative to preservation of tissue by the FF process is the use of FFPE methods (**Figure 1A**), which are routinely used by pathologists around the globe to preserve tissue by embedding in paraffin. The FFPE process preserves tissues by chemical fixation most often in 10% of formalin and is followed later by embedding in paraffin to form a tissue block for subsequent slicing. The combination of formalin fixation with paraffin embedding allows for long term storage of tissues.

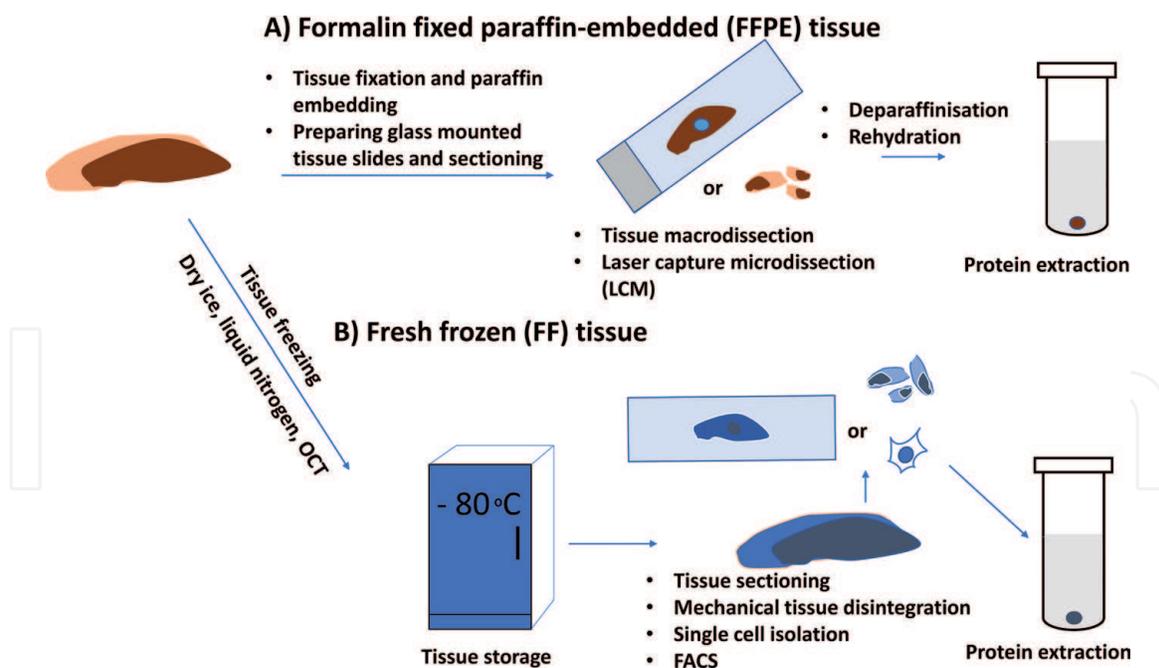


Figure 1.

Overview of tissue sample processing prior to proteomic experiments. A) Tissues can be preserved and stored for long periods of time by formalin fixation and paraffin embedding which has been routinely used for decades. Proteomic analysis of FFPE tissues can be accompanied with laser capture microdissection (LCM) which helps to retrieve regions of interest on tissue sections. Further steps involve isolation of the proteins from the sections in appropriate lysis buffers and further processing for protein analysis by MS. B) Tissues can be preserved by freezing and further stored at low temperatures. Tissue should be frozen as soon as possible after retrieval and tissue can be sectioned prior to proteomic sample preparation. Tissue sections are further prepared for protein isolation, but also single cell isolation protocols could be employed to retrieve limited cell subsets prior protein extraction.

Also, FFPE tissues are often used for histopathological studies, a routine process in examination of a patient's biopsies and clinical material [5]. Moreover, it is known that formalin leads to chemical modification of proteins in the fixed tissues causing cross-linking between proteins and modifications most often as methylation (+14 Da) as well as formation of methylene and methylol adducts to a lesser extent. As a consequence of formaldehyde induced modifications, the molecular weight or physicochemical properties of fixed proteins can be altered.

3. Preparation of the sample for bottom-up proteomics

Protein extraction and the subsequent preparation for LC-MS analysis represents one of the key steps in proteomics (Figure 2). While there have been numerous protocols reported, they have mainly focused on preparation from large amounts (i.e. micrograms to milligrams) of material, which limits their utility in the study of patient clinical samples. Notably, protein extraction from FFPE preserved tissues requires removal of formaldehyde-formed cross links, which is usually carried out by heating samples in a buffered solution at an elevated temperature (95°C or 100°C). The most common buffers used for protein extraction are ammonium bicarbonate, tris(hydroxymethyl)aminomethane (Tris), and Radioimmunoprecipitation assay (RIPA) buffer. Addition of detergents to the buffer composition (e.g. sodium dodecyl sulfate (SDS), sodium dodecyl cholate (SDC), RapiGest SF surfactant™ (Waters), PPS Silent Surfactant™ (Expedeon) have been routinely used to improve protein solubilization efficiency and thus enhance protein extraction. In addition to optimization of the extraction buffers many studies also optimized other parameters like incubation time of the extraction and/or addition of various proteases to improve protein coverage during subsequent LC-MS/MS analysis.

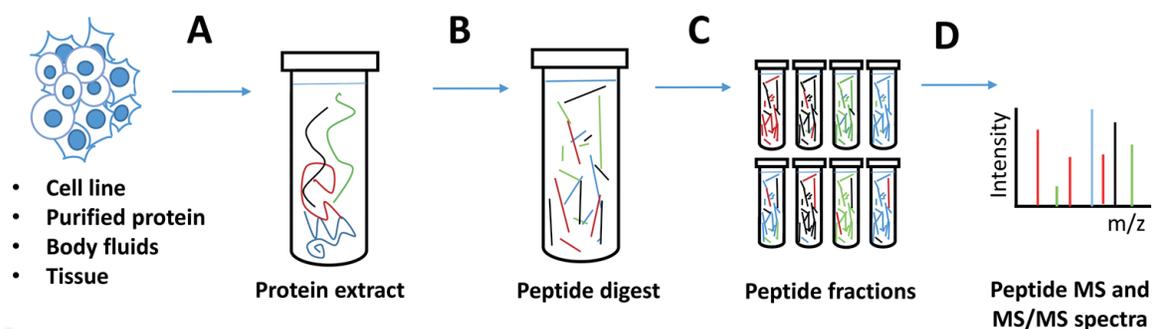


Figure 2.

Overview of sample preparation for bottom-up proteomic analysis by tandem mass spectrometry. A) Sample lysis: proteins are extracted from biological matrix in lysis buffer. Mechanical disintegration or sonification is used to homogenize rigid structures present within samples such as is common in mammalian tissue. B) Protein digestion: proteins are proteolytically digested into peptides, usually by the protease trypsin. C) Peptide fractionation: optionally, the complexity of the peptide sample is decreased by addition of fractionation steps orthogonal to methods used in the next step. D) Mass spectrometry analysis: desalted peptide samples are dissolved in an appropriate buffer and introduced into a tandem mass spectrometer. Most often reversed phase liquid chromatography separation is used in this final step to enable sequential introduction of peptides into the tandem mass spectrometer.

3.1 Detergents

Traditional detergents and chaotropes such as SDS and urea have been widely used for protein solubilization, however they are also well known to inhibit digestion at higher concentrations and are incompatible with reversed phase chromatography separation (RPLC) used to introduce samples for MS analysis. Therefore, their concentration must be kept low at the time of proteolysis in order to preserve the effectiveness of proteases used for protein digestion. Failure to do so often leads to incomplete protein solubilization and denaturation. Also, presence of detergents in the sample might interfere with later instrumental analysis, therefore there have been different purification methods developed for detergent removal to improve LC–MS outcome. The choice of the most effective procedure depends on the physicochemical properties of the detergent. Some of the procedures might include detergent removal on the basis of size exclusion (i.e. molecular weight cut-off filters) or with the use of spin columns containing appropriate resins for detergent removal. Moreover, heating of the sample in urea buffers often leads to covalent modification of proteins via carbamylation, which might affect peptide retention time during RPLC separation and if not accounted for will interfere with identification. In order to circumvent these problems caused by mass spectrometry incompatible detergents significant effort went into development of reagents that avoid these complications. To this end, acid labile detergents such as RapiGest SF surfactant™ (Waters) and PPS Silent Surfactant™ (Expedeon) were developed that could be easily removed after proteolysis by simple measures like decreasing the pH. For example, the MS compatible surfactant ProteaseMAX™ (Promega) surfactant enhances tryptic, chymotryptic and LysC digestion and then degrades during the course of a digestion reaction. Another compound, Invitrosol™ (Thermo Fisher Scientific) is a homogenous surfactant that does not impact tryptic digestion and elutes during RPLC in three peaks well separated from where peptides elute [6].

3.2 Sample digestion

Classical bottom-up proteomic sample preparation aims to turn protein extracts into peptides via a process of protein cleavage or digestion with proteases. Notably, proteins extracted from biological material tend to keep their native tertiary structure mostly held by non-covalent interactions of amino acid side groups [7].

It is thus essential to disrupt the tertiary structure and linearize the protein sequence to ease the accessibility of proteases to cleavage sites. Protein tertiary structure is frequently disrupted by chaotropic and denaturing reagents. Disulfide bonding contributes to tertiary structure as well via a covalent bond between cysteine side chain groups also termed an S-S bridge. Disulphide bonds are most often broken by use of reducing agents leaving free sulfhydryl groups available that allow the protein to unfold more fully. Dithiothreitol (DTT), tris (2-carboxyethyl) phosphine (TCEP), tris (3-hydroxypropyl) phosphine (THPP) and 2-mercaptoethanol (2-ME) are the most commonly used reducing agents. Sulphur containing reagents such as 2-ME and DTT break the S-S bridge by thiol-disulfide exchange, while phosphorus containing reagents form a phosphine oxide as a result of disulphide bond reduction [8]. Reduction is commonly followed by free sulfhydryl group alkylation to prevent disulphide bond reformation. In this chemistry a free sulfhydryl group performs a nucleophilic attack on the alpha carbon of an alkylating reagent creating a covalent bond between the alkyl group and cysteine. There is a wide palette of alkylating reagents that may be used, but in proteomic sample preparation the most commonly used reagents include iodoacetamide, iodoacetic acid, N-ethylmaleimide (NEM) and S-methyl methanethiosulfonate. Covalent modification of a free sulfhydryl group leaves a mass tag on each cysteine that must be considered as a mass shift to cysteine during interpretation of peptide tandem mass spectra. Alkylated proteins are then further processed by proteolytic cleavage, to shorter segments; peptides, which are then easily detected in a bottom-up experiment carried out by LC-MS/MS analysis. As mentioned above peptides may be produced by enzymatic methods but also chemical methods that can be either specific or unspecific (**Table 1**). In both cases there are a variety of protocols available to digest proteins into peptides for mass spectrometry-based proteomic analysis.

Bottom-up proteomics frequently relies on proteolytic enzymes that digest a protein at specific sites. Having predictable digestion rules for a given protease results in a faster database search process that also makes it computationally less demanding and more accurate. Trypsin is the most common protease in bottom-up proteomics cleaving peptide bonds at the C-terminus of arginine and lysine when not followed by proline [16]. Notably, maintaining an optimal temperature of 37°C at a pH optimum between 7 and 8 in the presence of Ca²⁺ ions in the digestion buffer is important for the reaction to proceed efficiently [17]. The optimal enzyme to substrate ratio is also important and for trypsin this is often from 1:20 to 1:100 (w:w). In some instances LysC endoproteinase, which is isolated from *Lysobacter enzymogenes*, is often combined with trypsin to provide cleavage at lysine C-terminus. This combination of multiple enzymes is used to enhance peptide sequence coverage by producing overlapping peptides. The addition of chymotrypsin and pepsin produce the most orthogonal peptides to trypsin. Chymotrypsin is a serine protease which cleaves a peptide bond at the C-terminus of large hydrophobic side chain amino acids such as phenylalanine, tryptophan, tyrosine and leucine. Chymotrypsin performs best in a 1:50 (w:w) enzyme to substrate ratio at basic pH and a temperature around 37°C. Chymotrypsin is also activated and stabilized by the presence of Ca²⁺ ions, therefore it is beneficial to use digestion buffers containing calcium ions (e.g. CaCl₂) [18]. Pepsin is an endopeptidase that is secreted in gastric chief cells as an inactive precursor called pepsinogen that becomes activated by cleavage of an N-terminal pro-segment in acidic conditions. The optimal enzymatic activity of pepsin is achieved at pH 1.5–2.5 and 37°C. Pepsin cleaves at the C-terminus of phenylalanine, leucine and rarely after histidine and lysine, unless they are adjacent to leucine or phenylalanine. Pepsin is frequently used for on-column protein digestion in hydrogen-deuterium exchange experiments (HDX), but also an application in off-line pressure assisted protein digestion has been reported [19].

Protease	Class	pH range/ion	t [°C]	Cleavage specificity	Example application	Reference
Trypsin	Serine	7–8/Ca ²⁺	37	Arg, Lys (C-term)	Primary central nervous system lymphoma	[9]
LysC	Serine	8.5	37	Lys (C-term)	Whole liver SDS lysates	[10]
LysN	Metalloproteinase	7-9/Zn ²⁺	Thermostable	Lys (N-term)	HEK 293 cells	[11]
Chymotrypsin	Serine	8/Ca ²⁺	37	Hydrophobic AAs (C-term)	Cerebrospinal fluid (CSF)	[12]
Pepsin	Aspartic	1.5-2.5	37	Preferentially Phe, Leu (C-term)	Human liver tissue	[13]
Thermolysin	Metalloproteinase	5.0–8.5 / Zn ²⁺	65–85	Ala, Met, Ile, Leu, Val, Phe (N-term)	Human liver tissue	[13]
AspN	Metalloproteinase	6.5–8.0 / Zn ²⁺	40	Asp (N-term)	Brain and liver tissue from C57BL/6 mouse	[14]
GluC	Serine	4.0, 7.8	37	Glu, Asp (C-term)	Cerebrospinal fluid (CSF), brain and liver tissue from C57BL/6 mouse	[12, 14]
ArgC	Cysteine	7.2–8.0/Ca ²⁺	37	Arg, Lys (C-term)	Cerebrospinal fluid (CSF), brain and liver tissue from C57BL/6 mouse	[12, 14]
CNBr	Chemical	—	—	Met (C-term)	Extracellular matrix of human mammary and liver tissue	[13, 15]

Table 1.

Proteases used for proteolytic digestion of protein extracts retrieved from biological material such as tissue, body fluids or cell extract. Table 1 presents the enzyme class, pH and temperature optimum, inorganic ion cofactor and specificity of protease. In addition a representative application and literature source is given.

GluC, ArgC, LysN, AspN are also popular proteases in bottom-up proteomics as they predictably produce complementary or orthogonal peptides to trypsin with different substrate affinities. GluC is a serine protease isolated from *Staphylococcus aureus* with specificity dependent on the digestion buffer composition. For example, performing proteolysis in phosphate buffers will lead to cleavage at the C-terminus of glutamic acid and aspartic acid, but only cleavage at the C-terminus of glutamic acid will be catalysed in ammonium acetate (pH 4.0) and ammonium bicarbonate (pH 7.8) buffers [20]. GluC is known to perform optimally under pH 4.0 and pH 7.8 at 37°C while it is stable in denaturing conditions. ArgC, isolated from *Clostridium histolyticum*, is a cysteine endopeptidase cleaving at the C-terminus of arginine and sometimes at the C-terminus of lysine. Its pH optimum is 7.6 and Ca^{2+} ions also enhance its activity. ArgC digestion has recently been considered an alternative to the conventional trypsin digestion as it cleaves at the C-terminus of arginine. LysN is a metalloprotease that cleaves at the N-terminus of lysine and it is resistant to denaturation allowing digests to proceed even at temperatures higher than mentioned above. AspN is a selective metalloproteinase isolated from *Flavobacterium meningosepticum* requiring zinc atoms for its catalytical activity [21]. Its endopeptidase activity is specific to the N-terminus of aspartic acid or cysteic acid. To maintain optimal enzymatic activity it is recommended to include ZnSO_4 in the digestion solution buffered between pH 6.5–8.0 at a temperature of 40°C. Combining AspN with trypsin digestion increases data quality and increases protein coverage [22]. WaLP and MaLP are less known proteases cleaving at aliphatic amino acids, which makes them popular for membrane proteomic applications. Meyer et al. demonstrated that combination of data from trypsin, LysC, WaLP and MaLP digestion leads to an increase in membrane proteome coverage by 101%, compared to coverage achieved by trypsin digestion alone [23].

Broad specificity protease digestion is less common to bottom-up sample preparation, nevertheless it is used to digest rigid protein structures that resist digestion using common proteases. Proteinase K is one such serine endopeptidase isolated from fungus *Engyodontium album* that cleaves protein sequences with a broad specificity and like others discussed above requires Ca^{2+} ions for activity. Generally, it cleaves at the C-terminus of aromatic or aliphatic amino acids and is able to digest proteins in their native state or in presence of detergent such as SDS and Triton-X 100, but works best at alkaline pH 7.5–12.0 and 37°C. Most frequently, it is used for nucleic acid purification to remove proteins, but it is also suitable for some proteomic applications such as non-specific digestion of membrane proteins, protease footprinting or prion digestion. As the name implies thermolysin is a thermostable metalloproteinase isolated from *Bacillus thermoproteolyticus*. Thermolysin requires zinc and calcium ions for proteolytic activity but remains active in temperatures from 65–85°C and between pH 5.0 to 8.5. It cleaves at the N-terminus of alanine, methionine, isoleucine, leucine, valine and phenylalanine and is often used to digest proteins that resist proteolysis using conventional proteases [24]. Papain and elastase have endopeptidase activity and broad specificity that while available are rarely used in bottom-up sample preparation. Elastase is a serine endopeptidase that cleaves at the C-terminus of small hydrophobic side chains such as glycine, valine, isoleucine and leucine. While, papain is cysteine endopeptidase that cleaves at the C-terminus of arginine and lysine if it is preceded by hydrophobic amino acid, but not succeeded by valine. Subtilisin is a serine endopeptidase isolated from soil bacteria (e.g. *Bacillus licheniformis*) that is known to non-specifically cleave the peptide bond with a preference for large uncharged amino acids, although amino acids with basic side chains can be accepted in alternate binding mode [25]. Subtilisin remains active and stable under denaturing and alkaline conditions ranging from pH 8–12 and Ca^{2+} ions stabilize subtilisin structure, therefore it is essential to include CaCl_2

in a digestion buffer. Subtilisin's use in bottom-up proteomics is quite limited due to its wide range of specificity, nevertheless it has been reported that it could be used to reveal previously hidden areas of the proteome [26]. Cathepsins form a large group of proteases with endopeptidase activity. Their use in proteomics is not frequent but nevertheless some uses have been reported. Cathepsin L is a cysteine protease located in lysosomes, it is physiologically involved in tissue remodeling and in diseases such as cancer metastasis. Cathepsin L is catalytically active at pH 3.0–6.5 in the presence of thiol compounds [27]. Digestion using Cathepsin L has been reported in research of histone N-termini. Cathepsin C is a N-terminal dipeptidase physiologically involved in activation of serine proteases and inflammatory cells [28]. Its use in proteomic sample preparation is limited, as its cleavage is unspecific. Nevertheless, it could serve as a potent tool to generate peptides orthogonal to conventional proteases.

Thrombin is a serine protease which is proteolytically activated during the clotting process from an inactive prothrombin precursor. It is exclusively specific towards the Leu-Val-Pro-Arg-Gly-Ser motif. Therefore, it is most often used to cleave a specific linker tethered to another peptide with this sequence motif inserted into recombinant fusion protein constructs. There is a wide palette of these type of protein tag removal endopeptidases; namely Factor Xa cleaving Leu-Val-Pro-Arg-Gly-Ser motif, Enteropeptidase cleaving Asp-Asp-Asp-Asp-Lys motif, TEV Protease cleaving Glu-Asn-Leu-Tyr-Phe-Gln-Gly motif, Rhinovirus 3C Protease cleaving Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro motif and several others [29]. Further details of protein tag removal proteases will not be discussed as it does not fall within scope of this chapter.

Finally, it should be noted that reproducible protein cleavage could be achieved even in non-enzymatic reactions mediated by chemical reagents. The most frequent chemical reagents to cleave peptide bond are dilute acids, such as hydrochloric acid, formic acid and acetic acid or other reagents such as cyanogen bromide (CNBr), hydroxylamine and 2-nitro-5-thiocyanobenzoate (NTCB) [30]. Exposure of proteins to dilute acids results in kinetically favored cleavage of peptide bonds at asparagine but with time others as well, while CNBr cleaves at less abundant methionine [31]. NTCB is specific towards cysteine, while hydroxyl amine reagent cleaves peptide bonds at asparagine and glycine. Generally, chemical mediated cleavage targets peptide bonds of less common amino acids producing long peptides useful in middle-down proteomics [30].

4. Technologies for analysis of limited sample amounts

Given that there is no technology to amplify proteins as may be done for nucleic acids with polymerase chain reaction, historically proteomics has faced limitations in terms of the amount of starting material required for success. Traditional proteomics approaches to sample preparation such as filter-aided sample preparation (FASP), in-gel digestion, and in-solution digestion typically require at least several micrograms of a protein sample, which can be complicated to retrieve from representative clinical samples that are by default limited in availability. Therefore, the traditional method of defining proteomes has generally produced knowledge on the underlying biology that reflect averages rendered from analysis of mixtures of cells of different types present in tissue.

As proteomics and the requisite mass spectrometry instrumentation have evolved, microscale proteomic pipelines that decrease the amount of protein required to sub-microgram levels have become available. Microscale proteomics pipelines rely on modifications of traditional proteomics pipelines frequently

accompanied with cell sorting, laser capture tissue microdissection (LCM) or single cell extraction methods. Microdevices such as nano-capillary columns, microfluidic chips, miniaturised ESI introduction interfaces and miniaturised enzyme reactors are often required [32]. Introducing microscale proteomics provides a clearer picture of reality as it substantially increases sensitivity, spatial proteome resolution and leads to better understanding of how protein networks coincide on microscopic level. Despite obvious benefits, microscale proteomics still requires special instrumentation making implementation of these protocols for the moment some what difficult across laboratories worldwide.

One recent promising such technology is nanoPOTS (nanodroplet processing in one pot for trace samples) (Figure 3A). The nanoPOTS platform is intended for processing small cell populations in nanoliter volumes. NanoPOTS benefits from downscaling the processing volumes that in turn substantially reduces surface associated sample losses. The final step of nanoPOTS is accompanied with solid phase extraction (SPE) that concentrates, desalts and efficiently introduces a sample to nanoLC fluidics. Recently, a modification of nanoPOTS termed microPOTS was reported that is a more adoptable variant not requiring a robotic platform [33]. It has been reported that nanoPOTS could identify >3000 proteins from 10 cultured mammalian cells, while microPOTS has been reported to reproducibly identify up to 1200 and 1800 proteins from 25 HeLa cells and 50 mm square mouse liver tissue, respectively [33]. Several nanoPOTS modifications have been reported since it was

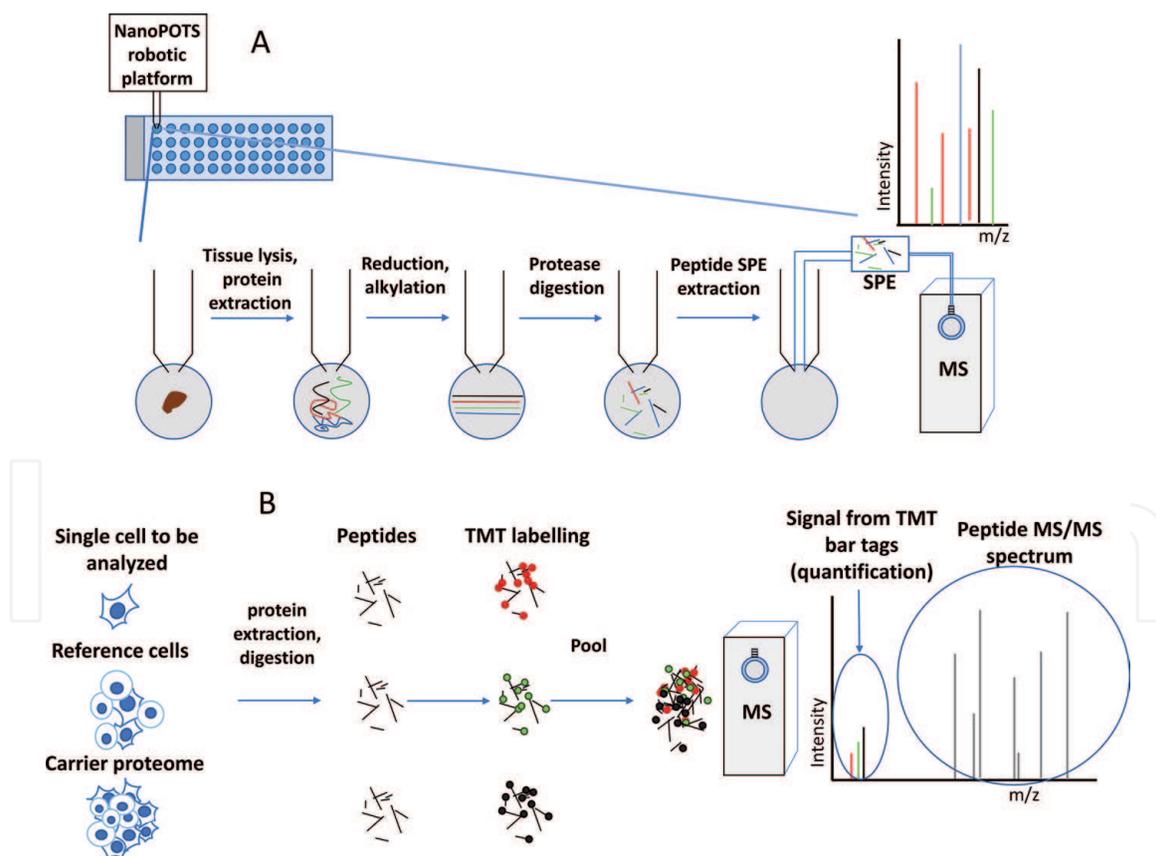


Figure 3.

Modern limited proteomic sample preparation approaches. (A) NanoPOTS; A limited proteomic sample preparation protocol that uses an automated robotic platform operating with nanoliter volumes. Sample is processed in a nano-well patterned slide. Sample preparation is based on principles of classical in-solution protein digestion. Protein digest is then transferred into SPE cartridge, where peptides are desalted and concentrated. Following, peptides are separated and analysed using mass spectrometry. (B) SCoPE-MS; a single cell proteome analysis platform. Carrier proteome is used to overcome sample losses accompanied due to peptide adsorption to surfaces. TMT labelling identifies the carrier and analysed proteomes. It could also serve for relative quantification of compared proteomes (SCoPE-MS2). Protein presence in the investigated sample and its quantity is determined based on reporter ion intensity.

introduced. For example, Zhu et al. claim that a combination of nanoPOTS with fluorescence activated cell sorting (FACS) could detect 670 protein groups from a single mammalian cell [34]. Later a combination of nanoPOTS, nanoLC separation operated at 20 nL/min and Orbitrap Eclipse and Tribrid mass spectrometer led even to a slight increase in sensitivity identifying ~1000 protein groups from a single HeLa cell [35]. Extraordinary low sample requirements predispose nanoPOTS to being useful for LC-MS/MS tissue imaging. Spatially resolved proteomic maps of a mouse blastocyst embedding into placenta have been produced using a combination of nanoPOTS and LCM. The nanoPOTS - LCM combination produced quantitative tissue images for >2000 proteins with 100- μ m spatial resolution which substantially outperformed classical protein imaging mass spectrometry (IMS) [36]. The universality of nanoPOTS is well documented in several publications summarising results from pancreas, liver brain tissue thin sections as well as plant samples.

Achieving submicrogram detection limits has also been reached by introducing a carrier proteome to decrease adsorption of the proteome of interest in combination with TMT labelling (**Figure 3B**). The carrier proteome spike-in helped the method known as Single-Cell-ProtEomics-by-Mass-Spectrometry (SCoPE MS) to overcome extensive losses due to adsorption of proteins to surfaces (e.g. LC columns) while the addition of TMT labelling identifies the carrier and analysed proteomes. Moreover, TMT labels enable relative protein quantitation of multiple samples/conditions per one LC-MS run. The SCoPE MS approach has enabled detection of >1000 proteins from a single mouse embryonic stem cell [37]. Specht et al. further exploited quantitative potential of TMT labels and claimed to reproducibly quantitate >1000 proteins in a SCoPE MS experiment investigating differentiating monocytes heterogeneity [38].

Introducing on-column immobilised protease digestion (IMER) downscales sample requirements up to the sub-microgram level, especially when combined with miniaturised column diameter. Utilising various nanostructured materials such as nanoporous material, nanoparticles, nanofibers and nanotubes succeeded in IMER nanobiocatalysis as it has led to enzyme stabilisation and increasing apparent enzyme activity per unit mass of immobilisation host [39]. Several sub-microgram proteomic setups combining IMER with downstream microfluidic platforms have been reported [40–42].

The microfluidic platform termed Open tubular lab-on-column combines LysC and trypsin enzymatic digestion on 20 mm inner diameter (ID) column with on-line connected nano LC-MS/MS system. Open tubular lab-on-column benefits from very narrow capillary ID and IMER column ID that prevent excessive peptide dilution and adsorption to fluidics. The authors detected a biomarker Axin 1 in 10 ng of HCT15 colon cancer cells [40]. Huang et al. characterised 348 proteins from 25 mice blastocysts on a platform termed SNaPP coupling enzymatic digestion on 150 mm ID IMER to nanofluidics [41]. Naldi et al. coupled SCX column-based IMER proteomic reactor to nano-proteomic platform capable of protein capture, reduction, alkylation, digestion and the first dimensional SCX peptide pre-separation followed by LC-MS/MS. These authors claim that the platform performs with as low as 200 ng protein starting material [42]. Moreover, the integrated Proteome Analysis Device (iPAD) couples a 10 port valve, digestion loop and SPE trap column in a microfluidic setup that is intended for micro sample preparation prior to mass spectrometry. The authors claim that the iPAD approach is capable of identifying 813 proteins in approx. 100 Duke's type C colorectal adenocarcinoma [43].

Capillary electrophoresis (CE) is an efficient and sensitive separation technique reliably resolving proteins or peptides. Historically, it has been less robust than nanoLC but recently this has begun to change. Specifically, the introduction of CE-ESI interfaces that do not lead to an excessive peptide dilution have made

CE-MS applicable in microproteomics [44]. Several reports describe various proteomic pipelines coupling CE to MS. An ultrasensitive electrokinetically pumped nanospray ionization source coupled with CE was able to identify 283 proteins from 80 ng of MCF7 breast cancer cells. Moreover, the detection limit of spiked-in angiotensin II in bovine serum albumin digest was 2 attomole/injection [45]. Although animal proteomics does not fall within scope of this chapter it is worth mentioning that CE-MS input allowed analysis down to 50 ng of *Xenopus laevis* eggs in a single protein extract. The authors of this study used linear polyacrylamide coating and sulfonate-silica hybrid strong cation exchange monolith for SPE followed by CE-MS [46]. Combining SPE with CE in 2D manner is a promising candidate for the future development of microscale CE-MS proteomics.

5. Conclusions and future perspectives

Developments in proteomics to identify clinically relevant proteins has been widely used in scientific research. Sample preparation has been considered as one of the key steps during analysis, and as such a variety of protocols to minimize variability and to obtain best sensitivity and protein recovery from the material have been used.

Constant development of technologies that could be applied in a medical context and potentially used for screening of patient samples have been rising in recent years. Technological evolution has also had an impact to provide platforms for proteome screening of limited cell numbers, i.e. some technologies have clearly demonstrated success on the single cell level. Cellular heterogeneity at the cellular level results during tumour development that can confound analysis. Therefore, advancement of the tools for profiling of cellular subpopulations or regions of tumours has great potential to provide novel insight in mechanisms of tumour growth. Moreover, integration of developed tools with machine learning algorithms to discover and map molecules that manifest pathological development will likely lead to a better understanding of mechanisms of oncogenesis and potentially uncover therapeutic targets.

Acknowledgements

This work was supported by the International Centre for Cancer Vaccine Science, carried out within the International Research Agendas program of the Foundation for Polish Science, co-financed by the European Union under the European Regional Development Fund. The University of Victoria-Genome BC Proteomics Centre is grateful to Genome Canada and Genome British Columbia for financial support for Genomics Technology Platforms (GTP) funding for operations and technology development (264PRO).

IntechOpen

Author details

Jakub Faktor¹, David R. Goodlett^{1,2,3} and Irena Dapic^{1*}

1 International Centre for Cancer Vaccine Science, University of Gdańsk, Gdańsk, Poland

2 Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada

3 University of Victoria Genome BC Proteome Centre, Victoria, BC, Canada

*Address all correspondence to: irena.dapic@ug.edu.pl

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] M. M. Morente et al., 'TuBaFrost 2: Standardising tissue collection and quality control procedures for a European virtual frozen tissue bank network', *Eur. J. Cancer Oxf. Engl.* 1990, vol. 42, no. 16, pp. 2684-2691, Nov. 2006, doi: 10.1016/j.ejca.2006.04.029.
- [2] S. R. Mager et al., 'Standard operating procedure for the collection of fresh frozen tissue samples', *Eur. J. Cancer Oxf. Engl.* 1990, vol. 43, no. 5, pp. 828-834, Mar. 2007, doi: 10.1016/j.ejca.2007.01.002.
- [3] H. Meng et al., 'Tissue triage and freezing for models of skeletal muscle disease', *J. Vis. Exp. JoVE*, no. 89, Jul. 2014, doi: 10.3791/51586.
- [4] B. Enthaler, T. Bussmann, J. K. Pruns, C. Rapp, M. Fischer, and J.-P. Vietzke, 'Influence of various on-tissue washing procedures on the entire protein quantity and the quality of matrix-assisted laser desorption/ionization spectra', *Rapid Commun. Mass Spectrom. RCM*, vol. 27, no. 8, pp. 878-884, Apr. 2013, doi: 10.1002/rcm.6513.
- [5] L. Giusti and A. Lucacchini, 'Proteomic studies of formalin-fixed paraffin-embedded tissues', *Expert Rev. Proteomics*, vol. 10, no. 2, pp. 165-177, Apr. 2013, doi: 10.1586/epr.13.3.
- [6] E. I. Chen, D. Cociorva, J. L. Norris, and J. R. Yates, 'Optimization of Mass Spectrometry-Compatible Surfactants for Shotgun Proteomics', *J. Proteome Res.*, vol. 6, no. 7, pp. 2529-2538, Jul. 2007, doi: 10.1021/pr060682a.
- [7] I. Rehman and S. Botelho, 'Biochemistry, Tertiary Structure, Protein', in *StatPearls, Treasure Island (FL): StatPearls Publishing*, 2020.
- [8] S. Suttapitugsakul, H. Xiao, J. Smeekens, and R. Wu, 'Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics', *Mol. Biosyst.*, vol. 13, no. 12, pp. 2574-2582, Nov. 2017, doi: 10.1039/c7mb00393e.
- [9] Y. Zhu et al., 'High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification', *Mol. Oncol.*, vol. 13, no. 11, pp. 2305-2328, 2019, doi: 10.1002/1878-0261.12570.
- [10] J. R. Wiśniewski, C. Wegler, and P. Artursson, 'Multiple-Enzyme-Digestion Strategy Improves Accuracy and Sensitivity of Label- and Standard-Free Absolute Quantification to a Level That Is Achievable by Analysis with Stable Isotope-Labeled Standard Spiking', *J. Proteome Res.*, vol. 18, no. 1, pp. 217-224, 04 2019, doi: 10.1021/acs.jproteome.8b00549.
- [11] N. Taouatas, S. Mohammed, and A. J. R. Heck, 'Exploring new proteome space: combining Lys-N proteolytic digestion and strong cation exchange (SCX) separation in peptide-centric MS-driven proteomics', *Methods Mol. Biol. Clifton NJ*, vol. 753, pp. 157-167, 2011, doi: 10.1007/978-1-61779-148-2_11.
- [12] R. G. Biringer, H. Amato, M. G. Harrington, A. N. Fonteh, J. N. Riggins, and A. F. R. Hühmer, 'Enhanced sequence coverage of proteins in human cerebrospinal fluid using multiple enzymatic digestion and linear ion trap LC-MS/MS', *Brief. Funct. Genomic. Proteomic.*, vol. 5, no. 2, pp. 144-153, Jun. 2006, doi: 10.1093/bfgp/ell026.
- [13] R. Chen et al., 'Glycoproteomics analysis of human liver tissue by combination of multiple enzyme digestion and hydrazide chemistry', *J. Proteome Res.*, vol. 8, no. 2, pp. 651-661, Feb. 2009, doi: 10.1021/pr8008012.
- [14] J. R. Wiśniewski and M. Mann, 'Consecutive proteolytic digestion in

an enzyme reactor increases depth of proteomic and phosphoproteomic analysis', *Anal. Chem.*, vol. 84, no. 6, pp. 2631-2637, Mar. 2012, doi: 10.1021/ac300006b.

[15] E. T. Goddard et al., 'Quantitative extracellular matrix proteomics to study mammary and liver tissue microenvironments', *Int. J. Biochem. Cell Biol.*, vol. 81, no. Pt A, pp. 223-232, 2016, doi: 10.1016/j.biocel.2016.10.014.

[16] S. J. Walmsley, P. A. Rudnick, Y. Liang, Q. Dong, S. E. Stein, and A. I. Nesvizhskii, 'Comprehensive Analysis of Protein Digestion Using Six Trypsins Reveals the Origin of Trypsin As a Significant Source of Variability in Proteomics', *J. Proteome Res.*, vol. 12, no. 12, pp. 5666-5680, Dec. 2013, doi: 10.1021/pr400611h.

[17] T. Sipost and J. R. Merkelt, 'An Effect of Calcium Ions on the Activity, Heat Stability, and Structure of Trypsin', p. 10.

[18] M. Kotormán, I. Laczkó, A. Szabó, and L. M. Simon, 'Effects of Ca²⁺ on catalytic activity and conformation of trypsin and α -chymotrypsin in aqueous ethanol', *Biochem. Biophys. Res. Commun.*, vol. 304, no. 1, pp. 18-21, Apr. 2003, doi: 10.1016/S0006-291X(03)00534-5.

[19] D. López-Ferrer et al., 'Pressurized Pepsin Digestion in Proteomics: An Automatable Alternative to Trypsin for Integrated Top-Down Bottom-Up Proteomics', *Mol. Cell. Proteomics*, vol. 10, no. 2, p. M110.001479, Feb. 2011, doi: 10.1074/mcp.M110.001479.

[20] S. B. Sorensen, L. Sorensen, and K. Breddam, 'Fragmentation of proteins by *S. aureus* strain V8 protease', p. 3.

[21] M. A. Porzio and A. M. Pearson, 'Isolation of an extracellular neutral proteinase from *Pseudomonas fragi*',

Biochim. Biophys. Acta BBA - Enzymol., vol. 384, no. 1, pp. 235-241, Mar. 1975, doi: 10.1016/0005-2744(75)90112-6.

[22] D. L. Swaney, C. D. Wenger, and J. J. Coon, 'Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics', *J. Proteome Res.*, vol. 9, no. 3, pp. 1323-1329, Mar. 2010, doi: 10.1021/pr900863u.

[23] J. G. Meyer, S. Kim, D. A. Maltby, M. Ghassemian, N. Bandeira, and E. A. Komives, 'Expanding Proteome Coverage with Orthogonal-specificity α -Lytic Proteases', *Mol Cell Proteomics*, vol. 13, pp. 823-835, Mar 2014, doi: 10.1074/mcp.M113.034710.

[24] J. P. Owen, B. C. Maddison, G. C. Whitlam, and K. C. Gough, 'Use of thermolysin in the diagnosis of prion diseases', *Mol. Biotechnol.*, vol. 35, no. 2, pp. 161-170, Feb. 2007, doi: 10.1007/BF02686111.

[25] T. P. Graycar, R. R. Bott, S. D. Power, and D. A. Estell, 'Subtilisins', in *Handbook of Proteolytic Enzymes*, Elsevier, 2013, pp. 3148-3155.

[26] J. Reinders, U. Lewandrowski, J. Moebius, Y. Wagner, and A. Sickmann, 'Challenges in mass spectrometry-based proteomics', *PROTEOMICS*, vol. 4, no. 12, pp. 3686-3703, Dec. 2004, doi: 10.1002/pmic.200400869.

[27] F. M. Dehrmann, T. H. T. Coetzer, R. N. Pike, and C. Dennison, 'Mature Cathepsin L Is Substantially Active in the Ionic Milieu of the Extracellular Medium', *Arch. Biochem. Biophys.*, vol. 324, no. 1, pp. 93-98, Dec. 1995, doi: 10.1006/abbi.1995.9924.

[28] D. Turk, 'Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular

serine proteases', *EMBO J.*, vol. 20, no. 23, pp. 6570-6582, Dec. 2001, doi: 10.1093/emboj/20.23.6570.

[29] D. S. Waugh, 'An overview of enzymatic reagents for the removal of affinity tags', *Protein Expr. Purif.*, vol. 80, no. 2, pp. 283-293, Dec. 2011, doi: 10.1016/j.pep.2011.08.005.

[30] L. Switzar, M. Giera, and W. M. A. Niessen, 'Protein Digestion: An Overview of the Available Techniques and Recent Developments', *J. Proteome Res.*, vol. 12, no. 3, pp. 1067-1077, Mar. 2013, doi: 10.1021/pr301201x.

[31] E. Gross, 'Nonenzymatic Cleavage of Peptide Bonds: The Methionine Residues in Bovine Pancreatic Ribonuclease', vol. 237, no. 6, p. 6, 1962.

[32] M. Safdar, J. Sproß, and J. Jänis, 'Microscale immobilized enzyme reactors in proteomics: Latest developments', *J. Chromatogr. A*, vol. 1324, pp. 1-10, Jan. 2014, doi: 10.1016/j.chroma.2013.11.045.

[33] K. Xu et al., 'Benchtop-compatible sample processing workflow for proteome profiling of < 100 mammalian cells', *Anal. Bioanal. Chem.*, vol. 411, no. 19, pp. 4587-4596, Jul. 2019, doi: 10.1007/s00216-018-1493-9.

[34] Y. Zhu et al., 'Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS', *Angew. Chem. Int. Ed.*, vol. 57, no. 38, pp. 12370-12374, Sep. 2018, doi: 10.1002/anie.201802843.

[35] Y. Cong et al., 'Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell', *Systems Biology*, preprint, Jun. 2020. doi: 10.1101/2020.06.03.132449.

[36] P. D. Piehowski et al., 'Automated mass spectrometry imaging of over 2000 proteins from tissue sections

at 100- μ m spatial resolution', *Nat. Commun.*, vol. 11, Jan. 2020, doi: 10.1038/s41467-019-13858-z.

[37] B. Budnik, E. Levy, G. Harmange, and N. Slavov, 'SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation', *Genome Biol.*, vol. 19, no. 1, p. 161, Dec. 2018, doi: 10.1186/s13059-018-1547-5.

[38] H. Specht et al., 'Single-cell mass-spectrometry quantifies the emergence of macrophage heterogeneity', *bioRxiv*, p. 665307, Dec. 2019, doi: 10.1101/665307.

[39] J. Kim, B. C. Kim, D. Lopez-Ferrer, K. Petritis, and R. D. Smith, 'Nanobiocatalysis for protein digestion in proteomic analysis', *Proteomics*, vol. 10, no. 4, pp. 687-699, Feb. 2010, doi: 10.1002/pmic.200900519.

[40] H. K. Hustoft et al., 'Open Tubular Lab-On-Column/Mass Spectrometry for Targeted Proteomics of Nanogram Sample Amounts', *PLoS ONE*, vol. 9, no. 9, p. e106881, Sep. 2014, doi: 10.1371/journal.pone.0106881.

[41] E. L. Huang et al., 'SNaPP: Simplified Nanoproteomics Platform for Reproducible Global Proteomic Analysis of Nanogram Protein Quantities', *Endocrinology*, vol. 157, no. 3, pp. 1307-1314, Mar. 2016, doi: 10.1210/en.2015-1821.

[42] M. Naldi, U. Černigoj, A. Štrancar, and M. Bartolini, 'Towards automation in protein digestion: Development of a monolithic trypsin immobilized reactor for highly efficient on-line digestion and analysis', *Talanta*, vol. 167, pp. 143-157, May 2017, doi: 10.1016/j.talanta.2017.02.016.

[43] Q. Chen, G. Yan, M. Gao, and X. Zhang, 'Ultrasensitive Proteome Profiling for 100 Living Cells by Direct Cell Injection, Online Digestion and

Nano-LC-MS/MS Analysis', *Anal. Chem.*, vol. 87, no. 13, pp. 6674-6680, Jul. 2015, doi: 10.1021/acs.analchem.5b00808.

[44] R. Gahoual, E. Leize-Wagner, P. Houzé, and Y.-N. François, 'Revealing the potential of capillary electrophoresis/mass spectrometry: the tipping point', *Rapid Commun. Mass Spectrom.*, vol. 33, no. S1, pp. 11-19, Sep. 2018, doi: 10.1002/rcm.8238.

[45] L. Sun, G. Zhu, S. Mou, Y. Zhao, M. M. Champion, and N. J. Dovichi, 'Capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry for quantitative parallel reaction monitoring of peptide abundance and single-shot proteomic analysis of a human cell line', *J. Chromatogr. A*, vol. 1359, pp. 303-308, Sep. 2014, doi: 10.1016/j.chroma.2014.07.024.

[46] Z. Zhang, G. Zhu, E. H. Peuchen, and N. J. Dovichi, 'Preparation of linear polyacrylamide coating and strong cationic exchange hybrid monolith in a single capillary, and its application as an automated platform for bottom-up proteomics by capillary electrophoresis-mass spectrometry', *Microchim. Acta*, vol. 184, no. 3, pp. 921-925, Mar. 2017, doi: 10.1007/s00604-017-2084-8.