Evaluation of sample preparation methods for protein analysis of liver organoids

Lise Midtøy

Master Thesis
Master's degree in Chemistry
60 credits

Bioanalytical Chemistry
Department of Chemistry
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

[June / 2023]

Evaluation of sample preparation methods for protein analysis on liver organoids

2023

Evaluation of sample preparation methods for protein analysis on liver organoids

Lise Midtøy

http://www.duo.uio.no/

Printing: Grafisk senter, Universitetet i Oslo

Abstract

For mass spectrometry (MS)-based proteomics, different sample preparation methods were evaluated for limited cell samples (10 000-200 000 cells), and the optimized method was used to analyze liver organoids (20-30 organoids, equivalent to approximately 50 000 cells). The sample preparation methods evaluated were urea-based in-solution-digestion (ISD-Urea), a detergent-free sample preparation method named sample preparation by easy extraction and digestion (SPEED), and a method with bead-based sample clean-up called single-pot, solid-phase enhanced sample preparation (SP3). The samples were analyzed with liquid chromatography (LC)-MS using a nanoLC column or a micropillar array column coupled to a trapped ion mobility time of flight (timsTOF)-MS.

The samples prepared by SPEED had the overall highest number of protein identifications among all the sample sizes, with a range of 1684-3503 proteins identified for 10 000-100 000 cells. A quantification method for estimation of protein yield was found difficult, but because of the superior performance regarding number of protein identifications, SPEED was chosen as method for analyzing liver organoids of different metabolic states using label-free-quantification. The results showed significant differences between the organoid samples, with many of the hypothesized protein profiles found. The number of protein identifications was in the range of 2140-3203 proteins for the samples consisting of 20-30 organoids each. The method was successfully implemented as a preliminary step for disease studies, e.g. non-alcoholic fatty liver disease (NAFLD).

Abbreviations

Abbreviation	Full name
µPAC	Micropillar array column
2D	Two-dimensional
3D	Three-dimensional
A280	Absorbance at 280 nm
ABC	Ammonium bicarbonate
ACN	Acetonitrile
BCA	Bicinchoninic acid
BPC	Base peak chromatogram
BSA	Bovine serum albumin
CID	Collision induced dissociation
C-terminus	Carboxyl-terminus
DDA	Data-dependent acqusition
DIA	Data-independent acqusition
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
FA	Formic acid

FDR	False-discovery rate
HR	High-resolution
HTH	Hybrid Technology Hub
IAM	Iodoacetamide
ID	Inner diameter
IGD	In-gel-digestion
iPSC	Induced pluripotent stem cells
ISD	In-solution-digestion
ISD-Urea	Urea based in-solution-digestion
iTRAQ	Isobaric tags for relative and absolute quantification
LC	Liquid chromatography
LFQ	Label-free quantification
LOD	Limit of detection
m/z	Mass-to-charge ratio
MP	Mobile phase
MS	Mass spectrometry
MS^2	Tandem mass spectrometry
NAFL	Non-alcoholic fatty liver
NAFLD	Non-alcoholic fatty liver disease
NAPI	National network for advanced proteomics

NASH	Non-steatohepatitis
OoC	Organ-on-a-chip
PASEF	Parallel accumulation-serial fragmentation
PBS	Phosphate-buffered saline
PCA	Principle component analysis
PSC	Pluripotent stem cells
PSM	Peptide spectrum match
PTM	Post-translational modification
RP	Reversed phase
RSD	Relative standard deviation
RT	Room temperature
SCP	Single-cell proteomics
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SP	Solid phase
SP3	Single-pot, solid-phase enhanced sample preparation
SPE	Solid phase extraction
SPEED	Sample preparation by easy extraction and digestion
TFA	Trifluoroacetic acid
TIC	Total ion current

TIMS	Trapped ion mobility spectrometry
timsTOF	trapped ion mobility spectrometry time-of-flight
TMT	Tandem mass tags
TOF	Time-of-flight
TX100	TritonX-100
UHPLC	Ultra-high performance liquid chromatography
UHR	Ultra-high resolution
UV	Ultra violet
UV-VIS	Ultra violet - visible
WR	Working reagent
WRB	Working reconstitution buffer

Selected definitions

List of definitions that may vary depending on the scientific field.

Blank sample: Blank sample consists of the same diluent (e.g. water) as the analytes are dissolved in, and is analyzed together with the samples. The blank sample does not contain the analytes in question.

Control sample: The control sample is prepared and analyzed in the same way as the samples. With the same matrix composition, the control sample does not contain the analytes in question.

Biological replicates: Biological samples come from different sources, e.g. cells grown in different wells, and reflect biological variations.

Sample replicate (n): Not to be confused with biological replicates. Sample replicates are prepared in the same way with the same concentrations/amount and matrix composition. Sample replicates define the precision and accuracy of the whole experiment.

Technical replicates (N): Technical measurements are repeated measurements from the same sample (replicate). Technical measurements will only define the precision and accuracy of the instrument.

Acknowledgments

The majority of the figures are partly or entirely created in BioRender. All the chemical structures are created in ChemDraw.

This thesis presents unpublished work (discussion and results are found in **Section 4.6**) that is a part of the LC/GC special edition for the 2023 HPLC conference.

Preface

Even though my name is on the front cover of this thesis, this would not have been possible without the guidance and motivation from my supervisors. You made the best support team a master student could ask for. I would like to thank Ph.D. candidate Ingrid Wilhelmsen for providing me with HepG2 cells and teaching me the importance of setting my boundaries, and Ph.D. candidate Kristina Sæterdal Kømurcu for always dreaming bigger than big. I am forever grateful to Professor Steven Wilson for always cheering me on and for the opportunity to present some of my findings at the 2023 HPLC conference in Düsseldorf. Thanks to Ph.D. candidate Stian Kogler for all the laughs and support. My biggest gratitude goes to Associate Professor Hanne Røberg-Larsen. Her open-door policy has been invaluable for me, providing a space for discussions of both academic and personal matters.

I would like to thank all the members of the Bioanalytical group that I have been a part of since my bachelor thesis, especially the ones I spend the last two years with Cristina, Eva, Marius, Gustav, Jon Erlend, Sander, Ilka, Jan Arve, and Ph.D. candidate Christine. Thank you for the academic discussions and moral support. Also, a special thanks to engineer Inge Mikalsen for all the 3D printed equipment, and for performing instrumental first aid.

Not only have I had my supervisors discuss the delicate nature of proteins; I have been very fortunate to get to work with several people. Thanks to Ph.D. Henriette Engen Berg for paving the proteomics road at our group and answering all my questions. Thank you, Ph.D. Tuula Anneli Nyman, Ph.D Bernd Thiede and Ph.D. Manuel Ramirez Garrastacho at the core facility under the National network of Advanced Proteomics Infrastructure for helping me with data analysis, setting up the µPAC column, and analyzing the samples. I would also like to thank the members of the Hybrid Technology Hub at the Department of Medicine, especially engineer Justyna Stokowiec for giving me an insight into all things molecular biology, and Postdoctoral research fellow Aleksandra Aizenshtadt for providing me with organoids.

Lastly, I would like to thank Linda from Forvei and my partner Kristoffer for their support. I would like to pass on their wise words to future students: Don’t compare yourself with others. Don’t let mistakes stop you from trying again. Don’t let grades and results (or lack there off) define you. There are no “good” or “poor” results. We are in this together, trying to make the world more understandable one chromatogram (and when in doubt, an MS^2 spectrum) at a time.

Table of contents

1 Introduction

1.1 Non-alcoholic fatty liver disease

1.1.1 A disease often associated with lifestyle factors

Non-alcoholic fatty liver disease (NAFLD) is a chronic liver disease caused by the accumulation of hepatic fat in the liver. It is estimated that about 30% of the world's population suffers from this disease [1]. NAFLD differs from alcoholic fatty liver disease, where excessive alcohol consumption causes the accumulation of hepatic fat [2]. Among the key factors behind the development of NAFLD are increased intake of glucose, fructose, and fatty acids. This is often associated with obesity, type 2 diabetes, or poor metabolic health [3]. As illustrated in **Figure 1**, NAFLD ranges from non-alcoholic fatty liver (NAFL) to the severe form of non-alcoholic steatohepatitis (NASH). If not reversed, NASH can lead to cirrhosis and hepatocellular carcinoma, and/or organ malfunction [3, 4]. The prevalence and development of NASH among patients with NAFLD remain unclear, meaning that the patients at risk are difficult to predict [5]. Today, a liver biopsy is the most specific test to determine the state of the liver. With the additional cost and risk of complications regarding the biopsy procedure, there is a quest for non-invasive, specific diagnostic tools to identify patients with NAFLD, and more importantly, to evaluate the disease severity [3]. To understand the underlying mechanism behind this disease, it is necessary with disease models that can mimic the complexity of the human biological system.

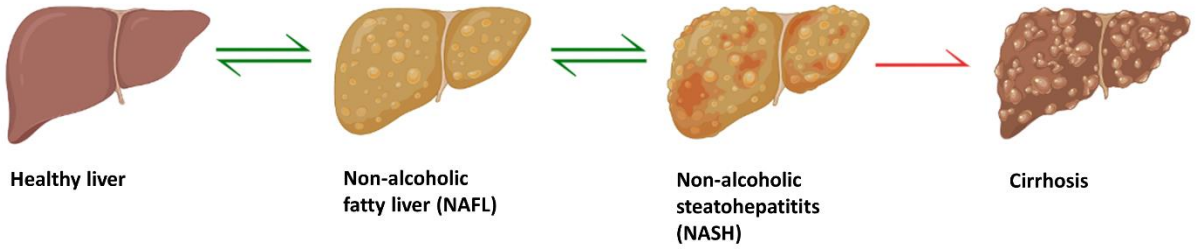

Figure 1. Development of different stages and severity of NAFLD. While NAFL is characterized by fatty liver, inflammation causes progression to NASH. The stages are reversible, but irreversible once progressed to cirrhosis. The figure is adapted from [6, 7].

1.1.2 Organoids as a disease model

Advancements in stem cell technology enable the generation of cellular models. Through step-wise biochemical and physical cues, stem cells can differentiate into any disease-relevant or tissue-specific cell type [8]. Stem cells can come from sources like adult stem cells from primary tissue or pluripotent stem cells (PSCs). PSCs are further categorized as embryotic stem cells or induced pluripotent stem cells (iPSCs). iPSCs are cells (e.g. skin cells) that are reprogrammed back to stem cells, which are ready to be differentiated into the model cells [9].

Two-dimensional (2D) cell cultures and animal models are commonly used as models for studying biological processes. However, both technologies suffer from oversimplification, as human development and diseases are inaccurately modeled in non-physiologic 2D cell cultures or non-human conditions, as in the case of animal models [10]. Because of these drawbacks, organoids, "mini-organs," are emerging tools in the field of medical research [11].

Organoids consist of cells able to self-organize through cell-cell and cell-matrix interactions and generate multicellular three-dimensional (3D) structures (**Figure 2**) [12]. Unlike 2D cell cultures, the 3D environment of the cells in organoids is similar to the cells of the native organ, thus enabling a near-physiologic representation [8, 11]. Organoids can be grown inside microfluidic chips, known as "organ-on-a-chip (OoC). The chip is designed to control nutrients, growth factors, and concentration gradient of chemicals, resembling the extracellular microenvironment more accurately, and making the model physiologically relevant [13, 14]. An issue with the OoC system compared to its conventional organoid culture counterpart, is that it generates low sample amounts, thus complicating experiments that usually require larger samples [9] as in the case of proteomics.

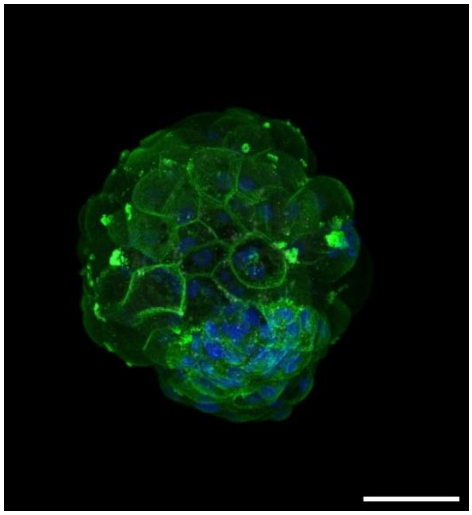

Figure 2. Organoid immunofluorecent confocal imaging showing expression of protein E-cadherin (green) in liver organoid. (Scale bar 50 µm.)

The use of organoids as tools to study disease modeling, drug discovery, personalized medicine, toxicology, physiology, and organ development is emerging [15], and was of 2017 awarded the "Method of the year" by *Nature Methods* [16]. However, there are drawbacks associated with organoids, and like 2D models and animal models, it is still a simplification of the human biological system. Among the limitations are the lack the maturity level, the range of cell types,

and heterogeneity as the organ the organoid model is mimicking. The 3D structure of organoids causes the cells deeply embedded in the organoids to have restricted access to nutrients compared to the cells on the surface of the organoids, which in the worst case will cause cell death [17]. Despite these drawbacks, the use of organoids for mimicking diseases such as NAFLD has great potential, as it might allow the study of diseases in a controlled human-like environment.

1.2 Proteomics

1.2.1 The biological importance of proteins

Proteins drive nearly every process that takes place in a cell and are the most abundant macromolecules, constituting approximately 50% of the dry weight of a cell [18, 19]. The 3D structure of the protein dictates its function. Proteins are the most versatile macromolecules, and there are five functional classes of proteins: metabolic enzymes, structural proteins, transport proteins, cell signaling proteins, and genomic caretaker proteins [20]. The most varied types of proteins are enzymes, which function as chemical catalysts for cellular reactions by lowering the activation energy and providing an optimal environment for product formation [18, 20].

1.2.2 Protein chemistry

Proteins are polymers made up of a set of 20 amino acids. Depending on the combination of the amino acids, countless proteins can be produced. Amino acids have a common structure where the α-carbon has an amino group, a carboxyl acid, a hydrogen atom, and a side chain (represented as R’ and R’’ in **Figure 3**). The different side chains have distinctive chemical properties and dictate how the proteins are folded, and thus the function of the protein.

In proteins, the amino group and carboxylic acid of neighboring amino acids are covalently linked to form a peptide bond. The peptide bond is formed by the loss of water through condensation reaction (**Figure 3**). The reaction is enzymatically controlled, making the peptide bonds stable under physiological conditions. When amino acids are covalently linked together, they are referred to as amino acid residue.

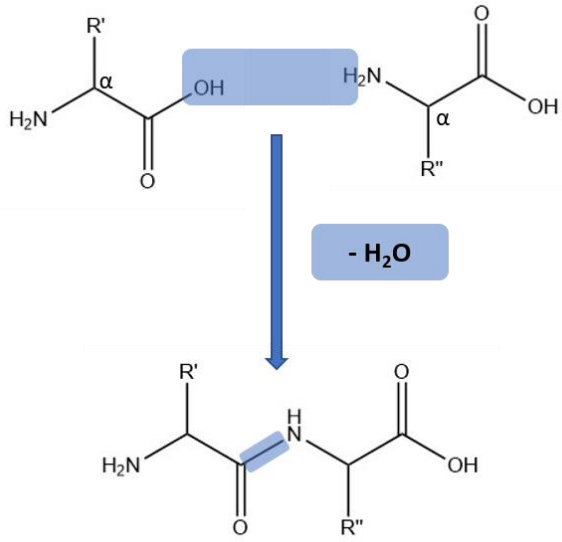

Figure 3. Two amino acids are covalently linked together by the loss of a water molecule. The peptide bond is colored blue. R' and R'' represent different side chains.

Proteins are folded, usually in a globular shape. The main contributors to the stabilization of the 3D structure are the noncovalent weak polar interactions: van der Waals interactions, hydrogen bonds, and electrostatic interactions. The amino acids with an uncharged and nonpolar side chain, are defined as the hydrophobic amino acids. Van der Waals interaction is the dominating force between hydrophobic amino acid residues, where the hydrophobic parts of the protein cluster together, minimizing surface area exposed to the aqueous environment. This clustering of hydrophobic residues is known as the hydrophobic effect. The hydrophilic amino acids have a polar and/or charged side chain. Depending on the pH of the environment, some of the hydrophilic amino acids can change their state from charged to uncharged and vice versa. The interaction dominating the hydrophilic amino acid residues are hydrogen bonds, which they form with each other, organic molecules, and water. Electrostatic interactions tend to be closer to the surface of the proteins. Some amino acids are amphipathic, meaning they have both polar and nonpolar characters [21, 22].

1.2.3 Proteomics is the study of the proteome

The study of the proteome, which constitute all the proteins expressed by a given cell or group of cells, is call proteomics [23]. Typically, proteomics involves a global analysis of the protein expression and properties of the sample. In addition to the global approach, proteomics can focus on the quantification of pre-defined proteins (targeted proteomics) or a particular set of post-translational modifications (PTMs) [15].

While the genome is static because the deoxyribonucleic acid (DNA) sequence is the same in every cell of the organism, the proteome is dynamic. Affected by external stimuli, the proteome changes at all times (**Figure 4**). The proteome even differs within one cell type, as cells are in different cellular states due to changes in metabolic and environmental conditions, leading to several possible protein expressions [24, 25].

Proteomics provides fundamental information about the biological system and processes. Proteins can be used as indicators or biomarkers, as altered proteins can be key drivers of a particular disease. Since proteins have a key role in virtually every biological process, proteomics can provide information about how a disease is developed and regulated. In addition to answering questions regarding diseases, proteomics is important when determining the drug response as proteins often are the main targets for drugs (e.g. a drug acting as an inhibitor or enhancer of protein function) [24, 26].

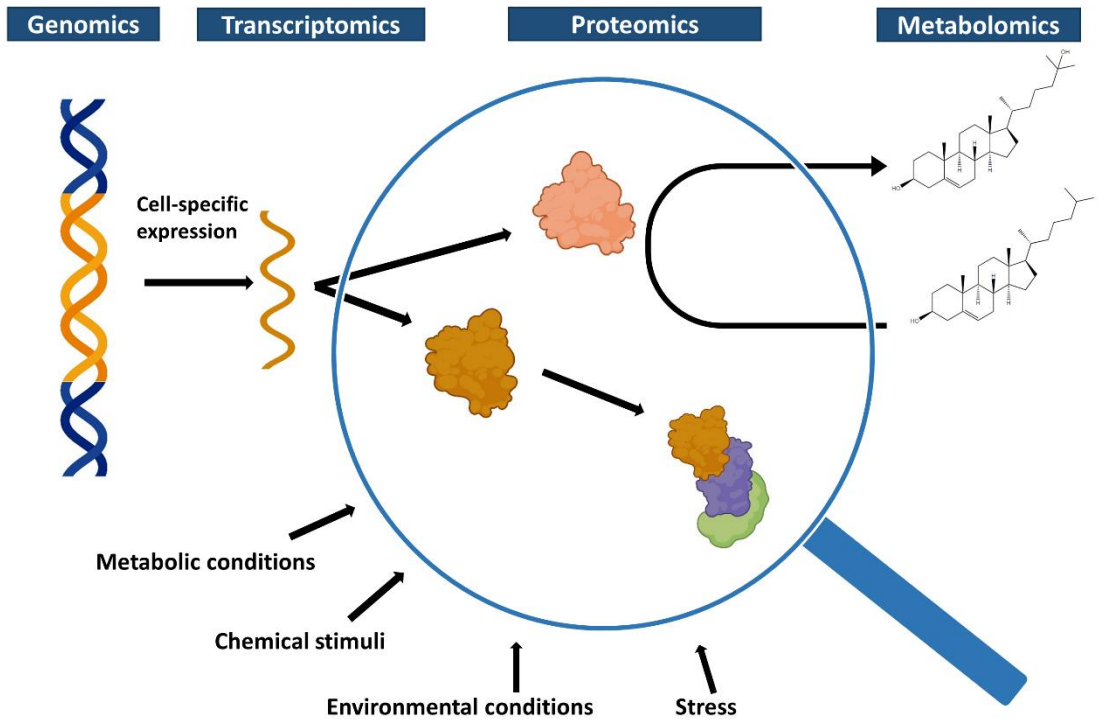

Figure 4. The field of "omics" and how they are intertwined. Genes are turned on due to cell-specific expression (genomics), and translated into proteins (transcriptomics). As proteins can contain PTMs, transcriptomics does not fully reveal the protein expression of the cell. Because the DNA sequence is the same in every cell of the organism, the genome remains predictable and static. Proteins map each step of a biochemical pathway and produce small polar compounds called metabolites. The proteome and metabolome, are dynamic and change continuously [24, 27].

1.2.4 Bottom-up and top-down proteomics

Liquid chromatography (LC) coupled to a mass spectrometer (MS) is considered the workhorse within proteomics due to high resolution and sensitivity, enabling a comprehensive proteome to be obtained [28-30]. The technique and principle of LC and MS will be described later in **Sections 1.7** and **1.8**, respectively.

There are two main approaches for MS-applied proteomics: Bottom-up and top-down proteomics (**Figure 5**). With bottom-up proteomics peptides (generated from proteins) are introduced to the LC-MS, while in top-down proteomics intact proteins enter the system. Bottom-up proteomics is the most used approach as peptides are easier to keep in solution, ionize more efficiently, and generate simpler MS-spectra compared to the top-down approach [31, 32]. The MS is efficient at obtaining data from peptides that are approximately 20 amino acid residues long, thus the peptides generated from enzymatic digestion are more compatible with the MS rather than proteins [33]. In addition, the sensitivity of the MS is lower for proteins than for peptides [33] due to the broad mass-to-charge (*m/z*) distribution of high molecular-weighted proteins, which reduces the signal [34, 35]. The advantages of top-down proteomics are specificity, enabling the identification of modified isomers, and a lower false-positive rate for protein identification. Yet the advantages come at the expense of complex instrumentation, advanced software to deconvolution the complex spectra, and higher experimental requirements [32].

The standard procedure for bottom-up proteomics involves cell lysis, protein extraction, and solubilization before the proteins are subjected to denaturalization, reduction, alkylation, and enzymatically digestion into peptides [36].

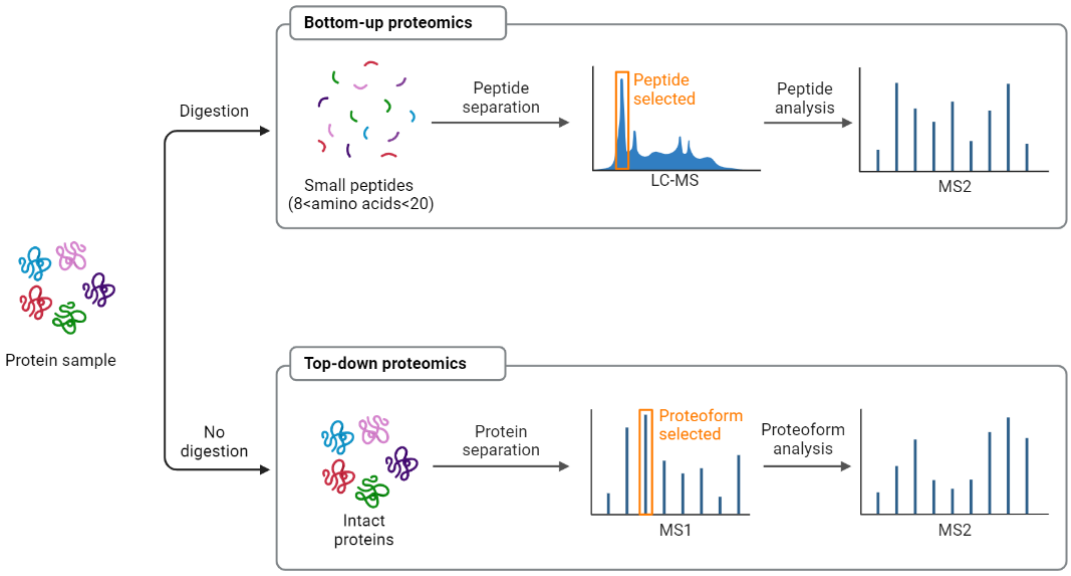

Figure 5. Bottom-up and top-down proteomics. In bottom-up proteomics, peptides are entering the LC-MS-system, while in top-down proteomics intact proteins are entering the system. While the top-down approach has a nearly full sequence coverage and all the PTMs intact, the protein ions can be very large and highly charged, setting high technical, computable demands. The bottom-up approach rarely has full sequence coverage, and PTMs can be lost. Peptides are often in the ideal mass range for MS. The figure is adapted from a BioRender template.

1.3 Extracting proteins from cells

Cell lysis is a method where the cell membrane is disrupted to release and extract cellular contents such as DNA, proteins, and other intracellular components [37]. The extract is referred to as the lysate. There are different ways to categorize the different lysis techniques. A simple way to distinguish the techniques is as physical disruption or solution (chemical) based cell lysis [38].

1.3.1 Detergents are used for chemical lysis, protein solubilization, and denaturation

Detergents have a hydrophobic and a hydrophilic (and possibly ionic) moiety, making detergents amphiphilic molecules. Because of this property, detergents can disintegrate the lipid bi-layer of the cell membrane by disrupting the hydrophobic and hydrophilic interactions. The detergents form micelles with the hydrophilic moiety face towards the aqueous (polar) solution, and the hydrophobic “tails” are facing inwards because of hydrophobic interactions. Since the membrane proteins have hydrophilic and hydrophobic regions, will they be incorporated into detergent-formed micelles with the disruption of the cell membrane [36]. The detergents are categorized by the type of charge the hydrophilic moiety carries (anionic, cationic, non-ionic,

and zwitterionic), where the non-ionic detergents cause minimal protein damage. Common non-ionic detergents are TritonX-100 (TX100) and NP-40. Anionic detergents, such as sodium dodecyl sulfate (SDS), are considered strong lysis agents. Because anionic detergents quickly denature proteins, are they not suitable for sensitive protein extraction where the native (3D) structure of the proteins needs to be intact [36]. In bottom-up proteomics, this is not of concern but rather considered a bonus as the proteins are to be denatured in order to efficiently be cleaved into peptides. However, detergents need to be removed in order to be compatible with downstream sample preparation and to avoid ion suppression in the MS [39].

Before enzymatic digestion, are the proteins unfolded to increase the cleaving efficiency and peptide yield. Denaturation is the process where the 3D structure of the protein is disrupted and unfolds, causing the protein to lose its function. Because the protein structures are stabilized mainly by noncovalent weak interactions, they quickly denature when the environment changes. Proteins denature when they are exposed to elevated temperature, pH changes, or when the environment is chemically changed [22]. Using detergents and chaotropic agents will in addition to cell lysis often denature proteins.

Chaotrophic agents, such as urea, guanidine, and ethylenediaminetetraacetic acid (EDTA) disrupt hydrogen bonds between water molecules, thus weakening the hydrophobic effect of proteins [37]. EDTA is also a chelating agent, which binds to metal ions. Metal ions are cofactors for different enzymes, and by including EDTA in the lysis buffer, enzymes such as DNases and proteases get inhibited [40]. A cocktail of protease inhibitors is often added to the lysis buffer because the proteins are subjected to spontaneous degradation by proteases when extracted from their biological environments [36].

1.3.2 Physical lysis is used to simplify downstream sample preparation

Physical cell lysis happens when the cell membrane is disrupted and broken down by external force. An example of physical disruption is freeze-thaw, where cells are repeatedly frozen and rapidly thawed at room temperature (RT). Ice crystals will form at the surface of the cells, and the deformation of the cell membrane aid in the disruption of the cell membrane [36, 37].

Sonication, the use of high-frequency waves (10-40 kHz), is another technique of physical disruption, where rapid change in pressure disrupts the cell membrane. A disadvantage is the formation of heat, therefore short pulses of sonication are recommended [36].

Mechanical lysis simplifies downstream sample preparation, as the chemical lysing agents often are incompatible with downstream workflow and analysis. Mechanical lysis is associated with high lysing efficiency [37, 38]. There are, however, limitations with mechanical lysis. In the absent of protease inhibitors, spontaneous protein digestion will occur. Membrane proteins are largely hydrophobic and difficult to handle in aqueous solutions. Without detergents, hydrophobic proteins will aggregate and precipitate [36].

1.4 Quantification of proteins in cell lysate

Protein quantification in the lysate can be of importance e.g. to determine the amount of enzyme needed (often trypsin) in bottom-up proteomics. Spectrophotometry uses light to measure the analyte concentration, and different spectrophotometric methods can be used for the quanfication of proteins.

The protein concentration is determined using the Beer-Lambert law (**Equation 1)**, where absorbance (A) is proportional to the protein concentration (c) given in mol/L. Molar absorptivity (or extinction coefficient) (ε) expresses how much light is absorbed at a specific wavelength. The unit is given as M^{-1} cm^{-1}. Path length (b) for the light is given in cm [41].

$$A = c \times b \times \varepsilon \qquad \textbf{Equation 1}$$

1.4.1 Measuring the absorbance at 280 nm

Aromatic amino acids (tryptophan, tyrosine, and to a lesser degree phenylalanine) and disulfide bridges absorb ultra violet (UV) light with a wavelength of 280 nm. Absorption measurement at 280 nm (A280) requires a solution of high purity and a higher protein concentration, due to a higher limit of detection (LOD), compared with other methods for protein determination [36]. A280 allows for rapid quantification without the need for additional selective reagents and the establishment of a calibration curve [42]. If the sample contains nucleic acids and nucleotides, the absorbance at 280 nm needs to be corrected since nucleotides contain an aromatic ring [36]. Common detergents used for cell lysis, e.g. TX-100, NP-40, and SDS, have high UV absorption

at 280 nm, making it difficult to detect proteins in these buffers [36]. Alternative approaches that selectively react with proteins and absorb light at different wavelengths exist, e.g. bicinchoninic acid (BCA) assay. Other methods for protein quantification is measuring tryptophan fluorescents (**Section 7.1.1 in Appendix**), and novel protein quantification methods such as turbidity measurement (**Section 7.1.2 in Appendix**).

1.4.2 Bicinchoninic acid assay

Colorimetric methods such as the BCA assay are often implemented for the selective detection of proteins in the lysate. In the presence of cysteine, tyrosine, tryptophan, and peptide bonds, Cu^{2+} is reduced to Cu^{+}. The reduced copper will form a colored complex with the BCA (**Figure 6**), enabling the detection of proteins with the maximum absorption of the complex at 562 nm. A point of consideration when using reducing reagents such as dithiothreitol (DTT), is that it will interfere with this assay because they also reduce Cu^{2+} [36].

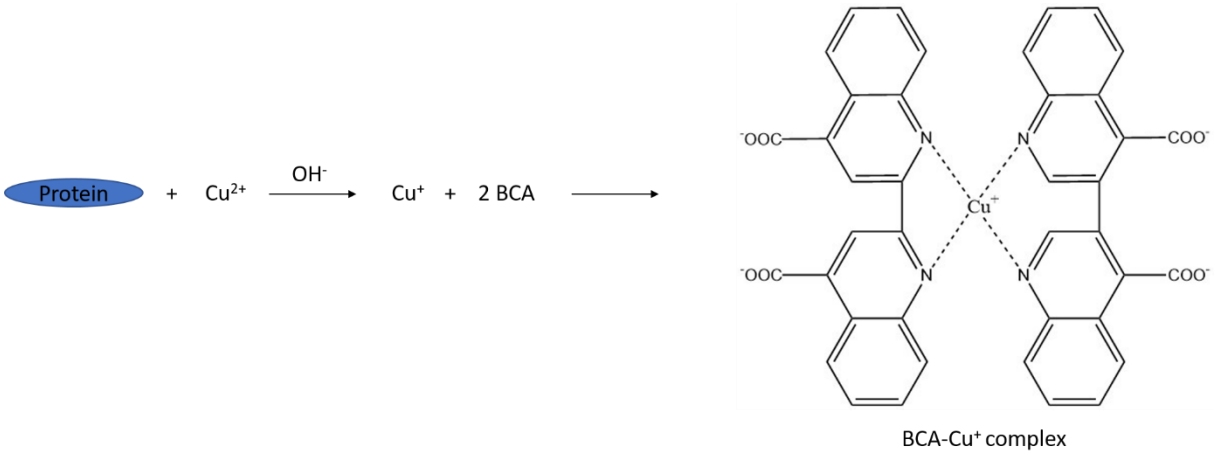

Figure 6. The formation of BCA-Cu+ complex. Cu^{2+} reduces to Cu^{+} in the presence of proteins. The reduced copper forms a colored complex with two BCA molecules. The figure is adapted from [36].

1.4.3 Microvolume UV-VIS spectrophotometer

The limitation with most UV-visible (UV-VIS) spectrophotometers is the relatively high volume necessary to quantify. When the sample is limited in sample size, this becomes an even bigger issue. Microvolume UV-VIS spectrophotometers are available on the market, such as the Nanodrop spectrophotometer from Thermo Fisher (**Figure 7**). From here on, microvolume UV-VIS spectrophotometer is referred to as Nanodrop. The sample is placed between a measurement pedestal and sampling arm which form a liquid column ("cuvette"), where light

from a xenon flash lamp passes through the liquid column. Since the pathway of 1 mm is significantly shorter in the Nanodrop than the standard 10 mm path length of the cuvettes frequently used in UV-VIS spectrometry, the LOD is increased allowing for more concentrated samples to be measured [43]. The sensitivity is higher for the traditional UV-VIS spectrophotometer, while the Nanodrop allows for lower sample consumption, at the expense of sensitivity. With the traditional UV-VIS spectrophotometers, the LOD is 20 µg/mL, 0.1-1 µg/mL, and 0.05-0.5 µg/mL for A280, BCA assay, and Bradford assay, respectively [36]. For Nanodrop the LOD, depending on the sample to reagents ratio (v/v), is 100 µg/mL, 20 µL/mL, and 15 µL/mL for A280, BCA assay, and Bradford assay, respectively (**Table 1**).

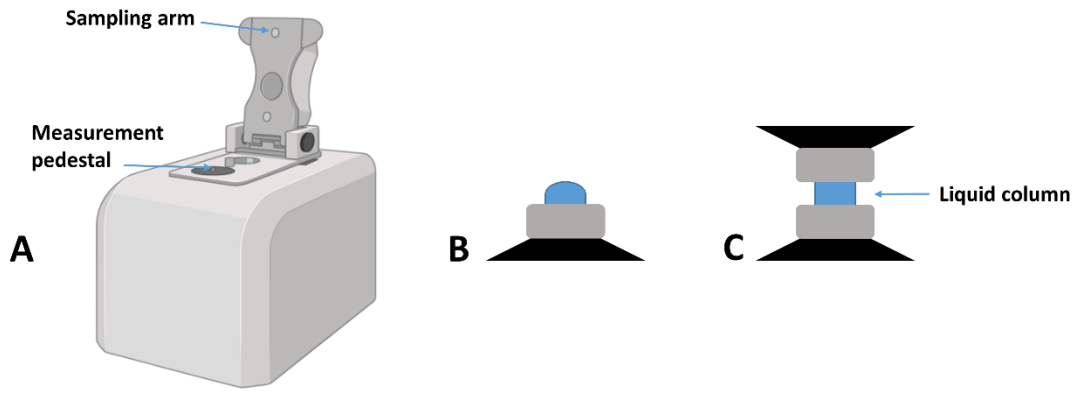

Figure 7. The Nanodrop UV-VIS spectrophotometer. A) The instrument with measurement pedestal and sampling arm. B) A sample aliquot on the measurement pedestal. The pedestal should be conditioned so that the droplet does not flatten out affecting the measurement. C) The sampling arm is lowered and a liquid column is generated where the light pathway is vertical.

Table 1. The detection range for different quantification methods for Nanodrop spectrophotometer instrument.

Method	Detection (µg/mL)	Minimum sample volume (µL)
A280*	100-40 000 [43]	2
BCA	20-200 (1:1 (v/v), sample:reagent) [44] 125-2000 (1:20 (v/v), sample:reagent) [44]	10
Bradford Assay	15-100 (1:1 (v/v), sample:reagent) [45] 100-1000 (1:30 (v/v), sample:reagent) [45]	10

* using bovine serum albumin (BSA) and molar absorptivity for BSA.

1.5 From proteins to peptides

The bottom-up proteomics strategies differ between in-gel-digestion (IGD) and in-solution-digestion (ISD). One common IGD method is sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The proteins are separated accordingly to their molecular weight using an electric field, and a gel as a support medium, enabling pre-separation and sample clean-up as low molecular weight impurities are not retained in the gel [46, 47]. However, there are disadvantages. Firstly, a higher concentration of protease is necessary, which can result in unwanted background noise from autolysis, which is the process where the protease cleaves itself [39]. Secondly, the handling of the gel increases the risk of contamination [47]. The gel requires a load of 20-50 µg proteins [46, 47], making the protocol unsuitable for limited cell samples. However, the following paragraphs (**1.5.1-1.5.3**) involve steps common in both IGD and ISD methods.

Furthermore, detergents need to be removed prior to enzymatic digestion as even small concentrations can inhibit proper enzymatic digestion. In addition, detergents are highly ionizable and will suppress the signal from peptide ions [48]. The clean-up can be performed by protein precipitation, e.g. by adding an organic solvent to the lysate (or other protein sources). By using cold acetone (-20°C) the proteins precipitate while lipids and small molecules remain in the supernatant [36].

After precipitation, proteins are often solubilized in 6-8 M urea to keep the proteins denatured and in solution. This concentration limits the activity of several enzymes, among them the enzymes responsible for protein degradation. To activate the enzyme such as trypsin, the sample needs to be diluted to <1 M Urea [36].

To prepare the proteins for enzymatic digestion the proteins need to be denatured by disrupting the weak nonpolar interactions (for chemically lysed cells usually performed at the cell lysis stage described in **Section 1.3.1**) and breaking the covalently disulfide bridges by reducing agents.

1.5.1 Reduction and alkylation

When the protein denatures, the disulfide bridges are still intact. Disulfide bridges are covalent bonds between the thiol group (-SH) of cysteine residues, that, in additional to noncovalent

weak interactions, stabilize the 3D structure of the proteins (**Figure 8A** and **B**). By reducing these covalent bonds, the proteins are more accessible for enzymatic digestion. DTT is a frequently used reducing agent. After reduction, an alkylating agent e.g. iodoacetamide (IAM) is added to avoid reformation and rearrangement of the disulfide bonds and to stabilize the denatured protein (**Figure 8C**) [36]. Reduction and alkylation take place with reagents in excess because of difficulties estimating the number of disulfide bridges in complex samples. To avoid over-alkylation, it is advised to quench the alkylation with low concentrations of thiols such as DTT [49, 50].

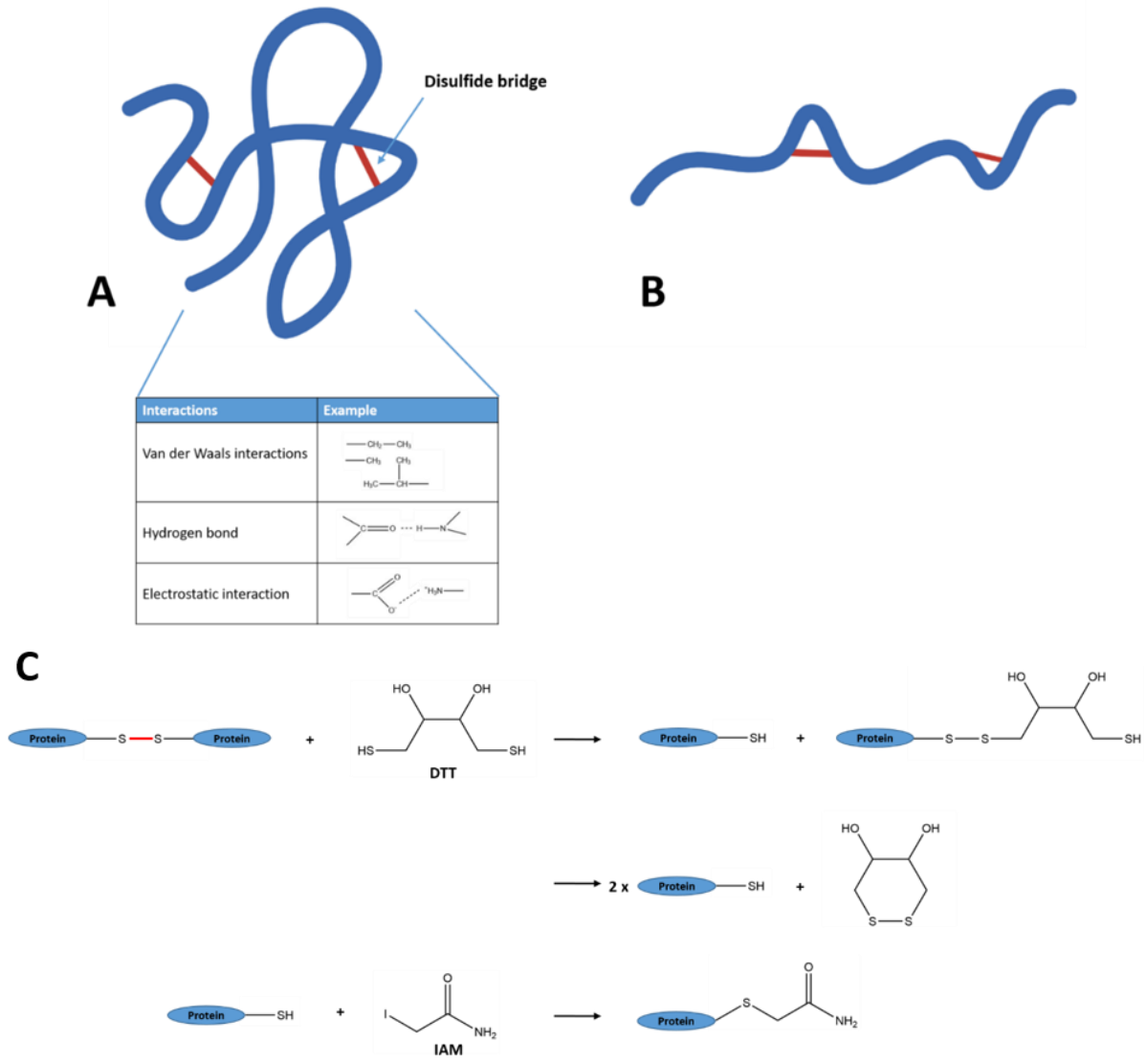

Figure 8. Protein stabilization and denaturation. A) The protein fold is stabilized by noncovalent weak polar interactions: Van der Waals interactions, hydrogen bonds, and electrostatic interactions. B) The covalent disulfide bridges (marked red) are still intact when the protein denatures. C) First the disulfide bridge is reduced using DTT, following S-alkylation («capping») of reduced cysteine thiols using IAM. The figure is adapted from [36].

1.5.2 Enzymatic digestion

Enzymatic digestion in bottom-up proteomics is performed by protease, which are enzymes that hydrolyze peptide bonds at specific sites, thus generating peptides for protein identification [36]. One of the commonly used proteases in MS-based proteomics is trypsin, an enzyme that cleaves the peptide bond at the carboxyl terminus (C-terminus) of arginine (Arg) and lysine (Lys), unless when followed by a proline (Pro) or when Arg or Lys is N-linked to aspartic acid (Asp) [39]. The stringent cleavage specificity creates peptides in the preferred mass range for the MS and ensures protein identification by acquiring mass spectra of the peptides and matching them against a database. In addition, the digest generated by trypsin yields peptides with a basic C-terminus residue, further enhancing positive ionization [51]. Trypsin can be considered a protein contaminant when digested into peptides in autolysis as it can suppress other peptide ion signals in the MS [52]. An optimum ratio between the protein and trypsin, substrate, and enzyme, ensures that the cleaving efficiency is kept high, yet preventing autolysis [39]. Studies have shown an optimum ratio of 1:50, yet for smaller sample sizes a ratio of 1:20 is sufficient [53, 54].

1.5.3 Desalting

Salts, buffers, and detergents interfere with the analysis, and sample clean-up, pre- and post-digestion are necessary to obtain MS data. Salts are present in the biological sample, e.g. sodium chloride, as well as added during the sample preparation. Detergents ionize well and are in excess relative to the peptides, causing ion suppression [33, 55]. Complete removal of detergent as SDS is considered difficult [48], and even with depletion of detergent before enzymatic digestion, desalting as the last step before LC-MS is often considered necessary.

Solid-phase extraction (SPE) is the most common sample clean-up strategy for bottom-up proteomics, with different types of extraction devices, sorbent, and sample capacity (amount of analyte that can be extracted) [55]. The extraction device, usually a tube, is filled with porous particles with a modified surface called sorbent. In proteomics, the reversed-phased SPE is frequently used, with C8 or C18 as the modified groups. The extraction principle is based on hydrophobic interaction and is analogous to reversed-phase chromatography [56].

As illustrated in **Figure 9**, the SPE procedure for reversed-phase material is usually carried out in the following steps: Firstly, the sorbent is conditioned using an organic solvent (e.g.

acetonitrile (ACN). Then the sorbent is rinsed with a polar solvent (e.g. water) with appropriate pH before the sample is applied, to eliminate polar interferences. The peptides are eluted with an organic solvent of appropriate pH and elution strength [56]. In cases when analyzing limited sample amounts, the use of SPE can cause sample losses [55]. In such cases, SPE material packed within a pipette tip can be employed. Commercially available SPE pipette tips are Omix by Agilent and ZipTip by Thermo Fisher Scientific. One of the disadvantages of SPE is poor retention of especially hydrophilic peptides on reverse-phase (RP) material, where they can be washed out with the washing solution [31].

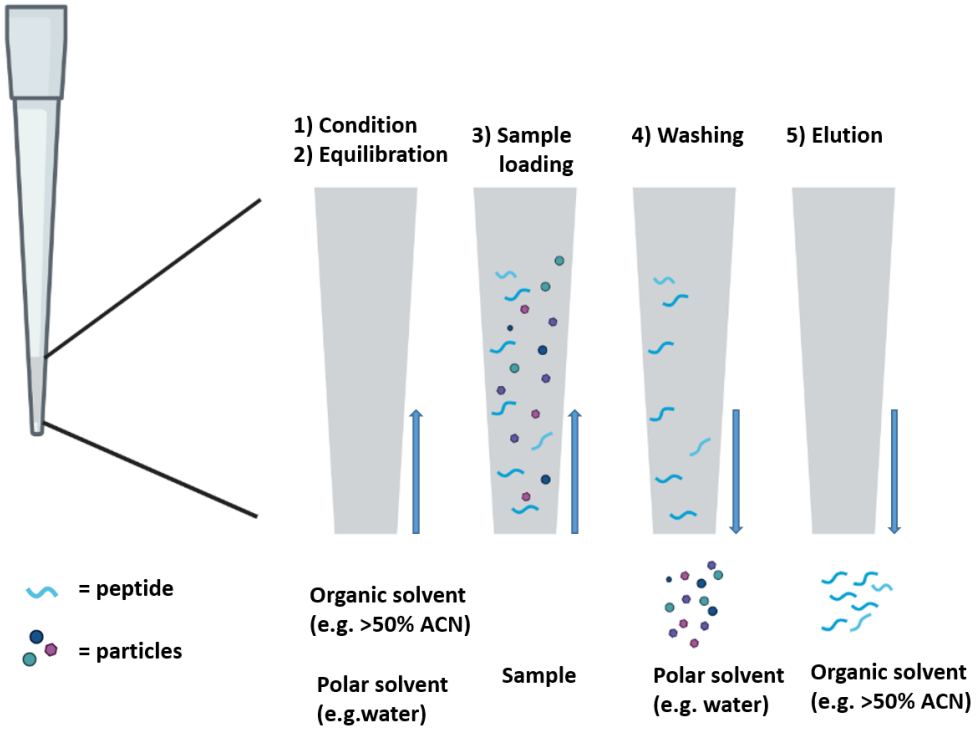

Figure 9. Common steps in solid phase extraction. The grey background is the total porous material with RP material. First, the material is conditioned with an organic solvent, e.g. ACN, before equilibration with a polar solvent. The sample is loaded. Then the material is rinsed for salts and polar contaminants. The sample is eluted in a new, clean vial with high organic sorbent.

1.6 Sample preparation methods for limited cell samples

1.6.1 Limiting sample size enables new insight into the proteome

Historically, global proteomics has been performed by bulk sample analysis consisting of millions of cells, providing a comprehensive protein expression [57]. The sample material when

working with cells is often scarce, as in the cases of clinical settings or working with primary cells. Research topics often require analysis of subcellular populations or single-cell [58], as bulk sample analysis does not take into account the heterogeneity of the cell sample, and therefore provides the average proteome profile [59]. The disadvantage of analyzing subcellular populations compared to single-cell is even samples of isolated cells of the same type will suffer from heterogeneity as the cells may be in different cellular states [60]. That means that the proteome of two single-cells of the same cell type may not be identical, causing the protein expression of single-cells to be lost (**Figure 10**) [61, 62]. Single-cell proteomics (SCP) enables a better understanding of the biological processes since the protein expression is attributed to the individual cell. In biomedicine, e.g. cancer research, the heterogeneity of the cell sample is an issue since the cells are in different microenvironments and thus have different functions and proteomes [26, 59]. As an organoid consists of a repertoire of cell types replicating the cell heterogeneity of an organ [17], the heterogeneity is not a severe issue as in SCP. However, since the material is often scarce the same consideration should be taken as for SCP and limited cell samples.

There are still technological limitations regarding limited samples and SCP. Recently there have been major improvements with regard to the sensitivity of the MS, and today the sample preparation is considered the bottleneck [63].

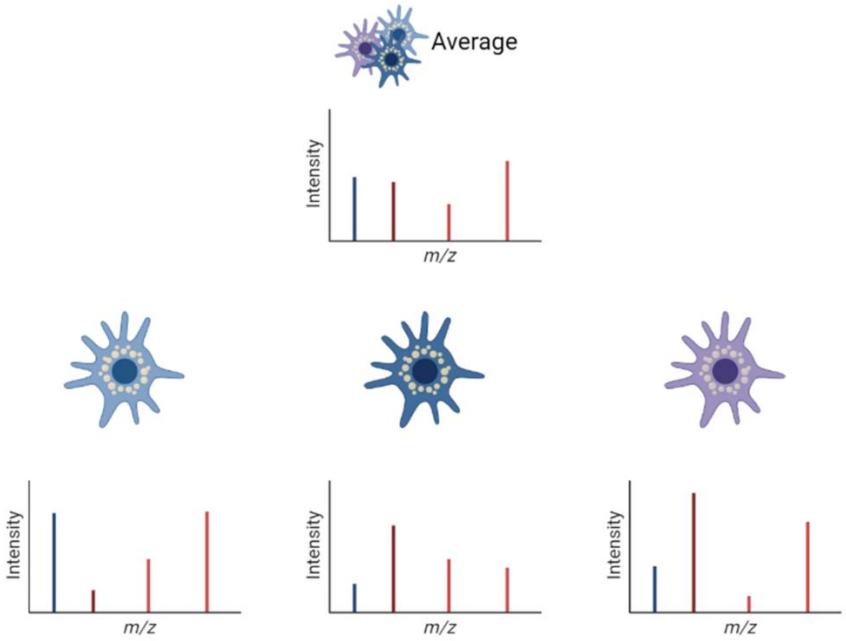

Figure 10. The MS[1] spectra in centroid mode of four proteins and how they differ in abundance between the single-cells and the bulk sample.

1.6.2 Method optimization using cost-efficient cells

Cell lines are commonly used in scientific research instead of primary cells, or when biomaterial is scarce because cell lines are cost-efficient and similar to the primary tissue. Since they are easy to culture, they provide a constant and unlimited supply of materials [64]. Immortalized hepatic tumor cell line, especially the HepG2 cell line, is popular within the scientific research field as it has retained most of the metabolic characteristics of hepatocytes [65, 66], which are the most abundant cell in the liver [67]. Another isolated cell line commonly used in research is the HeLa cells, which is the most used human cancer cell line [68]. Digested HeLa protein standard is often used as quality control for LC-MS analysis of proteomics samples [69, 70].

1.6.3 Considerations regarding limited cell samples

There is a tradeoff between sample size and protein coverage, yet recent advancements, with improved sensitivity and resolution, of the chromatographic separation and MS instrumentation allow for more proteins to be identified.

Today, the bottle-neck with SCP and limited cell samples lies in the sample preparation [59, 63]. As one single-cell contains around 200 pg of proteins, there are considerations to take compared to bulk samples. 2D systems and fractional schemes, used to improve the peak capacity, are typically avoided with limited samples as there is a significant loss of proteins between the dimensions [59]. Because of the hydrophobic nature of proteins, it is important to minimize sample loss to non-specific adsorption, by limiting the sample volumes, and the number of sample handling and transfer steps in the workflow [26, 63]. Several ISD approaches, e.g. the methods described in **Section 1.6.4** and **1.6.5**, have been developed to overcome the challenges associated with the removal of cellular debris, salts, and detergents which are incompatible with downstream analysis [26].

1.6.4 Single-pot, solid-phase-enhanced sample preparation

There are devices-based approaches to remove interfering compounds before enzymatic digestion. Single-pot, solid-phase-enhanced sample preparation (SP3) is a paramagnetic beads-based sample preparation technique for proteomics (**Figure 11**). The denatured proteins are aggregated and immobilized on the surface of carboxylate-coated beads [71, 72]. These beads are available on the market with different levels of hydrophobicity, such as the Sera-Mag

SpeedBeads by Cytica. By using beads of different levels of hydrophobicity, in theory, a better proteome coverage can be obtained and potential differences between the bead types can be eliminated. SP3 provides sample cleanup before enzymatic digestion, enabling efficient digestion, and minimal sample loss compared to the traditional SPE sample cleanup. It is easy to ensure compatibility between the sample capacity provided by the beads, sample volume input, and protein concentration. The recommended working volume is 20-100 µL with a protein concentration of 1-50 µg [73]. This makes SP3 a suitable technique for limited protein samples. Still, SP3 has its limitations, one of them being sample loss in the rinsing steps, especially in cases where the pH changes [74].

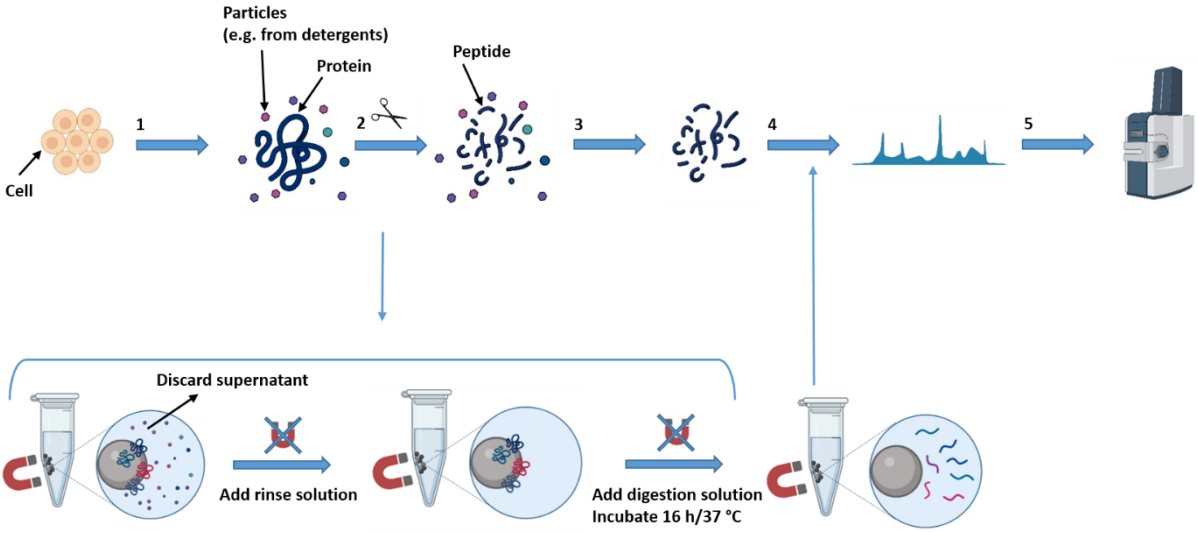

Figure 11. SP3 (bottom) compared with the standard in-solution method (top). The numbers represent the common steps in a sample preparation procedure where 1) cell lysis, 2) enzymatic digestion, 3) sample clean-up (this can also take place before enzymatic digestion), 4) separation of peptides with LC, and 5) analysis with high-resolution MS. The sample clean-up in SP3 takes place prior to enzymatic digestion using paramagnetic beads. The supernatant containing peptides is directly injected into the LC-MS system.

1.6.5 Sample preparation by easy extraction and digestion

Other methods focus on minimal sample handling and increased cell lysis efficiency for higher protein yield. Sample preparation by easy extraction and digestion (SPEED) is a method developed to overcome the challenges linked to the removal of detergents and chaotropic agents that are incompatible with MS and can interfere with enzymatic digestion. SPEED is a detergent-free procedure, using trifluoroacetic acid (TFA) for cell lysis and complete sample dissolution (**Figure 12**). For cell samples, the lysis takes approximately 2-3 minutes at RT, reducing the time of lysis significantly compared to common chemical lysis methods. The alkylation and reduction agents are combined, added to the samples, and incubated for a short

time at a high temperature (95 °C) before the enzymatic digestion. SPEED provides a rapid sample preparation with minimal sample loss, where all steps are performed within one vial [53]. One of the disadvantages of the method is the high sample volume obtained due to dilution during the sample preparation, with the potential loss of hydrophobic peptides [54].

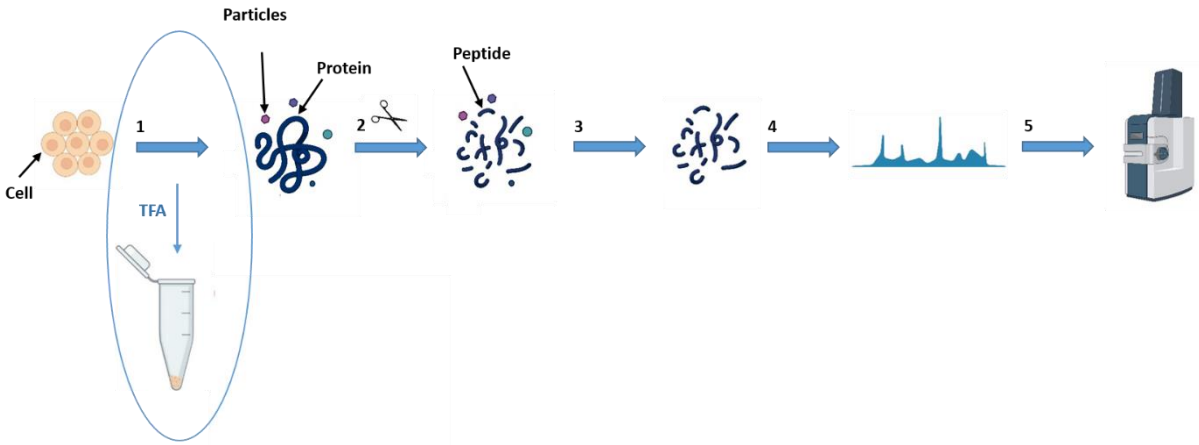

Figure 12. The SPEED method uses the same steps common in sample preparation, where the main difference is in the lysis method (circled in blue). SPEED uses TFA which is a strong acid to lyse the cell sample within minutes. By using acid and avoiding detergents, interfering compounds are limited, as illustrated by the fewer particles compared to the SP3 method (Figure 11).

1.7 Separation of peptides

Separation of a complex sample is necessary before detection. By having a dimension of separation prior to the MS, more peptides can be detected. Chromatography is a separation technique where compounds in a mixture are distributed differently between a mobile phase (MP) and a stationary phase (SP). The separation will occur because the compounds have different affinities to the SP. There are several separation principles in LC, depending on the interactions between SP and the compounds in MP. Reversed-phase (RP)-LC is the most common separation principle in proteomics [28, 75]. A liquid MP pumps the compounds through a column with SP containing surface-bound hydrocarbon chains, where the most commonly used in RP-LC is C18. The SP is usually bound to silica-based totally porous particles [76]. By increasing the organic content in the MP, the peptides are separated according to their hydrophobicity. Hydrophilic peptides will have a poor affinity to the stationary phase and elute from the column before the more hydrophobic peptides [33].

1.7.1 Nano-liquid chromatography increases sensitivity

The sample will be diluted during the chromatographic process, as shown in **Equation 2**. By reducing the ID of the column, the eluting band becomes more concentrated because of reduced radial dilution (**Figure 13)**. In addition, will lower flow rate required by the miniaturized system to avoid backpressure, increase ionization efficiency in the electrospray ionization (ESI) (Described in **Section 1.8.1**) [76]. Since the compounds are entering the detector in concentrated bands, the sensitivity (signal intensity) is increased when coupled with a concentration-sensitive detector like the ESI-MS. Because biological samples often are small in sample size and protein amounts, narrow columns coupled with concentration-sensitive detector such as ESI-MS is often employed in proteomics [77]. These columns, with inner diameter (ID) ≤ 100 μm are termed nanoLC columns [76, 77]. As with the conventional RP-LC columns, the particle-packed column is dominating the nanoLC-marked [75]. In addition to particle-packed columns, there are several formats in nanoLC such as monolithic (spongelike structure), open tubular columns, and micropillar array columns (μPAC), which provides low backpressure, and increased sensitivity and efficiency [77].

Dilution can be expressed as follows:

$$D = \frac{C_0}{C_{max}} = \frac{\varepsilon\pi r^2(1+k)\sqrt{2\pi LH}}{V_{inj}} \qquad \textbf{Equation 2}$$

Where C_0 is the concentration of the sample, C_{max} is the concentration of the sample during detection, ε is the porosity of the particles, the inner radius of the column, L is the column length, H is the plate number and V_{inj} is the injection volume. As illustrated in **Equation 2**, the most efficient way of reducing dilution is by reducing the ID of the column [76].

If the column ID changes, and every other parameter of the column and the $V_{inj,}$ stays the same, a dilution factor between the two columns can be expressed with **Equation 3**:

$$D = \frac{ID_1^2}{ID_2^2} = \frac{ID_{conv}^2}{ID_{nano}^2} \qquad \textbf{Equation 3}$$

Equation 3 can be used to compare the dilution when switching from the conventional column to nanoLC column, expressed ID_{conv} and ID_{nano}, respectively. A column with an ID of 0.05 mm will be nearly 1800 times less diluted compared with a column with an ID of 2.1 mm. The minimized dilution represents a theoretical gain in sensitivity [28, 78].

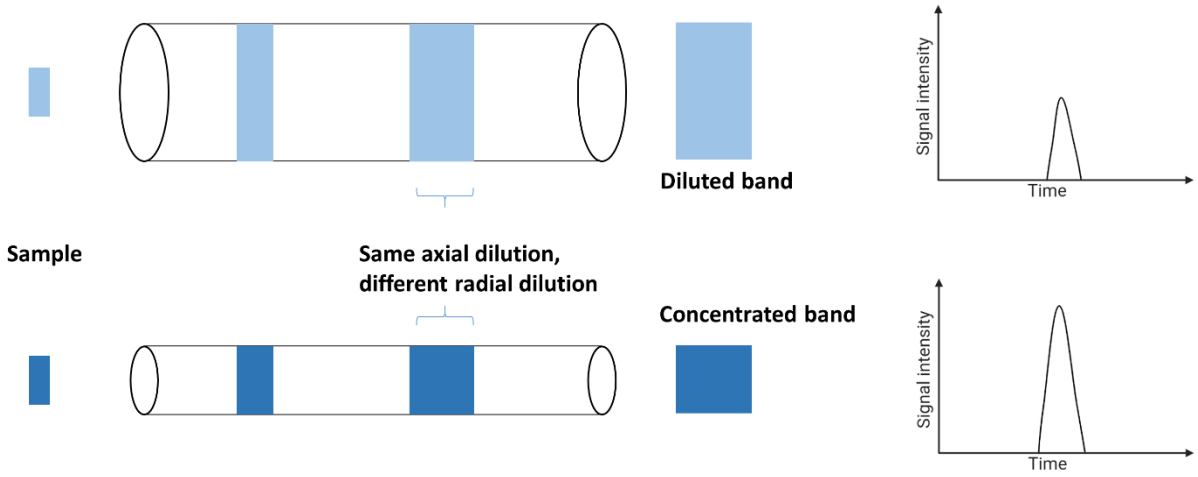

Figure 13. How sensitivity is dependent on the ID of the column. A packed column with particles of the same size has the same efficiency, but different sensitivity.

1.7.2 Micropillar array column increases efficiency

Physical processes cause band broadening, where the extent of band-broadening can be described as efficiency [79]. Plate number (N) (**Equation 4**) or plate height (H) (**Equation 5**) can express column efficiency. A column with high efficiency will provide sharp and narrow peaks, and thus increase signal intensity.

The relation between N and H is as follows [79]:

$$N = \frac{L}{H} \qquad \textbf{Equation 4}$$

Where L is the column length. A column with high efficiency has a high N and low H and can be obtained with longer columns and/or limiting the band-broadening contributions.

The contribution of band-broadening can be expressed by the van Deemter equation [79]:

$$H = A + \frac{B}{u} + Cu \qquad \textbf{Equation 5}$$

Where u is the linear flow rate. Eddy diffusion (A) happens when the same compounds will have different path lengths through the column because of irregularly sized and placed particles. Longitudinal diffusion in the MP, also known as axial dilution (**Figure 14**) is expressed with B, and C is the resistance to mass transfer in MP and SP. Stagnant MP is neglected in nanoLC systems as this is only valid with larger particles [79]. NanoLC is more technically demanding

[77], and all possible sources of extra-column dead volume need to be reduced substantially as the band-broadening effect is more prominent in nanoLC compared to conventional LC [80].

As illustrated in **Equations 4** and **5**, higher efficiency can be obtained by increasing the column length. With packed and monolithic columns, this comes at the expense of higher backpressure. The long column of several meters provided by open tubular columns has the disadvantages of low sample capacity and being technically demanding [77, 81]. A means to overcome the limitation of high backpressure, yet still have a high efficiency is the µPAC column, which is commercially available by PharmaFluidics, now acquired by Thermo Fisher Scientific. The column consists of ordered pillars with a porous or solid core, coated with C18 for RP separations. The pillars with the solid core are analogous to the core-shell particles often used in ultra-high-performance liquid chromatography (UHPLC) [82]. The highly homogenous ordered pillars reduce eddy dispersion compared to packed columns (**Figure 14**) [83]. However, wall and race-track effects (turns) can contribute to band-broadening [84]. Improved design has minimized the losses due to the race-track effect [84] and there are improvements aiming at minimizing the wall effect [85]. With a column length of 50-200 cm, pillars with a diameter of 5 µm and interpillar distance of 2.5 µm, µPAC provides increased peak capacity (how many theoretical peaks that can be separated within a retention window), efficiency, and sensitivity compared to particle-packed nanoLC columns [83, 86]. The pitfalls include the big price tag of the commercial µPACs [77].

Figure 14. Efficiency compared between packed column (left) and µPAC column (right). The figure is adapted from [77].

1.8 Detection of peptides using mass spectrometry

Because the sample is highly complex, the peptides will coelute even with a LC-column of high peak capacity. Mass spectrometer is a universal detector that separates and detects gas phase ions by their *m/z*, thus enabling the peptides to be separated by their molecular mass (**Figure 15A**). An MS consists of an ion source, mass analyzer(s), and a detector. The task of the ion

source is to transfer the analytes from the liquid phase into gas phase ions, while mass analyzer separates the ionized compounds. Some mass analyzers can subject the ions to collision-induced dissociation (CID), where the ions will collide with a buffer gas, usually argon, causing fragmentation. From an MS run, a full scan spectrum (MS^1) can be obtained, where *m/z* is detected at a given time point as the peptides are eluting from the chromatographic system. The fragments caused by CID, give rise to a tandem MS (MS^2 or MS/MS) spectra. The ions entering the MS are the precursor ions, that fragment into smaller product ions by CID (**Figure 15A**). From now on, the peptide ions are annotated as precursor ions. The MS^2 spectra are necessary to determine the amino acid sequence and the structural information of the peptides and proteins (**Figure 15B**). The spectrum is considered the "fingerprint" of the peptide, as these spectra are often reproducible and can be compared to a library or dataset in order to qualitatively determine the proteins [87-89].

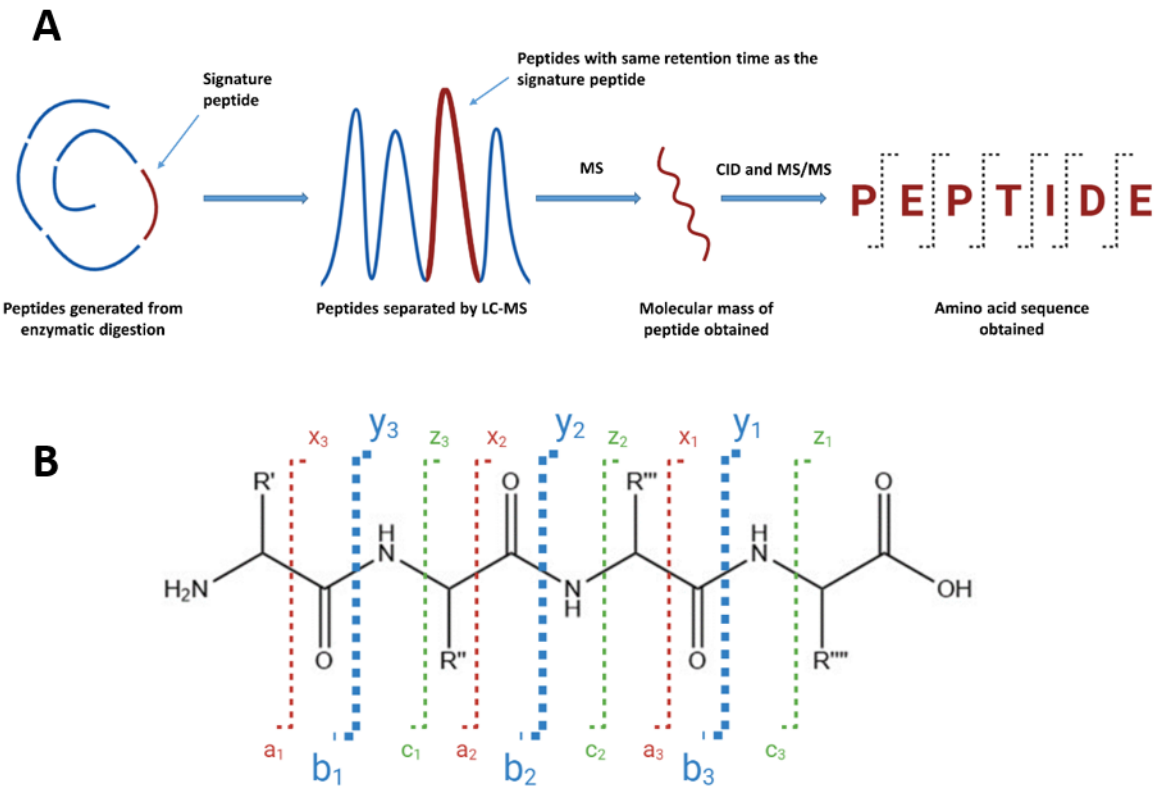

Figure 15. A) A revisit of the principle behind bottom-up proteomics. From proteins and peptides in sample to peptide identification with MS^2 (in the figure annotated as MS/MS). In order to identify a protein, its signature peptide need to be detected [51]. A signature peptide is unique for the protein. The MS^2 makes it possible to determine the amino acid sequence of the peptide. B) Fragmentation pattern of peptides during CID. The b and y ions (peptide bonds) are the most abundant ions in low-energy CID which is the most used fragmentation mode in peptide sequencing. The figure is adapted from [90].

Due to sample complexity, high-resolution (HR)-MS is widely used in proteomics. Mass resolution is a measure of the ability of the instrument to separate one precursor ion (with a given *m/z*) from another [91]. Fourier-transformation-based mass analyzers, like Orbitrap, and time-of-flight (TOF) are considered HR [92]. Quadrupole mass analyzers are often used in combination with the before-mentioned mass analyzers. Proteomics on a limited sample amount requires an MS sensitive enough to detect low protein concentrations. Because the abundance of different proteins (and thereby peptide ions generated) varies by several orders of magnitude, the instruments need to have a high dynamic range of at least 10^8 [25]. The dynamic range is defined as the difference between the highest and lowest signal intensities observed within the same MS/MS scan [89]. The technical limitation makes answering biological questions more difficult as the low abundant proteins play an important role in the regulation of the biological system [25]. The peptides generated from bottom-up sample preparation generate a high number of precursor ions entering the MS. High complexity samples demand MS instruments which obtain a high number of MS^2 data [89]. Challenges with MS-based SCP are a high amount of missing data and low reproducibility from one MS run to another [93].

1.8.1 Electrospray ionization

ESI is a common ion source used in proteomics. Although considered an ion source, the analytes are ionized prior to entering the LC-ESI-MS system by adjusting the pH in the MP. By the inlet of the ESI chamber, there is an emitter connected to an electrode. The emitter introduces the sample and MP to the ion source. In addition, nebulizing and drying gas, mostly N_2, are introduced to facilitate droplet formation. When an electrical potential is applied in the positive mode, the cations are pushed out of the emitter toward the counter electrode at the MS inlet, the surplus net charge creating a Taylor cone, which ejects charged droplets. As the repulsion forces between the cations within the droplet exceed the surface tension, the droplet will divide into smaller droplets. This is known as Coulomb fission, and the process is repetitive, yielding smaller and smaller droplets [76]. The gas phase ions are generated by evaporating from the droplet (ion evaporation model) or are left over after complete solvent evaporation (charge residue model) [94].

The miniaturization of the LC-MS system allows for a reduced flow rate, which increases the ion efficiency of the ESI. The downscaled ESI is known as nanoESI and is similar to the

conventional ESI. The biggest difference is with the emitter, which in nanoESI has a narrow ID allowing for lower flow rates leading to the formation of small droplets (**Figure 16**) [76]. Because the droplets are small, there is no need for employing a nebulizing gas with nanoESI. The smaller droplets provide an increased formation of gas-phase ions occurs before entering the MS inlet, allowing more ions to be detected and analyzed by the MS. As the ionization efficiency is increased, higher sensitivity is achieved [95].

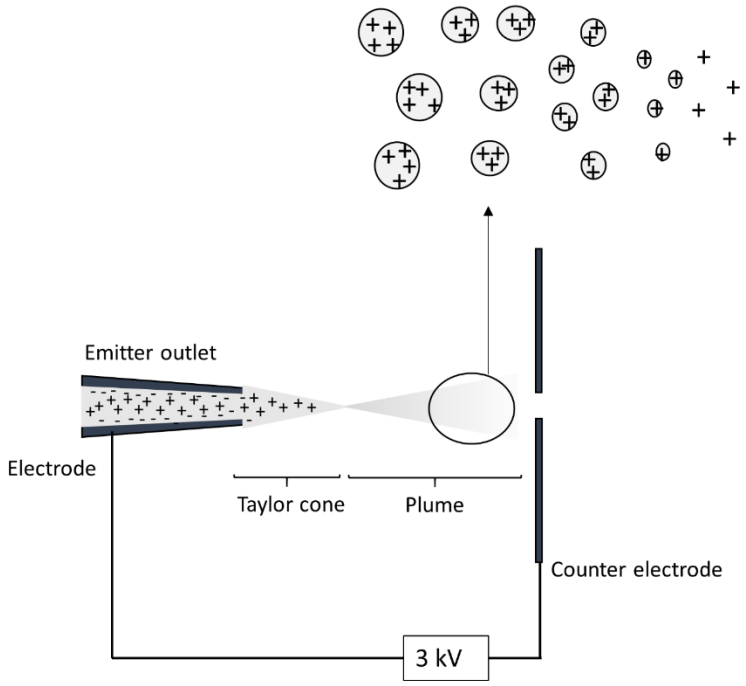

Figure 16. The nanoelectrospray ionization with the formation of gas phase ions in positive mode. Adapted from [76].

1.8.2 Trapped ion mobility time-of-flight mass spectrometry

The trapped ion mobility spectrometry time-of-flight (timsTOF) MS consists of a trapped ion mobility spectrometry (TIMS) device, followed by a quadrupole, collision cell, and time-of-flight (TOF), see **Figure 17**. TIMS provides an extra dimension of separation and enables new insight into the proteome [96].

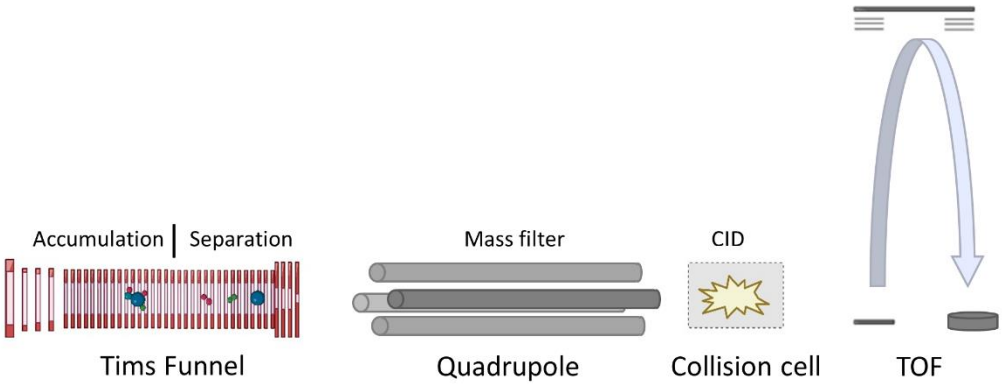

Figure 17. A schematic figure of the timsTOF-MS. First, the analyte ions enter the tims funnel where they are accumulated before separation by their ion mobility. Then ion packets are sent to the quadrupole where they are separated by the m/z. CID in the collision cell ensures an MS^2 fragmentation. Ions are separated by the m/z in the TOF funnel before detection.

TIMS separate gas phase ions accordingly to their ion mobility. The ion mobility is defined as the steady-state drift velocity of the ion under the influence of an applied electric field. In other words, ions are separated by their *m/z* as well as their shape. When entering the TIMS funnel, a gas flow pushes the ions forward. Larger ions, which have the lowest mobility, will experience more drag-force because of the higher collisional cross-section. The ions are held back by an electric field, increasing in strength along the funnel. Ions are at rest when the drag force on the ions is opposite of the force from the electric field. The ions of low mobility need a stronger electric field to counteract the high drag force it experiences, meaning the low-mobility ions will be trapped near the funnel exit and hit the detector before the high-mobility ions [97].

The TIMS funnel consists of two sections, where trapping of ions takes place in the first section, while simultaneously separating the ions according to their ion mobility in the second section. The ions are introduced in a stepwise manner, by decreasing the electric field strength, to the quadrupole where they are fragmented. Because of the TIMS device and the extra dimension of separation, it provides, even isobaric peptides (peptides with the same *m/z*) and isomers of PTMs can be separated [98].

The quadrupole is a mass analyzer, which is capable of scanning, filtering, and fragmenting ions depending on the electrical potential and mode applied. Quadrupole is the only mass analyzer able to mass-select ions [99]. The quadrupole consists of four rods parallel to each other, where the opposite pairs are connected electrically. By applying the right values of a direct current and a radio frequency, ions with specific *m/z* will stay in between the rods. The rods change polarity periodically, giving the ion a stable trajectory toward the detector. This way the quadrupole acts as a mass filter, discarding all the ions with unstable trajectories [100].

Quadrupole is often incorporated in hybrid MS instruments with mass analyzers that by themselves cannot perform tandem MS. This includes HR-MS as Orbitrap and TOF. By coupling quadrupole with these mass analyzers, both selectivity and high resolution can be achieved [89].

Time-of-flight (TOF) MS separates the ions after fragmentation by their time of flight through a field-free tube, where lighter ions are detected before the heavier ones. The mass resolving power is dependent on the path length of the field-free tube. In timsTOF, the effective length of the tube is 3.5 m [101]. Ultra-high-resolution (UHR)-TOF is often orthogonally oriented, providing a longer flight path for the ions and thus increasing the resolving power compared to the linear TOF. UHR-TOF has the fastest high-resolution spectra collection and the highest mass range among the mass analyzers [102].

The disadvantage with instruments utilizing mass-selective quadrupole is that only one precursor ion can be scanned at a time while discarding all the other ions that were accumulated together. To overcome this, parallel accumulation-serial fragmentation (PASEF) is used in conjunction with TIMS allowing for the fragmentation of multiple precursors within the same scan, thus increasing the efficiency of MS^2 acquisition. The quadrupole is set to the *m/z* of a precursor eluting from the TIMS funnel. When the precursor ions have eluted, the quadrupole rapidly switches to the *m/z* of the next precursor. In this way, PASEF selects several precursors within one scan [98, 103]. The parallel trapping and separation increase the duty cycle by up to 100%, meaning that in a cycle nearly 100% of the incoming ions can be used [97, 104].

1.8.3 Data acquisition strategies for proteomics

There are three acquisition strategies: One targeted method using multiple reaction monitoring, and two untargeted methods using either data-dependent acquisition (DDA) or data-independent acquisition (DIA). To recover a complete proteome profile, untargeted data acquisition methods are used. In DDA the whole mass range is repeatedly scanned in MS^1 during the chromatographic gradient. The full scan spectrum generates a list of the N most abundant precursors, usually the top 10-20, at every point in time subjected to CID and MS^2 analysis [25]. The dependence on the signal intensity will cause co-eluting, low-abundant peptides to not be selected for fragmentation, thus decreasing the proteome coverage. Often, proteins of diagnostic and biological importance are in low abundance, leading to a limited understanding of the biological systems. Another pitfall of DDA is the reduced intensity of the

full scan spectrum since the MS uses more time acquiring the MS^2 spectra [105]. DIA can obtain the MS^2 spectra for all the precursors by having a widened isolation window, enabling simultaneous fragmentation of several precursors, and acquiring spectra from all precursor ions in the sample. Since several precursors are fragmented in the same isolation window, complex MS^2 spectra are generated. These spectra need to be deconvoluted in order to determine the peptide sequence, making peptide identification more challenging [106]. DDA is the most used strategy for global proteomics [106] as this has fewer technical and computational requirements compared to DIA.

1.8.4 Database search

The data needs to be processed to assign the peptides to their corresponding proteins. This is accomplished using software like MaxQuant [107], Proteome Discoverer [108], Scaffold [109], or PEAKS [110]. The peptide needs to be unique for the protein in order to be assigned. These peptides are called signature peptides [51]. The software offers statistical and graphical tools and sets restricting parameters for the search. Using Scaffold as an example, protein threshold, peptide threshold, a minimum of signature peptides, and false discovery rate have to be set in order to use the search tool [109, 111]. The software uses a database search engine, among the most widely used being Mascot [112], where the acquired MS^2 spectra are searched against theoretically computed MS^2 spectra of peptides. A high peptide spectrum match (PSM) score means that there is a good correlation between the experimental and theoretical spectrum, and indicates that the spectrum is assigned to the right peptide. The data is also searched against an altered sequence databased to establish the false-discovery rate (FDR) [112]. When the proteins are identified, their subcellular locations, cellular function, protein class and pathway can be annotated through the Gene Ontology project [113].

1.8.5 Quantitative proteomics strategies

Comparing spectra against a database provides qualitative information about the proteins in the samples, yet this information alone is often insufficient and quantitative information is necessary to obtain, e.g. quantitate a cellular response. Due to the overwhelming number of proteins, it is challenging to use internal standard, which is commonly used in targeted approaches to eliminate biases introduced in sample preparation (e.g. lysis, digestion, clean-up

steps). Hence without the use of internal standards, quantification in global proteomics is considered relative quantification or semi-quantitative [54].

There are two quantification approaches in proteomics: label-free quantification (LFQ) and labeled approaches (**Figure 18**). A common way to label peptides is by isobaric labeling using e.g. tandem mass tags (TMT) or isobaric tags for relative and absolute quantification (iTRAQ) [114]. In the labeled approach, the samples are combined (multiplexed) before the LC-MS run, making the total runtime shorter and increasing the signal intensities since the combined peptide concentration can be above the LOD of the instrument. TMT chemical labeling is based on the substitution of a bipartite adduct, which binds to the N-termini of peptides and lysine side chains generated after enzymatic digestion in a bottom-up workflow [115, 116]. The bipartite adduct consists of a reporter and a mass balancer group, where the mass of the adduct is constant. What varies in mass is the mass and reporter group individually by using isotopes of C, N, and O (^{13}C, ^{15}N, and ^{18}O, respectively). The signal for a specific *m/z* in MS^1 is of the precursor from all the samples. The bond between the balance group and the amine-specific reactive group, and the bond between the balance and reporter group, will break during CID. The neutral balancer group is lost, while the charge remains in the reporter group. The reporter group ions are responsible for peaks in the MS^2 spectra. These peaks are used to identify and quantitate the individual peptides from the multiplex set [115]. The limitations of tagging peptides include the cost of reagents, additional sample preparation steps, and the limitation of the number of samples depending on the labeling kit [117].

The most straightforward method concerning sample preparation is the LFQ. The samples are analyzed one by one in different MS runs. LFQ is usually performed by spectral counting which measures the frequency of a peptide identification for a protein, or summarized peak areas/intensity for the peptides belonging to the protein [25, 117]. In contrast to labeled approaches, is LFQ affected by run-to-run variations (such caused by parameters as temperature, column condition, and pressure) [118].

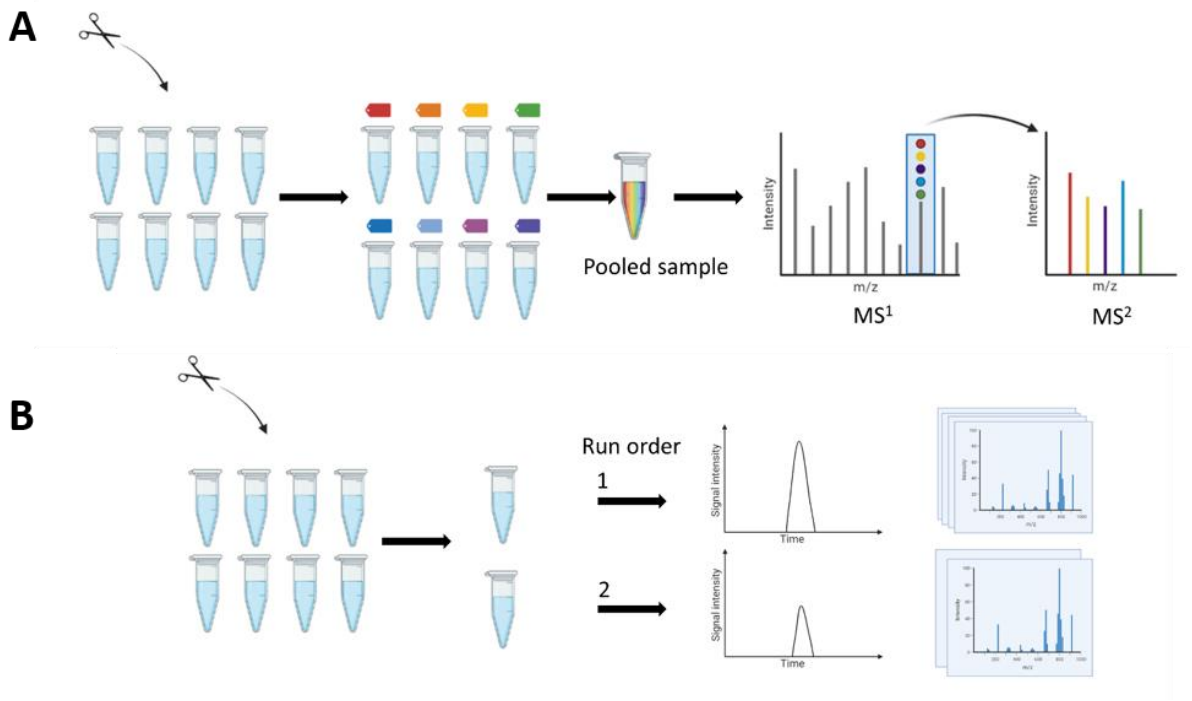

Figure 18. A) Labeled approach. After enzymatic digestion the labels, here illustrates as different colored tags, are added to each sample. The tags bind to the N-termini of peptides and the samples are pooled together. The relative quantification is performed after one run. The figure is adapted from a BioRender template. B) LFQ. After sample preparation the samples are analyzed with LC-MS one by one, here illustrated with two samples. The relative quantification is performed by comparing signal intensities and/or spectra counting. The figure is adapted from [119].

2 Aim of study

The aim of this study was to optimize and establish a sample preparation method for analyzing limited biological samples for proteomics. Sample preparation methods such as ISD-Urea, SPEED, and SP3 were evaluated using hepatic cells (HepG2). The spectrophotometric methods A280 and BCA Assay were also evaluated for protein quantification. The samples were analyzed with nanoLC and µPAC columns, and timsTOF-MS. Finally, the chosen method was applied for liver organoids as a first step toward NAFLD studies (**Figure 19**).

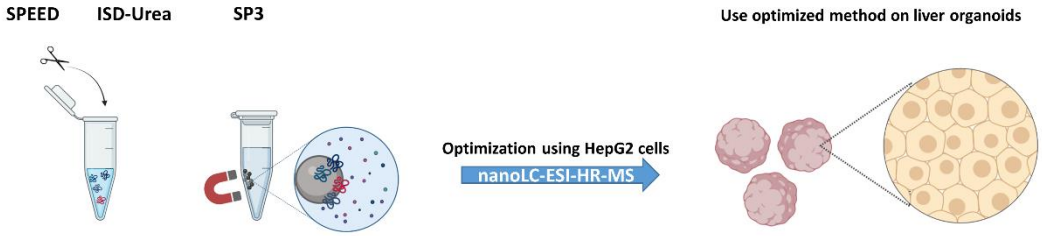

Figure 19. Visualization of the aim of the study in this work. SPEED, ISD-Urea, and SP3 were compared by using HepG2 cells and tested on liver organoids, as a preliminary step towards NAFLD studies.

3 Experimental

3.1 Consumables, materials, and small equipment

Falcon centrifuge tubes (15 mL and 50 mL) and autosampler vials (0.3 mL) were from VWR. Safe lock tubes (1.5 mL) were from Eppendorf (Hamburg, Germany). The low-binding safe lock tubes (0.5, 1.5, and 2.0 mL) and pipette tips were from Sarstedt (Nümbrecht, Germany).

Omix C18 10 µL pipette tips were from Agilent (Santa Clara, CA, USA). Sera-Mag SpeedBeads were from Cytica (Emeryville, CA, USA), while the magnetic rack, Dynal MPC™-E, was from Invitrogen under Thermo Fisher Scientific (Waltham, MA, USA).

Thermo-shaker PSC24 was used for all the incubation steps and was from Grant instruments (Shepreth, United Kingdom). The test tube shaker was from Heidolph instruments (Schwabach, Germany). Minicentrifuged Mini Star was from VWR. The vacuum centrifuge Concentrator plus and centrifuge 5424 R were from Eppendorf. The ultrasonic bath was from Limpieza por Ultrasonidos - ATU (Paterna, Spain).

3.2 Chemicals

High-performance liquid chromatography (HPLC) water, from here on defined as water, ACN, ethanol and formic acid (FA) (≥99%) were from VWR (Radnor, PA, USA). TFA (99.8%), IAM, DTT, TX100 SigmaUltra, SDS, SIGMAFAST™ Protease inhibitor tablets, NaCl, EDTA titriplex (99.995%), benzonase nuclease, Trizma base (Trisbase) and ammonium bicarbonate (ABC) were from Sigma-Aldrich (St. Louis, MO, USA). Tris-HCl buffer (1 M) was provided by Oslo University Hospital. BCA™ Protein Assay Kit – Reducing Agent Compatible, Pierce™ HeLa protein digest standard and Gibco™ Dubecco's phosphate-buffered saline (PBS) were from Thermo Fisher Scientific. Trypsin Gold, mass spectrometry grade was from Promega (Madison, WI, USA).

The HepG2 cells and liver organoids were kindly provided by Ph.D. candidate Ingrid Wilhelmsen and Ph.D. Aleksandra Aizensthtadt, respectively, at Hybrid Technology Hub (HTH).

Water was filtrated through a Milli-Q ultrapure water purification system by Millipore (Bedford, MA, USA), generating type 1 water.

3.3 Solutions

3.3.1 Lysis buffers

TritonX-100 (1%) lysis buffer

A solution of 5 M NaCl was made by dissolving 29.2 g of NaCl in 100 mL of type 1 water. EDTA (0.1 M) was prepared by dissolving 0.37 g of EDTA in 10 mL of type 1 water. The TX100 (1%) buffer was prepared in a 50 mL Falcon tube by adding 2.5 mL of 1 M Tris-HCl (pH 8), 1.5 mL of 5 M NaCl (29 g in 100 mL water), 43 mL of type 1 water, 2.5 mL of 0.1 M EDTA (0.37 g EDTA in 10 mL water) and 0.5 mL TX100. The buffer was frozen in 10 mL tubes at -20°C until use. When needed, an aliquot was thawed on ice and a protease inhibitor tablet was added. The buffer containing protease inhibitor was then again aliquoted into 1 mL tubes. If precipitation was observed after thawing, the buffer was discarded and a new tablet of protease inhibitor was dissolved in thawed aliquot of TX100.

Sodium dodecyl sulfate (1%) lysis buffer

The buffer contained 500 µL of 10% SDS, 50 µL 1 M DTT (77 mg in 500 µL water), 500 µL 1 M Tris-HCl (pH 8), one protease inhibitor tablet and 3950 µL of water. The 10% SDS solution was made by adding 50 mg SDS in 500 µL of water.

3.3.2 Solutions for reduction and alkylation

The solution of DTT and IAM were aliquoted into 50 µL aliquots and stored at -80 °C. The DTT solutions contained either 200 mM (30.8 mg) in 1000 µL 0.1 M Tris-HCl (pH 8) or 0.5 M (77.1 mg) dissolved in 1000 µL water. The IAM solutions contained either 200 mM (37.0 mg) in 1000 µL 0.1 M Tris-HCl (pH 8) or 1 M (92.5 mg) in 500 µL water. The 200 mM solutions were used for ISD-Urea, while the other solutions were used for SP3 and SPEED. If concentrations of <0.5 M or <1 M of DTT or IAM, respectively, were needed, the stock solution was diluted with water.

3.3.3 Trypsin

The trypsin stock solution was prepared by dissolving 100 µg of trypsin in 100 µL of 50 mM acetic acid, making a stock solution of 1 µg/µL trypsin. Aliquots of 10 µL were stored at -80°C until use. The aliquots did not exceed more than 5 freeze-thaw cycles.

3.3.4 Other solutions

A stock solution of 200 mM ABC buffer was prepared by dissolving 1.58 g ABC in 100 mL type 1 water. Urea (6 M)/ABC (100 mM) were prepared on the day of the experiment for ISD-Urea by dissolving 31.6 mg urea in 100 µL of 100 mM ABC. The neutralization solution (2 M) for the SPEED method was made by dissolving 2.4 g of Trisbase in 10 mL water. For the micro solid phase extraction, the conditioning and elution solutions were 0.1% TFA in 50% ACN, and the equilibration and washing solution were 0.1% TFA in water. An ampule (20 µg) HeLa digest standard was dissolved in 200 µL 0.1% FA, yielding a solution of 100 ng/µL peptide solution. Aliquots of 20 µL were stored at -80°C until use.

3.4 Sample preparation for limited cell samples

How to prevent contamination during sample preparation, see **Section 7.2.1 in the Appendix**.

3.4.1 Redistribution of cells

The HepG2 cells were cultivated by Ph.D. candidate Ingrid Wilhelmsen, snap-frozen after harvesting, and delivered in a pellet of 4 700 000 cells (**Section 7.2.2 in the Appendix**). Cells were redistributed at the Department of chemistry by the following procedure.

After thawing on ice, the cells were resuspended in 2350 µL PBS, making a solution of 2000 cells/µL, before dividing them into 3 aliquots of 1 000 000 cells, 2 aliquots of 500 000 cells, and one aliquot of 700 000 cells. After aliquotation, the cells were pelleted by centrifugation at 20 000 g for 5 min at 4ºC. The supernatant was discarded before storing the cell pellet at -80°C until use. When the time for the experiment, the cell pellets were thawed on ice and resuspended in PBS.

For the initial screening samples of approximately 50 000, 100 000, 150 000, and 200 000 cells (n=1) an aliquot of 500 000 cells was dissolved in 250 µL PBS and distributed accordingly to

the number of cells. For the second evaluation the samples consisted of approximately 10 000, 25 000, 50 000 and 100 000 cells (n=3) 500 µL PBS was added to three vials with aliquots of 1 000 000 cells and distributed accordingly to the number of cells. In **Table 2** the redistribution of cells (1 000 000 cells) is prior to lysis given, while the distribution of lysate is given in **Table 3**.

Table 2. Each method had 550 000 cells to make three replicates of 10 000, 25 000, 50 000, and 100 000 cells, meaning that one replicate consists of 185 000 cells. One pellet of 1 000 000 cells was used to make one replicate for each method, where the order of distribution was changed for each replicate. The pellets for SP3 and ISD-Urea samples were lysed and distributed according to their cell number equivalent, while SPEED samples were distributed before lysing.

	Replica 1: 1 000 000 cells (500 µL PBS)		**Replica 2: 1 000 000 cells (500 µL PBS)**		**Replica 3: 1 000 000 cells (500 µL PBS)**	
Aliquot 1	SP3	185 000 cells (93 µL PBS) for each aliquot.	SPEED	185 000 cells (93 µL PBS) for each aliquot.	ISD-Urea	185 000 cells (93 µL PBS) for each aliquot.
Aliquot 2	SPEED		ISD-Urea		SP3	
Aliquot 3	ISD-Urea		SP3		SPEED	

For the samples from the second evaluation (n=3), the cells were distributed after lysis for the SP3 and ISD-Urea method, while in the SPEED method, due to the nature of the method, the cells were distributed before lysis.

Table 3. The distribution of cell lysate or PBS according to the cell number equivalent.

SP3		ISD-Urea		SPEED	
185 000 cells (150 µL lysis buffer)		185 000 (50 µL lysis buffer)		185 000 cells (93 µL PBS)	
# Cells	**Solution (µL)**	**# Cells**	**Solution (µL)**	**# Cells**	**Solution (µL)**
10 000	8	10 000	2.7	10 000	5
25 000	20.5	25 000	6.8	25 000	12.5
50 000	40.5	50 000	13.5	50 000	25
100 000	81	100 000	27	100 000	50

3.4.2 Mechanical lysis

Samples of 250 000 cells were lysed mechanically in either Protease inhibitor (1X) (n=2) or PBS (n=2). The samples were lysed with either ultrasonication or freeze-thaw cycles. Ultrasonication was performed by on and off cycles for a total of 4 min in an ultrasonic bath. Freeze-thaw was performed by adding 300 µL protease inhibitors (1X) or PBS and rapidly thawed to RT, then frozen at -80°C for 45 min. The cycle was repeated 3 times. The samples were centrifuged at 10 min/ 16000 x g/4 °C.

3.4.3 Urea-based In-solution-digestion

The method for cell lysis and in-solution-digestion for limited samples was kindly provided by Ph.D. Henriette Engen Berg, and been previously used (with some variations) in [120-122]. From here, this method is called ISD-Urea (note that urea is not used in the lysis buffer, but rather to keep the proteins denatured and in solution). Volumes of all reagents are described as the method is universal for all sample sizes with < 50 µg proteins.

Chemical cell lysis procedure

Samples of the initial screening (n=1 for each) were dissolved and lysed with 300 µL TX-100 buffer, while the sample consisting of 185 000 cells (n=3 each) for the second evaluation were lysed in 50 µL TX-100 buffer before being divided into samples of 10 000-100 000. The samples were placed in an ultrasonic bath for 30 s on, and 60 s off, for three cycles before being centrifuged for 20 min at 16 000 x g. The samples were placed on ice for 30 min. The supernatants were transferred to new clean vials. The supernatants of the samples with 185 000 cells were divided accordingly to the cell number (10 000, 25 000, 50 000, and 100 000 cells) before further sample preparation.

Protein precipitation

Neat acetone (-20 ºC) was added to the lysate in a 4:1 (v/v) ratio. After adding the acetone, the samples were homogenized with a vortex mixer and kept at – 80 ºC for 20-30 minutes followed by centrifugation at 14 000 x g for 10 minutes at 4 ºC. The supernatant was discarded and the protein pellets were left to dry in the fume hood for approximately 30 min, before storing the protein pellets at -80 ºC until use [120, 123].

Denaturation, reduction, and alkylation

The protein pellet was dissolved with 25 µL of 6M urea/100 mM ABC. For the reduction of proteins, 1.25 µL of 200 mM DTT was added to the samples, followed by incubation for 30 min at 30 ºC and 250 rpm. Alkylation was carried out by adding 3.75 µL of 200 mM IAM, before incubation for one hour at room temperature at 250 rpm. Due to IAM being light-sensitive, the samples were kept in the dark during incubation. To quench the alkylation, 5 µL of 200 mM DTT was added and incubated for 30 min at 30 ºC and 250 rpm.

Trypsination

Prior to the trypsination, 80 µL of 50 mM ABC was added. For the samples of 50 000-200 000 (n=1 of each) cells and 10 000-100 000 cells (n=3 of each) 1 µg and 0.5 µg trypsin, respectively, were added. The samples were incubated overnight at 37 ºC and 250 rpm. To terminate the protease activity, 5 µL FA (50%) was added.

3.4.4 Single-pot, solid-phase enhanced sample preparation

The procedure is based on the method by Hughes *et. al* [73], in combination with the lysis method from Ph.D. Berg [54].

Lysis

Samples of the initial screening (n=1 for each) were dissolved and lysed with 300 µL 1 % SDS buffer, while the sample consisting of 185 000 cells (n=3 each) for the second evaluation was lysed in 150 µL 1 % SDS buffer. The samples were heated in a Thermoshaker to 95°C for 10 min at 700 rpm. The samples were ultrasonicated in an ultrasonication bath for 30 s on, and 60 s off, for three cycles before centrifugation at 16 000 x g for 10 min at 4 °C. The supernatants were transferred to new vials, and 1 unit benzonase per 100 000 cells [73] was added before incubation for 30 min at 37°C at 700 rpm. The samples of 185 000 cells were divided accordingly to the cell number (10 000, 25 000, 50 000, and 100 000 cells) before further sample preparation.

Reduction and alkylation

IAM was added to a final concentration of 20 mM before incubation for 30 min at 700 rpm at RT. DTT was added to a final concentration of 5 mM and incubated for 15 min at 700 rpm at RT. The samples were incubated in the dark. The volumes of stock solutions are found in **Table 13 in the Appendix**.

Paramagnetic beads

The original stock solution from the manufacturer was 50 µg beads/µL solution. Because of limited stock solution, the stock solution for the hydrophilic beads was combined with the stock solution for the hydrophobic beads, giving a stock solution of 25 µg hydrophilic beads/ µL solution and 25 µg hydrophobic beads/ µL solution, in total a 50 µg beads/µL solution. The beads were washed by transferring 20 µL bead stock solution in an Eppendorf tube, before placing the vial on a magnetic rack for 5 min. The supernatant was discarded. The beads were reconstituted in 500 µL water, and placed on the magnetic rack for 5 min. The supernatant was discarded, and the beads were reconstituted in 50 µL water, providing a bead solution of 20 µg/µL.

The beads were transferred to the samples, providing a concentration of 0.5 µg beads/µL sample. The protein binding was induced by adding 50 µL EtOH. The sample tube was shaken carefully by hand. Incubation was performed for 5 min at 25°C and 1000 rpm, before placing the samples on the magnetic rack for 3-5 min until the beads had settled to the tube wall. The supernatant was discarded. Sample clean-up was performed by adding 180 µL of 80% EtOH and gently mixing the sample, before placing the samples on the magnetic rack. When the beads were settled on the tube wall after 3-5 min, the supernatant was discarded. The washing step was repeated in total three times.

Trypsination

The enzymatic digestion was promoted by adding 1 µg trypsin for the samples of the initial screening (n=1 for each), or 0.5 µg trypsin for the samples of the second evaluation (n=3 for each), in 100 mL 100 mM ABC. The beads not covered by the sample, were submerged in the solution with the pipette tip. Before pipette mixing the samples, the samples were sonicated for 30 s. The samples are incubated overnight at 37°C and 1000 rpm followed by centrifugation at 20 000 x g for 1 min at 4 °C. The samples were placed on the magnetic rack and the supernatants were transferred to new clean vials, and the digestion was stopped by adding TFA (in total 2%). As desalting was required to use the LC-MS at the core facility, sample clean-up was performed accordingly to **Section 3.5**.

3.4.5 Sample preparation of easy extraction and digestion

The procedure is based on the method by Doellinger *et. al* [53], with modifications of the reduction and alkylation by Ph.D. Berg. The volumes of solutions and reagents depend on the starting volume of the sample. For cell samples of <100 000 cells, a starting volume of 1 µL was estimated.

Denaturation, reduction, alkylation, and trypsination

The cells were lysed with TFA, in a 1:4 (sample/TFA) ratio. To neutralize the sample 2 M Trisbase was added in a volume ten times the volume of TFA. DTT was added to a final concentration of 10 mM, and incubated for 25 min at 56°C and 700 rpm. IAM was added to a final concentration of 20 mM, and incubated for 30 min at RT and 700 rpm. To quench the alkylation, 5 mM DTT was added. The samples were diluted 1:5 with H_2O. Digestion was

promoted by adding 1 µg or 0.5 µg of trypsin and incubating at 37°C at 700 rpm overnight. Before desalting, TFA was added to a final concentration of 2%. The volumes of stock solutions are found in **Table 14 in the Appendix**.

3.5 Sample cleanup

The method was kindly provided by Ph.D. Bernd Thiede at the core facility under the National network of Advanced Proteomics Infrastructure (NAPI).

Micro solid-phase extraction

Before desalting, the sample volume was reduced by placing the samples in a vacuum centrifuge for one hour at 30°C. When pipetting solutions and samples, the plunger was depressed and fixed in that position between aspirations to avoid introducing air into the C18 material. Conditioning of the Omix Tip was done by aspirating 30 µL (3 × 10 µL) conditioning solution and discarding the solvent. The same procedure was done for the equilibration. The sample was loaded by pipetting 10 µL at a time and slowly eluting the sample. To avoid contamination for future samples, 35 µL washing solution was pipetted to a clean vial and 30 µL (3 × 10 µL) washing solution was aspirated from this vial, and subsequently discarded. The sample was eluted with 10 µL elution solution in a new vial. The samples were completely dried in a vacuum centrifuge and reconstituted in 4 µL 0.1% FA water.

3.6 Organoid samples

3.6.1 Comparison between SP3 and SPEED method

Liver organoids derived from the H1 cell line were provided by Ph.D. Aleksandra Aizenshtadt from HTH. Three samples, defined as biological replicates, were pooled and washed within total of 200 µL ice-cold PBS before separating the pooled sample into two new clean vials. The samples were centrifuged at 400 x g and 4°C for 5 min. The pellets were not equal in size and the samples were again pooled and transferred to two vials. The samples were centrifuged at 16 000 x g and 4°C for 5 min so ensure that proteins would pellet together with the organoids in case the lysis had prematurely begun. The supernatant was discarded, and one sample was

prepared accordingly to the SP3 method (**Section 3.4.3**), the other sample accordingly to the SPEED method (**Section 3.4.4**).

3.6.2 SPEED-prepared samples

The samples consisted of liver organoids of different metabolic states were generated from the iPSC cell line (WTSIi013-A, Wellcome Trust Sanger Institute) by three differentiation protocols. Protocol 1, here called control, represents a standard differentiation protocol [124], and protocols 2 and 3 represent differentiation protocols [125] that lead to more metabolic mature liver organoids. Each sample consisted of 20-30 organoids (equivalent to 50 000 cells).

Each of the organoid samples was thawed on ice for approximately one hour and washed with 200 µL ice-cold PBS and suspended by gently rotating the vial. The samples were centrifuged at 400 x g and 4°C for 5 min, before discarding the supernatant. The rest of the sample preparation followed the protocol described in **Section 3.4.4,** and **Section 3.5** for sample clean-up with reconstitution of the sample in 10 µL 0.1 % FA.

3.7 Quantitative determination of total protein concentration in the lysate

3.7.1 Absorbance at 280 nm

The method is from [43] using the Nanodrop UV-VIS spectrophotometer and the A280 in the Nanodrop software. Before measuring the samples, the buffer should be measured at 280 nm to evaluate if it is suitable for this method. Aliquots of 2 µL solution or samples were used for all the measurements. Between all measurements, a lint-free wipe and water were used to clean the pedestal. First, water was applied to the sample pedestal and a blank measurement was performed to set the baseline by using the *Blank* function. Then the buffer was measured. If the spectrum varied more than 0.04 from the baseline at 0.280 nm the buffer was considered not suitable for this method and other quantification methods were considered. If suitable, a new aliquot of buffer was applied, and a *Blank* was performed before measuring the samples.

3.7.2 Bicinchoninic acid assay

If the buffer was not suitable for using A280, the protein content was quantified using the Pierce BCA Protein Assay Kit. For measuring the UV-absorbance a Nanodrop 2000 spectrophotometer from Thermo Fisher Scientific was employed. The procedure is adapted from the producer [126] and modified by Ph.D. candidate Maria Schüller [120] to be compatible with 10 µL sample amounts.

Preparation of BCA assay reagents

The working reconstitution buffer (WRB) was prepared by mixing 600 µL reconstitution buffer and 600 µL water. The WRB (1000 µL) was added to an ampule of the compatibility reagent and vortexed for 30 s at high speed. The compatibility reagent solution was stored at 4 ºC until use the same day. The excess solution was aliquoted to 200 µL per vial and stored at – 80 ºC immediately. This way of storage was confirmed suitable by the technical support of Thermo Fisher [127]. The BCA working reagent (WR) was made on the day of the experiment by mixing BCA reagent A and BCA reagent B in a ratio of 50/1 (v/v). The WR developed a green color upon mixing.

Preparation of BSA calibration solution

The bovine serum albumin (BSA) stock solution containing 2 000 µg/mL BSA provided by the assay kit was diluted with PBS in a concentration range of 125-200 µg/mL to make the calibration solutions. The composition of the calibration solutions is found in **Table 15 in the Appendix**.

Protein quantification with Nanodrop 2000 spectrophotometer

For quantification, 10 µL of calibration or sample solution was used. If the sample contained IAM and DTT, 10 µL of CRS was added and the sample was vortexed at low speed for 20 s and incubated at 37 ºC for 15 min in a Thermoshaker at 250 rpm. After incubation, 100 µL of WR was added and vortexed well. The tubes were incubated for 30 min at 37 ºC with 250 rpm. The tubes were then cooled at RT for 5-10 min before measuring the Cu-BCA complex at 562 nm with the Nanodrop. Prior to the measurements, the pedestal on the Nanodrop was cleaned with 0.5 M HCl, followed by type 1 water and reconditioning with the reconditioning kit by Thermo Fisher Scientific. For the measurements, 2 µL of the BCA assay prepared

calibration or sample solution was pipetted onto the pedestal. The standard curve was established with two measurements per calibration solution, here defined as technical replicates, using increasing concentrations of BSA. After establishing the calibration curve, a blank sample consisting of lysis buffer without proteins was measured, following the samples with an increasing number of cells. For each of the samples, three technical replicates were measured. To correct the background interference, the average absorbance of the blank sample was subtracted from the average absorbance of the sample.

3.7.3 Turbidity measurement for SPEED-prepared samples

Turbidity measurement was performed with Nanodrop by pipetting 2 µL of sample and measuring absorbance at 360 nm using the UV-Vis method in the Nanodrop software. The baseline was established by 2 µL of 1:10 TFA/Trisbase. The protein concentration was estimated by multiplying the sample absorbance with a coefficient of 0.79 µg/µL (**Section 7.1.2 in the Appendix)** [53, 128].

3.8 nanoLC-timsTOF setup

The samples were analyzed by LC-MS using a timsTOF Pro by Bruker (Bremen, Germany) which was coupled online to a nanoElute nanoflow LC system via a CaptiveSpray emitter via nanoelectrospray ion source, using a CaptiveSpray emitter with ID of 20 µm. The MP A contained water with 0.1% FA, and ACN with 0.1% FA was used as MP B. Injection volume was 2 µL for all the samples. Parameters for the MS are found in **Section 7.2.6 in the Appendix**. HeLa digest (100 ng) was used to evaluate the performance of the system.

Packed nanoLC column

Packed C18 nanoLC column was used for the method optimization. The peptides were separated on a Pepsep C18 column (25 cm x 75 µm, 1.5 µm) by Bruker with a main gradient of 2-35% MP B over 60 min with a flowrate of 300 nL/min. The column oven was set on 50°C.

µPAC

To make the µPAC column compatible with the nanoElute nanoflow LC system, additional two connectors (Valco reducing union 1/16'' to 1/32'', and 1/16'') and a 30 cm x 20 µm capillary

tube were added to the system (**Figure 20A and B**). HeLa Digest standard (100 µg on column) was injected with main gradient of 2-35% MP B over 120 min, 180 min and 240 min, where the 180 min main gradient was chosen for the organoids sample. The peptides from the organoid samples (4 technical replicates, (N=4)) were separated on a 200 cm C18 µPAC column by PharmaFluidics (now Thermo Fisher), using a main gradient of 2-35% MP B over 180 min at a flow rate of 300 nL/min, with pressure restriction of 340 bar (**Figure 20C**).

Data analysis

Data analysis for evaluation of the sample preparation method was performed by using the Scaffold software, with the parameters set with protein threshold of 99.9 %, minimum 2 signature peptides and peptide threshold of 95 %.

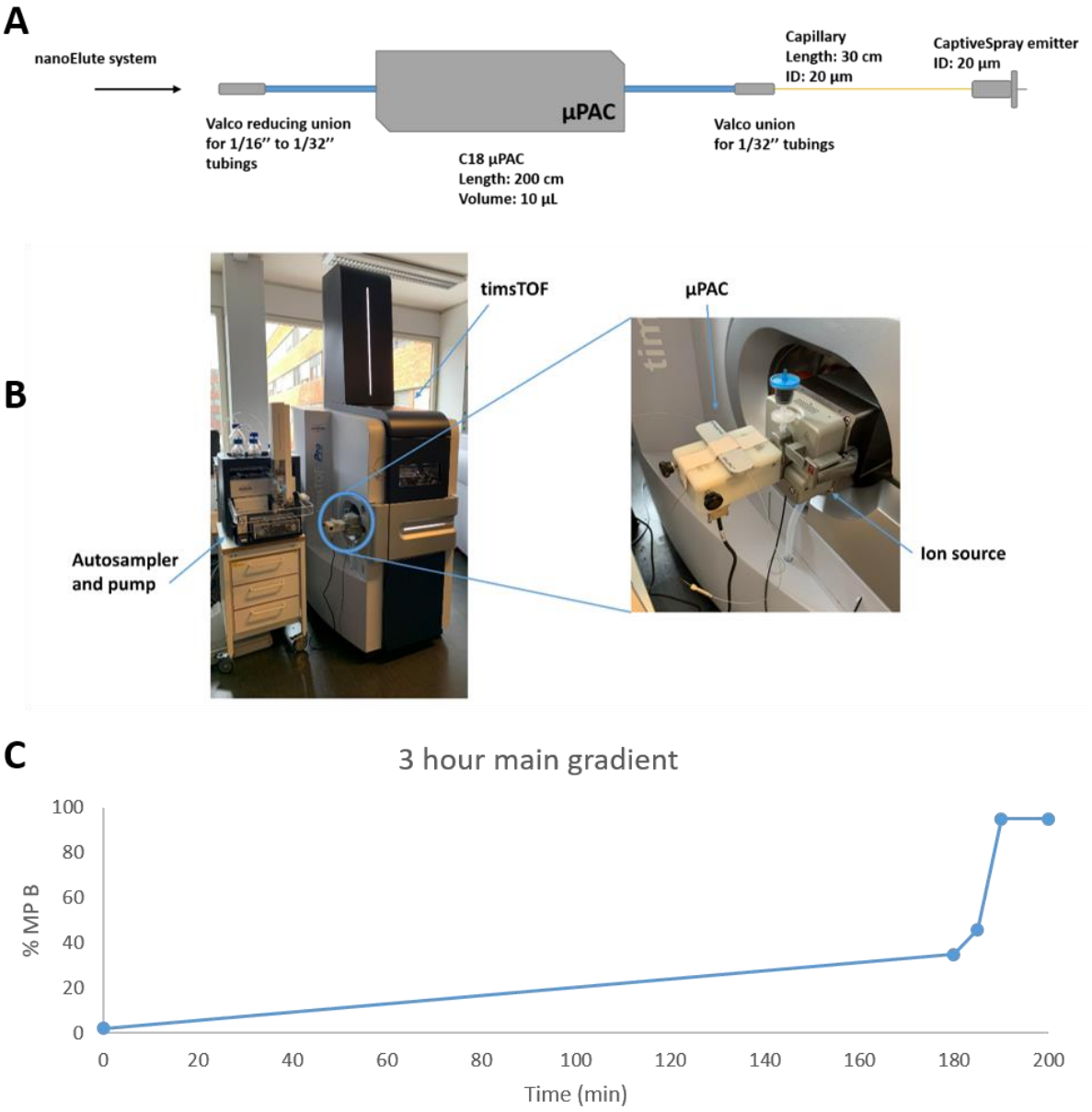

Figure 20. A) Schematic figure of the µPAC column with the fittings necessary to with the nanoElute system. B) The LC-MS system is shown on the left. On the right the µPAC column with the ion source are shown.C) The 180 min gradient used for the sample. The 120 min and 240 min gradient have the same gradient program were only the length of the main gradient differed.

4 Results and discussion

Because sample preparation is considered the bottleneck for proteomics on limited samples, the aim of this study was to establish a sample preparation method for analyzing limited sample amounts, e.g. organoids and subcellular populations. An ideal sample preparation for limited samples should consist of a lysing method with high protein yield, and few sample transfer steps to minimize the sample loss. As the sample preparation methods are non-automated, meaning that the robustness of the method is highly dependent on the skills of the operator, factors such as handling time and easy-to-follow methods should also be of considerations.

Narrow and/or long column formats provides increased sensitivity and efficiency compared to their conventional LC counterpart. Because limited samples are of high complexity, different nanoLC formats were used, together with an HR-MS. Initially, the samples were planned to be analyzed in-house with Inifinity nanoflow LC by Agilent or EASY nLC system by Thermo coupled with Q-Exactive MS. For the initial nanoLC-ESI-MS platform see **Section 7.3 in the Appendix**. Due to technical problems with the LC-system and the Q-Exactive MS, with prolonged repair time, it was decided that the samples were to be analyzed at the NAPI consortium with an established nanoLC-ESI-MS platform for proteomics. This platform minimized possible variations due to a non-optimized analytical system, making the comparison between different sample preparation methods more reliable as the variations in the number of protein identifications are due to the sample preparation rather than the performance of the LC-MS platform. Ph.D. Thiede and Ph.D. Garrastacho at the facility analyzed the samples and were responsible for the maintenance of the LC-MS. The platform using µPAC discussed in **Section 4.6** was coupled to the platform by the author, with assistance from engineer Inge Mikalsen and Ph.D. Garrastacho. The number of protein identifications and unique proteins were found using the Scaffold software. The LFQ and statistical processing were performed by Ph.D. Tuula Nyman and Ph.D. Thiede at the core facility.

The results and discussion section is divided into three main parts: protein yield (**Section 4.1-4.3**), evaluation of sample preparation methods (**Section 4.4-4.5**), and application of the optimized sample preparation on liver organoids (**Section 4.6**). The limitations of this work are discussed in **Section 4.7**.

4.1 Estimation of protein yield with mechanical lysis

Sample losses for limited samples are far greater compared to bulk samples, with sample clean-up considered a significant contributor to sample loss. Wu *et. al* found that there was 15% sample loss with 50 µg protein sample, while for 2 µg samples, the loss increased to 89% [58, 59]. The use of mechanical lysis eliminates the need for sample clean-up of MS-incompatible detergents, and sample preparation for SCP often employs this lysis method [129, 130]. Hence, the protein extraction efficiency of mechanical lysis described in **Section 3.4.2** was investigated by measuring protein concentration. To prevent protein degradation, protease inhibitor cocktail diluted with PBS was added. The solution had an absorbance of 0.02 at 280 nm and was therefore suitable with A280. Because it was hypothesized that mechanical lysis may have a lower protein yield due to the lack of detergents to lyse membrane proteins, it was decided to use the BCA assay instead, as it is more sensitive and selective. The mechanical lysis with 3 freeze-thaw cycles and sonication showed that by adding protease inhibitors the protein yield increased (**Table 4**) compared to the samples without protease inhibitors. The lack of proteins detected in the samples without protease inhibitors can be due to the spontaneous degradation of proteins.

Mechanical lysis on organoids for proteomics studies is less explored. Cells in organoids are less exposed because of the 3D assembly. Mechanical lysis may cause incomplete lysis of organoids [54], as detergents often are necessary to extract membrane proteins, break protein-lipid interactions as well as achieve protein solubilization [131, 132]. The sonication was performed in an ultrasonic water bath. Ideally, the sonication should be performed using a sonication probe since this device can concentrate the sonication waves better than the sonication bath [37]. Nevertheless, the low protein extraction yield (0.9 µg for 250 000 cells) using sonication implies that this is not a suitable method for lysing organoids. Although the protein extraction yield for the freeze-thaw method (13 µg for 250 000 cells) was higher than sonication, it is still not in the expected range (50 µg for 250 000 cells [59]). Hence, the mechanical lysing methods were not further investigated.

Table 4. The protein yield by mechanical lysis with and without protease inhibitor (PI). The samples consisted of approximately 250 000 cells (n=3).

# Cells	Lysing method	With PI (Y/N)	Measured conc. (µg/mL)	Calculated protein amount (µg)
0	Freeze-thaw	N	Below detection	-
0	Sonication	N	Below detection	-
250 000	Freeze-thaw	N	Below detection	-
250 000	Sonication	N	Below detection	-
0	Freeze-thaw	Y	Below detection	-
0	Sonication	Y	Below detection	-
250 000	Freeze-thaw	Y	43	13
250 000	Sonication	Y	3.0	0.9

To sum up: The samples with protease inhibitor had a protein yield, where freeze-thawing provided the highest protein yield of 13 µg. However, due to low protein yield and possible ineffective lysis of organoids, was it decided that mechanical lysis was not a suitable method for the lysis of organoids.

4.2 Estimation of protein concentration in cells lysed with ISD-Urea method

The protein concentration in lysate prepared with the ISD-Urea method described in **Section 3.4.3** was estimated using the BCA Assay.

Quantification of protein yield before enzymatic digestion is important to ensure high digestion efficiency, and to prevent autolysis and ion suppression by adding the right amount of protease. Quantification allows for relative quantification of samples by using isobaric tags, which

enables several samples to be analyzed in one run (multiplexing). This is especially important in DDA where LFQ will be affected by the run-to-run variations, which may cause different top N precursor ions at the same retention time to be selected for fragmentation [133]. In addition, multiplexing saves instrument time and MP.

4.2.1 Quantification of proteins using BCA Assay

To evaluate the protein concentration yield using detergent-based cell lysis method, the BCA assay was used on samples with different cell amounts. Samples consisting of 50 000, 100 000, 150 000 and 200 000 HepG2 cells lysed with 300 µL TX-100 yielded 77 µg/mL (relative standard deviation (RSD) 2 %), 38 µg/mL (RSD 1 %), 31 µg/mL (RSD 3%) and 31 µg/mL (RSD 1 %) (**Table 5**) of proteins in the cell lysate, respectively. It was expected that the number of cells would be proportional to protein concentration. The contrary results may indicate that the lysing efficiency is increased when there is a higher volume of lysis buffer. In a manual for a bottom-up proteomics kit using SDS-buffer from Thermo Fisher, it states that cell lysis is promoted by adding a five-pellet volume of lysis buffer [134]. The 300 µL lysis buffer added would in this case be more than sufficient to lyse the cells with the same efficiency, as this was more than five times the volume of the pellet in all the samples. If assuming that the vials are marked in the wrong order, it still would not explain the differences in protein concentration, e.g. two of the samples have the same protein concentration of 31 µg/mL, although the samples should differ with 50 000 cells. This indicates that the method for distributing the cells may not be ideal. This can be because cells tend to sediment and aggregate fast. When distributing the cells, a less volume of PBS should be used to minimize the heterogeneity of the cells in the sample. Technical replicates, N, are measurements of the same sample, and will only define the accuracy and precision of the instrument. In addition to the three technical replicates, there should have been a minimum of three sample replicates to determine the accuracy of the quantification.

Table 5. The protein yield after cell lysis. The measured concentration is given by the Nanodrop software, and the calculated concentration is the measured concentration of the blank sample subtracted from the measured concentration of the sample. (N=3)

# Cells	Measured conc. (µg/mL)	Calculated conc. (µg/mL)	Amount (µg)	RSD (%)
0	455	0	0	4
50 000	538	77	22	2
100 000	493	38	11	1
150 000	486	31	9	3
200 000	486	31	9	1

In addition to the uncertainty regarding the number of cells, the protein quanfication using BCA and Nanodrop has its limitations. It was observed that the sample droplet ("cuvette") flattened out and could "break" during measurements. The droplet tends to evaporate faster when it is flattened out, causing variable results. In addition, the b contribution in Beer's law (**Equation 1**) decreases, yielding a higher concentration of the analyte. Hence, to avoid this problem, the pedestal was properly cleaned and reconditioned before and after each set of measurements as the hydrophobic surface of the pedestal is prone to compromise when working with detergents.

Matrix matching is important, due to possible interferences that may affect the analysis. Hence, the calibration solutions should be diluted with the same buffer as the sample. The BSA standard (2000 µg/mL) provided in the BCA assay kit was of the same concentration as the highest concentrated calibration solution. Ideally, the volume-to-volume-ratio between the standard and the diluent should be high so that the matrix effect would be of the same extent in every calibration solution. The provided standard was diluted with PBS, hence all the calibration solutions were diluted in this instead of TX100.

To evaluate the time-dependency of the BCA reaction, calibration solutions were measured at different time points (**Table 6**). Standards ranging from 250-1500 µg/mL were measured after 45-135 min incubation in RT. The relative error of the measurements was 4-37%. The reaction is time-dependent, hypothesized due to a complex mechanism involving secondary interactions

and modulated conformations of proteins [135]. Hence, the time-dependent reaction compromises the precision and accuracy of the quantification. It is recommended to measure the calibration solutions and sample after 5-10 min in RT [126]. Because the samples are measured manually, one by one, it is difficult to reproduce the incubation time for all the samples as it takes about 2 min to measure three technical replicates for each sample, hence the need for estimating changes over time (e.g. 45 min). If more samples were to be measured, a plate reader using 2 µL cuvette plates would be a better choice in term of reproducibility. Hence, all the samples were measures within 15 min to minimize this contribution to the concentration.

Table 6. The protein concentration in cell lysate was performed using BCA assay and measured at different incubation times (N=3).

Theoretical conc. BSA (µg/mL)	Conc. blank (µg/mL)	Measured conc. BSA (µg/mL)	Relative error (%)	Incubation time RT (min)
250	0	280	10	45
750		890	19	45
1500		1570	4	45
750		950	28	75
750		1030	37	135

It was observed that the blank samples (containing the same lysis buffer as the sample) developed a purple color similar to the samples. This indicates that there are some interferences in the buffer with the assay, causing a reduction of the cobber. To take into account the matrix effects, the concentration of the blank was subtracted for all the samples. After personal communication with the technical application team at Thermo Fisher, it was concluded that even with the subtraction of the blank sample, it is hard to determine if the measurements are reliable as there should not be a significant visible color formation of the blank sample [136].

A typical protein yield for one mammalian cell is 200 pg [59], which is roughly equal to 200 µg proteins for 1 000 000 cells depending on the cell line. This is in the same range as the results

shown in **Table 5**, although the cell distribution is not optimal. Due to possible interferences in the lysis buffer, and the method was originally aiming for cell samples (towards SCP), it was decided to disregard using spectrophotometric methods to estimate the protein concentration in the detergent-based cell lysate (SP3 samples and the rest of ISD-Urea samples). The amount of trypsin was estimated by the number of cells. The group of Thiede uses, as a rule of thumb, 5 µg trypsin for 1 000 000 cells and this was used as a reference point for the amount of trypsin added in the samples instead of Nanodrop and absorbance measurement with BCA.

To sum up: *Using BCA Assay with Nanodrop provided a method for quantification of protein content in detergent-based cell lysate. The protein quantification indicated that the method for distributing the cells may not be ideal. Due to possible interferences between the lysis buffer and BCA reagents, it was decided to disregard protein quantification and estimate the amount of reagents for bottom-up proteomics based on the theoretical number of cells in the samples.*

4.3 Estimation of protein concentration in SPEED-prepared cell samples

The estimation of protein concentration in SPEED-prepared samples was evaluated using BCA assay, A280, and turbidity measurement.

4.3.1 BCA Assay for protein quantification on SPEED-prepared samples

The BCA assay has previously been used on SPEED-prepared organoid samples to estimate the protein concentration, with 15-18 times dilution of the organoid samples [54]. However, the experiments were not reproducible on cell samples due to the absorption of the blank samples was equal to or higher than that of the samples, even when samples were diluted 100-fold to limit the possible interference. Similar results were obtained with organoid samples, with the concentration of proteins being higher in the lysis buffer compared to the samples (**Table 7**). When measuring the protein concentration of 250 000 cells, the concentration was above the detection limit. The high absorbance of the blank is probably due to the salt concentration of the TFA-Trisbase used for cell lysing, as it will interfere with the BCA assay [128]. The dilution of samples in previous experiments [54] to decrease the ionic strength was not reproducible in these experiments.

Table 7. The absorbance and protein concentration of cell and organoid samples with different dilution factors. The protein concentration was calculated by Nanodrop Software.

Sample	Dilution factor	Reduced? (Y/N)	Frozen (Y/N)	Conc. blank (µg/mL)	Conc. range samples (µg/mL)
50 000-200 000 cells	10	Y	Y	71	59-85
250 000 cells	0	Y	Y	-	Out of range (>2000 µg/mL)
20-30 organoids	100	N	Y	11	9-11
20-30 organoids	18	Y	Y	107	87-102

The BCA assay kit was provided with a compatibility reagent solution as reducing agents reduce Cu^{2+} to Cu^{+} and interfere with the assay. Because a stronger color formation was observed when the compatibility reagent solution was added, was it investigated if this would affect the quantification. Since all the organoid and cell samples in this work were frozen prior to quantification with BCA assay, was it investigated if freeze-thawing, in addition to the reduction of proteins, would impact the quantification. To examine this, 10 µg BSA (equivalent to protein content in 50 000 cells) was diluted 18-fold according to [54] and 1 µL of this sample was dissolved in 1:10 TFA/Trisbase. The sample that was not frozen before quantification gave a protein yield of 12 µg of protein, which was almost the targeted value (+2 µg) (**Table 8**). However, when the experiment was repeated the same result could not be obtained as the blank again had higher absorption compared to the sample, indicating that the accuracy, precision, and repeatability of this method are poor, hence SPEED digestion is not compatible with BCA assay.

Table 8. Measured protein concentration of BSA using BCA assay in 10 µg protein samples, with combinations of reduction and freezing.

Theoretical amount BSA (µg)	Dilution factor	Reduced? (Y/N)	Frozen? (Y/N)	Measured conc. blank (µg/mL)	Measured conc. sample (µg/mL)	Calc. amount BSA (µg)
10	18	Y	N	111	103	-
		N	Y	80	81	0.8
		Y	Y	180	124	-
		N	N	33	48	12
10*	18*	N*	N*	39*	38*	-

* Measurement performed on another day.

Bradford assay is another colorimetric assay for protein quantification, which was not evaluated because it was initially thought that 150 µL of the sample was necessary for quantification, compromising the already limited sample volume [39, 137]. However, there is a method with the Nanodrop in mind using 10 µL of sample volume compatible with the Bradford assay [45]. Compared with BCA, Bradford assay is less protein selective as the quantification depends on the amount of basic and aromatic residues to form the colored complex [138] and the assay is not able to detect small proteins (<3 kDa) [45], while the colored complex formation by the BCA assay depends on the amount of peptide bonds. However, BCA assay is incompatible with buffers containing cobber-reducing agents or if the salt concentration is too high. Alkaline buffers can cause interference with the Bradford assay and it is recommended to have a Trisbase concentration of <2 M [138, 139], thus the SPEED method with limited sample volume may be compatible with the Bradford assay. Due to time restrictions, this was not evaluated.

Although the BCA assay for protein quantification before enzymatic digestion was not successful, the digested cell samples were analyzed by LC-MS. The amount of trypsin was estimated based on the cell count in the sample. The LC-MS analysis showed that there were proteins in the cell samples, and the relatively low absorbance of the samples compared to control samples were not due to absents of proteins (**Section 4.4** and **4.6**).

To sum up: The BCA assay was incompatible with the SPEED method. Other colorimetric methods such as Bradford assay should be evaluated.

4.3.2 Turbidity measurement

Since the TFA and Trisbase are incompatible with the BCA assay, the original method used tryptophan fluorescence and turbidity measurements (**Section 7.1.1** and **7.1.2 in the Appendix**, respectively) as quantification methods. The author of the SPEED method preferred turbidity over tryptophan fluorescence since it, according to the author [53], provided a simple and fast measurement, where the sample aliquot used to quantify could be collected and used further. A standard to determine the coefficient for the instrument was not available during this work.

After personal communication with the author of the SPEED method [128], it was suggested to measure the samples at 360 nm using a UV-VIS spectrometer and use 0.79 µg/µL as the coefficient. Even though the coefficient is instrument-dependent, in their case the coefficient was 0.79 µg/µL for multiple different UV-VIS instruments used in their study [53, 140]. When using the Nanodrop the samples absorbed at 360 nm showed less protein yield than expected. The sample size of 50 000 cells should theoretically contain approximately 10 µg protein, and calculation using 0.79 µg/µL as a coefficient gave approximately 0.32 µg protein (**Table 9**). For 200 000 cells the number was approximately four times higher, yielding 1.35 µg protein. This indicates a linearity, but the coefficient is not 0.79 µg/µL for Nanodrop and it is therefore necessary with a standard containing suspension. The different coefficients between conventional UV-VIS spectrophotometer and Nanodrop can be due to the light path of Nanodrop being less than the light path of the cuvettes used in the original method.

Table 9. The absorbance at 360 nm and calculated protein amount (µg) in the lysate.

# Cells	Absorbance (360 nm)	Calculated protein amount (µg)
50 000	0.009	0.32
200 000	0.054	1.35

Tryptophan fluorescence, on the other hand, is a suitable quantification method with the SPEED method. The original method used tryptophan standard to establish the calibration curve [53]. However, there are protein labeling kits on the market, e.g. Fluoreporter^{TM} FITC protein labeling kit by Thermo Fisher Scientific, which labels the proteins with a fluorescence tag [141]. The limitation of such kits is that the proteins have to be >30 kDa to be detected, making this type of quantification not absolute as small proteins are excluded. Also, an instrument able to detect fluorescence was not available during this work. However, measuring the protein concentrations using fluorescence methods has to be evaluated for the SPEED method.

To sum up: The turbidity measurements showed linearity, while the results were lower than expected, indicating that the coefficient used to calculate the protein amount was not right. To use this method with Nanodrop, the instrument-dependent coefficient needs to be determined. Tryptophan fluorescence measurement may be a suitable method for the SPEED method. However, these methods were not examined as there were no fluorescence instruments available.

4.3.3 A280 for protein quantification of SPEED lysed cells

The A280 is frequently used as a quantification method for less complex samples. However, it is not compatible with most detergents because they absorb light at 280 nm [36], and was not considered a quantification method for the detergent-based methods. Due to the difficulties with the before-mentioned methods for protein quantification of the SPEED method, the A280 was evaluated. The evaluation of the buffer compatibility with the A280 requires that the absorbance of the buffer does not deviate more than 0.04 in absorbance from the baseline at 280 nm [43]. A stable and repeatable baseline was not achieved with TFA-Trisbase. This may be due to the high salt concentration, or that Trisbase absorbs UV light at 280 nm.

Since TFA does not absorb at 280 nm and is a suitable buffer to use with A280, the measurement of protein concentration using A280 after adding TFA was considered and evaluated. However, TFA has a low surface tension [142] and will not form a droplet on the Nanodrop pedestal used without added water. Another issue is the volatility of TFA causing the droplet to evaporate, leaving the droplet with a higher protein concentration. A sample containing 200 000 cells diluted 20 times yielded 75 µg (625 µg/mL) of proteins. Theoretically, 200 000 cells would yield approximately 40 µg of proteins. It is difficult to evaluate if the sample contained more than 200 000 cells or if the high protein amount was a result of droplet evaporation. The author

also found the method to be difficult as little sample was left after quantification, thus decreasing the already small sample size for downstream sample preparation.

To sum up: A method for quantification of proteins in SPEED-lysate was not found in this work. The use of Bradford assay, turbidity measurement, and tryptophan fluorescence should be further investigated. As previous studies have successfully implemented SPEED for limited cell samples, it was decided to estimate the protein concentration by the number of cells and continue the sample preparation.

4.4 Evaluation of sample preparation methods with limited cell samples based on the number of protein identifications.

IGD is viewed as a cornerstone in bottom-up proteomics, providing separation of the sample prior to LC-MS analysis, increasing the dynamic range of the analysis, efficient sample clean-up, and thus providing in-depth protein analysis [47]. Because 20 µg of proteins equal approximately 100 000 cells and the limited samples are of less, IGD was discarded as a method for this thesis.

The aim of the initial screening was to evaluate if the LC-MS system was sensitive enough to be able to detect proteins with the SPEED and ISD-Urea as sample preparation methods. In addition, SP3 was evaluated as Ph.D. Berg concluded that using a lower sample volume (<10 µL) should be investigated between SP3 and SPEED [54], as SP3 provided a method with minimal sample loss by limiting surface contact, and eliminating the need for SPE sample-clean up before LC-MS analysis. Due to practical reasons, one sample replicate (n=1) per sample size was prepared and analyzed. The samples were evaluated by the average number of protein identifications from the *Statistical* function in the Scaffold software (**Table 16-Table 18 in the Appendix**). With samples containing more than one replicate, the number of unique proteins among the replicates could be found using the Venn-diagrams in the *Quantify* function (**Figure 35 in the Appendix**).

Initial screening

The samples in the initial screening consisted of 50 000, 100 000, 150 000, and 200 000 cells. Only the proteins in ISD-Urea treated samples were quantified before enzymatic digestion using BCA assay. For SPEED and SP3 there are therefore no other indications of the start concentration of proteins other than a rough estimate based on the approximate number of cells. The initial screening revealed protein identifications of 197 (SP3 with 50 000 cells) to 3122 (SPEED with 200 000 cells) (**Figure 21** and **Table 10**).

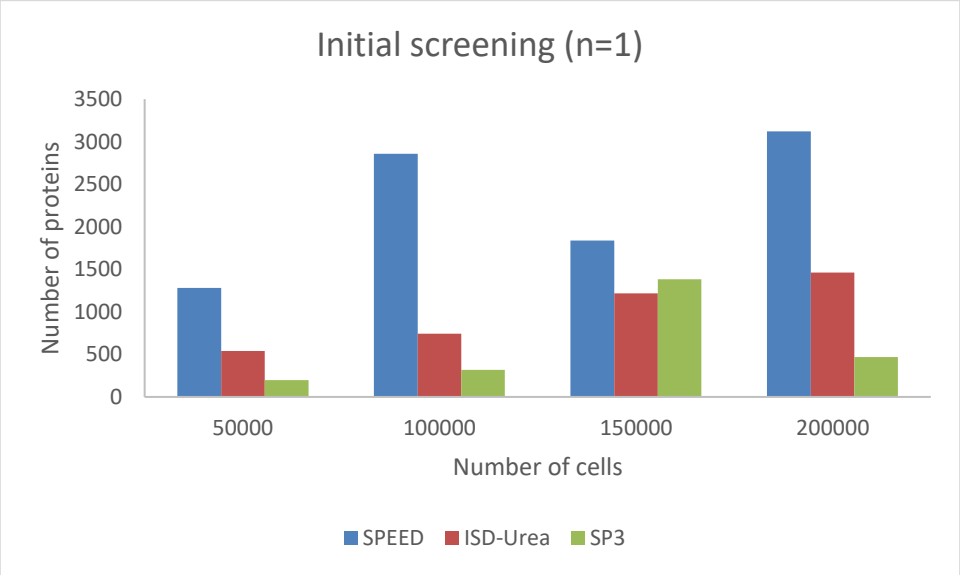

Figure 21. Number of protein identifications based on number of cells in the sample. (n=1)

To sum up: Proteins were identified with all the sample preparation methods, where the SPEED-prepared samples had the highest number of protein identification among all the sample sizes. It was decided to continue with all the sample preparation methods for further evaluation.

Second evaluation

Since proteins were detected in all the samples, was it decided to downscale the sample sizes, repeat samples containing 50 000 and 100 000 cells, and increase to three replicates (n=3) for all the samples. This experiment is here defined as the second evaluation (**Figure 22**). For the ISD-urea and SP3 samples, the cell pellets were chemically lysed with TX100 and SDS, respectively, and then different volumes of lysate were distributed accordingly to the sample size estimate, while in the initial screening, the cells were distributed before lysing. By dividing

the lysate, it was hypothesized that the distribution of cells should be more homogenous. Due to the nature of the SPEED method, the cells were distributed as previously described before lysing the cells. The samples from the second evaluation showed the SPEED-prepared samples to have the highest number of protein identifications among all the sample sizes, with an average ranging from 1684 proteins (10 000 cells) to 3504 proteins (100 000 cells). For the ISD-Urea-prepared samples, the protein identifications averaged at 266-627 protein identifications, while for the SP3-prepared samples averaged at 961-1490 protein identifications (**Table 10).** Raw data and representative chromatograms are found in **Section 7.4 in the Appendix**.

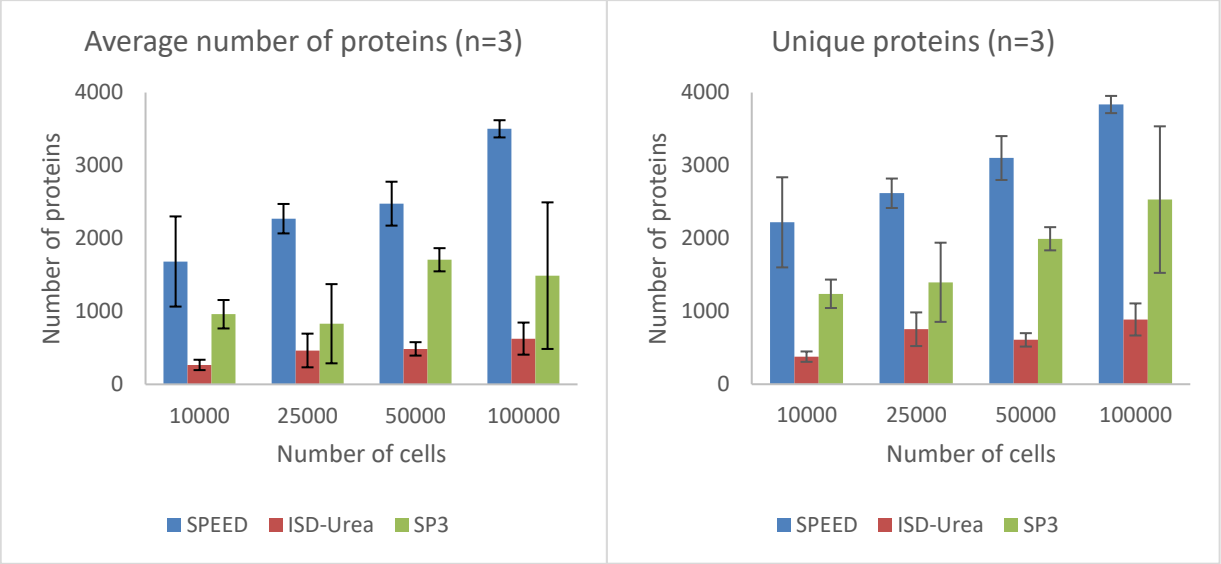

Figure 22. Left:The average number of protein identifications in the sample based on cell amount. Right: The number of unique proteins found in the samples. Standard deviations are shown as error bars. (n=3)

Table 10. The average number of protein identifications in the preliminary experiments using HepG2 cells.

#cells(10^3) Method	10	25	50	50*	100	100*	150*	200*
ISD-Urea	266	464	486	540	627	743	1218	1461
SPEED	1684	2271	2477	1283	3503	2858	1839	3122
SP3	961	831	1708	197	1490	319	1383	468

* n=1, initial screening

ISD-Urea-prepared samples had the overall poorest performance

The conventional ISD sample preparation method, ISD-Urea, is the most time-consuming of the three ISD methods. While SP3 and SPEED can be processed within one vial, the ISD-Urea requires an additional sample transfer step after protein precipitation, leading to potential sample loss. ISD-Urea is criticized for its low sample recovery [58]. In the initial screening, ISD-Urea had a better performance compared to SP3 regarding the number of protein identification, except for the sample with 150 000 cells. In the second evaluation, the ISD-Urea had the overall poorest performance and the chromatograms were of lower intensity (**Figure 34 in the Appendix**). The ISD-Urea-prepared samples were the last samples in the second evaluation to be analyzed with LC-MS. HeLa digest (100 ng) was injected before and after this run to evaluate the performance of the system since it with 39 samples took 2-3 days to run. The HeLa yielded 2907 proteins at the start of the run, while at the end of the run 2321 proteins were detected, meaning that the performance of the system was poorer at the end of the run compared to the start, which may have contributed to the lower protein yield for the ISD-Urea-prepared samples.

Despite the samples may have been affected by the performance of LC-MS, the low protein yield was considered to be significant. The last sample to run prior to the ISD-Urea samples yielded approximately 3000 more proteins compared to the ISD-Urea sample of the same size (**Table 10**). In addition to poor protein yield, the method was the most time-consuming among the sample preparation method (**Figure 23**). Processing one sample took in a total of about four hours before enzymatic digestion and sample clean-up. The additional steps and long processing time makes this method highly dependent on the operator's skills and impair the reproducibility of the method.

Hence, ISD-Urea was not considered the optimal method for limited cell and organoid samples.

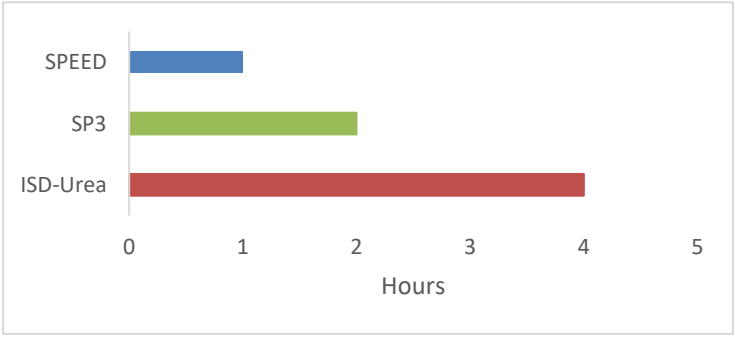

Figure 23. Handling time for one sample prior to the enzymatic digestion and clean-up. SPEED had the shortest time of one hour, while SP3 had two hours and ISD-Urea four hours.

The SP3-prepared samples had the second to the highest number of protein identifications

The lowest number of protein identification among all the samples and ISD methods was found using SP3 for the sample with 50 000 cells (n=1) with 197 proteins in the initial screening. SP3 had the lowest number of protein identifications in all the sample sizes except for samples with 150 000 cells. In the second evaluation, SP3-prepared samples had the second highest number of protein identifications among all the sample sizes. One replicate with 25 000 cells and one replicate with 100 000 cells had substantially less number protein identifications, with 292 and 496 protein identifications (**Table 16 in the Appendix**), respectively, compromising the average number of protein identifications. This indicates that the sample clean-up either by the beads or the SPE had a significant protein loss. This was reflected in the chromatograms of these low abundant samples (**Figure 31 in the Appendix**), as they showed fewer peaks compared to the other chromatograms of the SP3 and SPEED-prepared samples. One of the biggest advantages of the SP3 method is the sample clean-up prior to the enzymatic digestion which eliminates the need for sample clean-up with e.g. SPE before LC-MS analysis which is associated with protein loss. However, SPE using Omix-tips was performed for all the samples, including the SP3-prepared ones, as Ph.D. Garrastacho and Ph.D. Thiede at the core facility has experienced on several occasions beads clogging the column. With two clean-up steps during the sample preparation, significant sample losses may have taken place.

The method was by mishap not prepared as described in [73], where the recommended bead amount should be 10:1 beads/protein. The concentration used was 0.5 µg beads/µL sample, which is the lowest recommended bead concentration. An ideal sample preparation should be easy to follow with minimal steps, and the author of this work found that not to be the case using this method.

Nevertheless, it was decided to evaluate SP3 for organoid samples as it provided higher protein yield compared to ISD-Urea, and the lysate is easy to quantify before enzymatic digestion using BCA assay.

The SPEED-prepared samples had the highest number of protein identifications

The samples prepared with SPEED showed the highest protein identification in all sample sizes, the highest found in samples consisting of 100 000 cells (n=3) yielding 3503 proteins (**Table**

10). In addition to a high protein yield, SPEED provided the samples with the highest purity, expressed as a high %ID. SPEED had a %ID of 58-65% compared to 26-51% and 7-17% for the SP3 and ISD-Urea, respectively (**Table 16-Table 18 in the Appendix**). The high %ID may be due to the detergent-free lysis since detergents ionize well and can cause ion suppression of the peptide ions [31]. The SPEED sample had the shortest handling time of one hour before enzymatic digestion, making this an easy method to follow for the operator.

A challenge with proteomics on limited samples is the need for minimum sample volume during sample preparation to reduce surface contact and avoid loss of hydrophobic proteins or peptides. Using SPEED, the sample is heavily diluted (200x) during sample preparation. The biggest sources of dilution are the neutralization of TFA (10x the volume) and dilution with water (5x the volume). It should be investigated if the samples need to be diluted to that extent.

The author of SPEED method compared SPEED with SP3, among other, with cell and tissue containing 1 and 20 µg proteins and found SPEED to outperform the other methods regarding protein identifications [53]. The study also underlines that the performance of the methods are sample-type dependent, highlighting that there are no one-size-fits all method. In a comprehensive study comparing samples of 50 µg protein from HeLa cells with different methods including SP3, ISD-Urea, and SPEED, it was concluded that the SPEED method was favorable because of the short handling time and low consumable cost, yet still providing a comprehensive proteome [71]. Another study compared 50 µg proteins in cell lysate with SP3 and SPEED. With 2 hour gradient, 2212 proteins were identified with SPEED compared to SP3 with 1991 proteins [54]. These studies comparing SPEED with other methods have come to similar conclusions regarding the protein identification yield. The protein content of the cell samples was not quantified (except for the ISD-Urea-prepared samples in the initial screening), and absolute quantification of the proteins in the cell lysate of the SPEED samples was not found in this work. Therefore, a direct comparison between this work and the previous studies is difficult. However, the protein content in the samples of this thesis was probably <50 µg, estimated by the approximate number of cells.

Despite the lack of quantification method for the SPEED method, was it decided to prepare organoid samples with this method as it provided a superior performance with the highest number of protein identifications and unique proteins among all the cell samples.

Control samples indicate carryover in the LC-MS system

One sample without cells, was prepared accordingly to each method, here defined as a control sample. Because there were difficulties with the quantification of lysate in some of the methods, was it important to have a control sample. The control samples were used to evaluate the contribution of contaminants introduced during the sample preparations. In the initial screening 0-6 proteins were detected in the control samples, whereas in the second evaluation, there was detected 13-90 proteins (**Table 11**). Highly abundant proteins, e.g. 60 kDa heat shock proteins and actin cytoplasmic 2 were found in almost all the control samples that detected proteins. This indicates that there is some carryover in the LC-MS system, even with a blank sample of 30 min gradient employed after a high concentration sample. Strangely, serum human albumin was found in all the samples of the second evaluation, which is the most abundant protein in blood plasma [143]. The HepG2 cells were not fed with serum human albumin, making the source of this protein unknown, and supporting the possible carry-over issue. Keratin was not found in the control samples of the initial screening but was detected in all the control samples of the second evaluation. Working with three times the number of samples compared to the initial screening may have contributed to the introduction of contaminants, as more time was spent on the laboratory or gloves may have not been replaced often enough. There was also more traffic in the lab at the point of the second evaluation, exposing the samples to more keratin. Suggestion on how to limit contamination is found in **Section 7.2.1 in Appendix**.

Table 11. The number of protein identifications in the control samples.

	SP3	ISD-Urea	SPEED
Initial screening (n=1)	6	0	0
Second evaluation (n=1)	13	90	49

To sum up: ISD-Urea provided the lowest number of protein identifications among the samples, and it was concluded to be not suitable for limited samples. The SP3-prepared samples had a higher number of protein identifications compared to ISD-Urea, however, the method was not as easy to follow. The reproducibility is highly dependent on the operator's skills. A

quantification method for SPEED-prepared samples was not found with the instrumentations at hand. The lack of quantification was the biggest drawback of the SPEED method. Nevertheless, SPEED provided the highest number of protein identification, the shortest handling time, and easy to follow method. SP3 and SPEED were chosen as methods to examine organoid samples.

4.5 Comparing SP3 and SPEED for organoid-analysis

To evaluate the performance of SP3 and SPEED on liver organoids, one organoid sample was prepared with each method. Because of the limited amount of organoids available, one replicate was used. As SPEED-prepared samples could not be quantified, three samples of organoids were pooled and divided in two. The SP3 sample was quantified using BCA assay, and it was estimated that the SPEED sample would have approximately the same protein yield after lysis. The SP3 sample had 7.9 µg proteins in the sample, and beads were added in a 10:1 beads/sample ratio as described by the author of the original method [73]. SP3 did not find any unique proteins, compared to the 3862 unique proteins, in addition to the 442 common proteins, found in the SPEED-prepared organoid sample (**Figure 24**).

Figure 24. Venn diagram comparing unique proteins found in the SP3 and SPEED-prepared samples containing 7.9 µg proteins each. The overlapping area (colored orange) represents the proteins found in both groups. (n=1)

The low protein yield, despite having increased the bead content was found unexpected. After discussing the issue with a support scientist at Cytiva [144] was it recommended to decrease the stock solution to 10 µg/µL beads (compared to 20 µg/µL used in this work) as one of their customers experienced an improved protein recovery in their study [145]. The low protein yield can be due to contaminated beads or that they were more than two years old, thus compromising the protein recovery [144]. Discussion about the SP3 method is found in **Section 4.4**.

To sum up: The organoid sample prepared with SPEED had a higher protein identification yield. Despite difficulties with quantification it was decided to be the method of choice for the organoid samples. Instead, a relative quantification using LFQ after LC-MS analysis can be employed.

4.6 µPAC LC-MS analysis of SPEED-prepared organoids

The optimized SPEED method was used on samples consisting of 20-30 organoids, equal to approximately 50 000 cells. Relative quantification with LFQ was used since quantification of protein content was not possible. Using SPEED without protein quantification excluded the possibility of isobaric labeling which is the preferred choice for relative quantification of limited samples which increases the peptide concentration in the analytical column, and minimizes the effect by run-to-run variations. **Figure 25** and **Table 17** show that with DDA the number of unique proteins was not equal to the average protein yield indicating run-to-run variations.

In contrast to the previous experiments in this work, these organoid samples were analyzed using a µPAC column, which has a theoretically improved peak capacity compared to the packed column. When using the 25 cm packed nanoLC column with one hour gradient, a number of 2321-2907 (n=2) proteins were identified with 100 ng HeLa digest. However, with the µPAC column and longer gradient (2-4 hours), 2642-2769 proteins were identified with 100 ng HeLa digest (**Table 20 in the Appendix**). The µPAC column did not provide a higher number of identifications compared to the packed column. The packed column was placed in a column oven set at 50°C which may have contributed to improved peak capacity. The µPAC column did not fit in the column oven and an elevated temperature could not be obtained. Longer gradients of up to 5 hours are recommended for µPAC columns to take advantage of the column length [54, 146]. To connect the µPAC column to the nanoLC-MS system at the core facility, additional two connectors and a 20 µm ID and 30 cm long capillary tube were used. The capillary tube may have contributed to band broadening after the analytical column, and ideally, the column should be directly coupled to the emitter.

HeLa digest standard (100 ng, n=1) with the main gradient of 120, 180, and 200 min were analyzed to evaluate which gradient to use for the organoids samples (**Table 20 in the**

Appendix). The difference between the 180 min and 200 min was only 43 protein identifications, and it was decided to use the 180 min gradient since the instrument time was limited.

The samples were of liver organoids with different metabolic states, here called control (without metabolic properties), protocol 2, and protocol 3. Four technical replicates (N=4) were analyzed. There number of protein identifications for the organoid samples were on average 2140, 2619, and 3203 proteins for control, protocol 2, and protocol 3, respectively. Raw data and representative chromatogram are found in **Section 7.5 in the Appendix**. The number of unique proteins among the replicates was 2485, 2975, and 3638 for control, protocol 2, and protocol 3, respectively (**Figure 25** and **Table 12**).

Table 12. The average number of protein identifications (N=4).

Metabolic state	Control	Protocol 2	Protocol 3
Proteins (average)	2140	2619	3203
Unique proteins	2485	2975	3638

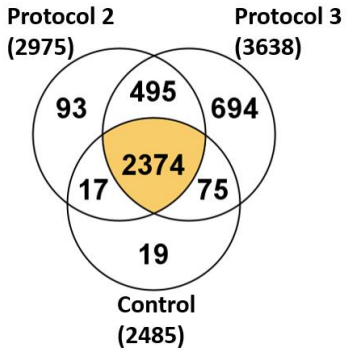

Figure 25. Venn diagram of the unique proteins found in the samples, where the orange overlapping area represents the proteins found in every sample. The small areas in between two groups show the common proteins found in these groups.

Ph.D. Tulla Nyman and Ph.D. Thiede performed the statistical data processing and LFQ at the core facility. It was hypothesized that increased metabolic activity for the organoids generated by protocols 2 and 3 compared to the control. The protocol 2 organoids would have increased cholesterol biosynthesis, TCA cycle, oxidative pathway, fatty acids beta-oxidation, and gluconeogenesis, while organoids from protocol 3 would have increased glucose uptake,

glycogen storage, lipogenesis, and bile acid synthesis. Principle component analysis (PCA) was used to evaluate the performance of the system and if the organoids samples were significantly different. The PCA plot (**Figure 26A**) demonstrates reproducible results as the technical replicates are clustered together, and significant differences between the three organoid samples as they are spread across the plot.

The proteins in the sample were not quantified during the sample preparation. However, there are reasons to believe that the amount of proteins is nearly the same, as the samples contained 20-30 organoids and was treated in the same way. Hence, relative quantification was performed by LFQ. The heat map (**Figure 26B**) visualizes the relative abundances of proteins identified and demonstrates a clear difference between the organoid samples. The volcano plot is a scatter plot (**Figure 26C**) that illustrates which proteins (marked as green or red dots on the plot) that were significantly down- or up-regulated compared to the control. These proteins were put in the KEGG pathway and showed that organoids from protocol 2 largely met the hypothesized traits, while proteins from protocol 3 had fewer unique traits (**Figure 26D**). This can be due to the fact that many of the proteins (e.g. beta-catenin, APC, etc.) in the hypothesized pathways are in very low concentrations and have been challenging to quantify even with targeted approaches [147].

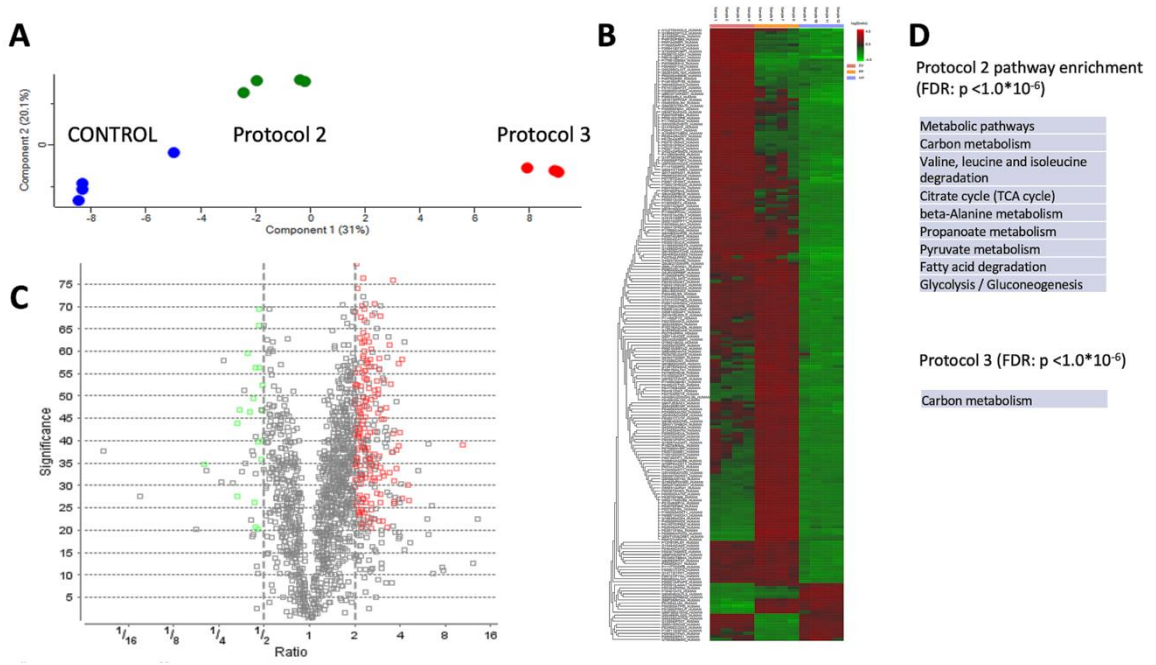

Figure 26. A) PCA of samples from organoids generated by standard protocol (control), protocol 2, and protocol 3. The clustering of samples shows good reproducibility of the system, yet significant differences between the samples. B) Heatmap of the relative abundances of proteins for control, protocol 2, and protocol 3 (n = 4) samples. The tree left on the heatmap is structured accordingly to the protein family. C) Volcano plot of all samples showing the proteins that deviate from the control sample. Red dots represent the proteins that are upregulated and the green dots represent the proteins that are downregulated. D) List of upregulated pathways relative to control samples.

To sum up: The PCA plot and LFQ showed a significant difference between the liver organoid samples of different metabolic states. The SPEED method in combination with µPAC-ESI-MS was successfully implemented as a preliminary step towards NAFLD studies.

4.7 Limitations of the study

The methods were evaluated with the three sample replicates of HepG2 cells in the second evaluation. Due to the limited amount of organoid samples, having three sample replicates is often not possible. However, for further proteomics studies on organoids, at least three biological replicates should be analyzed to establish the biological differences between the organoids at different metabolic states, and if possible, have three sample replicates to establish the reproducibility and accuracy of the method.

In this work, SPEED was proven to be difficult to quantify. Turbidity measurement and tryptophan fluorescence should be further investigated as quantification methods for SPEED-prepared samples. Also, the colorimetric method as Bradford assay could be considered.

Quantification allows for direct comparison between the methods, as well as peptide labeling to eliminate the run-to-run variations associated with LFQ.

In addition to the sample preparation, optimization of the LC-MS system can increase the number of protein identifications and proteome coverage. For proteomics samples, specialized nanoLC SP materials [28], gradient time of several hours [28, 148], trap column [132], and/or the use of other additives such as DFA [54] are often employed to increase efficiency and sensitivity. As the cell and organoid samples were analyzed in a core facility, limited access to improvement of the method was possible and the focus for this thesis was on evaluating sample preparation methods rather than optimizing the nanoLC-ESI-MS method to increase the protein yield.

5 Conclusion and future perspective

The protein yield for samples in the range 10 000-100 000 cells were 266-629, 831-1708, and 1684-3503 proteins for ISD-Urea, SP3, and SPEED, respectively. The SPEED-prepared samples had the highest protein yield, however, it was proven difficult to quantify the protein content in the samples. Hence, organoid samples were pooled and divided into two equal fractions, where one fraction was lysed and prepared accordingly to the SP3 method and quantified using BCA assay, while the other fraction was prepared by the SPEED method, estimating equal lysis yield as for the SP3-prepared fraction. For 7.9 µg proteins, 442 proteins were detected for the SP3 fraction and 4304 for the SPEED fraction, implying a better lysis method for SPEED. Both cell and organoid samples prepared with SP3 method had less number of protein identification, but are compatible with the BCA assay for protein quantification. An ideal sample preparation consists of an easy protein quantification method while yielding a high number of protein identifications. Despite the quantification difficulties, SPEED was chosen as the method for analysis of liver organoids at different metabolic states, and was successfully implemented using µPAC-ESI-MS for analysis and LFQ. The results showed significant difference in the protein expression between the organoid samples, where number of protein identifications was 2140-3203 proteins and many of the hypothesized protein traits were found.

A revisit of **Figure 19** in Aim of study (**Section 2**), here illustrated in **Figure 27**, shows that an optimized sample preparation method was successfully implemented on liver organoids. The thesis is a preliminary study and first step towards NAFLD studies where "obese" and "healthy" liver organoids are to be compared, and their crosstalk with pancreatic islet organoids.

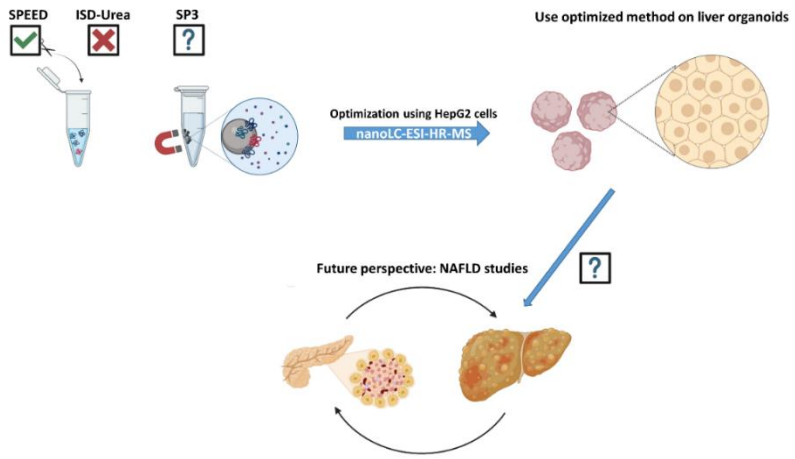

Figure 27. A revisit of Figure 19 in Aim of study. SPEED protocol had the highest number of protein identifications, while SP3 provided a quantification method using BCA assay. ISD-Urea had the lowest number of protein identifications. This thesis is a preliminary study towards NAFLD studies.

6 References

[1] K. Riazi, H. Azhari, J.H. Charette, F.E. Underwood, J.A. King, E.E. Afshar, M.G. Swain, S.E. Congly, G.G. Kaplan, A.-A. Shaheen, *The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis*, The Lancet Gastroenterology & Hepatology, 7 **(2022)** 851-861

[2] N. Chalasani, Z. Younossi, J.E. Lavine, A.M. Diehl, E.M. Brunt, K. Cusi, M. Charlton, A.J. Sanyal, *The diagnosis and management of non-alcoholic fatty liver disease: Practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association*, Hepatology, 55 **(2012)** 2005-2023

[3] N. Stefan, H.-U. Häring, K. Cusi, *Non-alcoholic fatty liver disease: causes, diagnosis, cardiometabolic consequences, and treatment strategies*, The Lancet Diabetes & Endocrinology, 7 **(2019)** 313-324

[4] Z.M. Younossi, *Non-alcoholic fatty liver disease – A global public health perspective*, Journal of Hepatology, 70 **(2019)** 531-544

[5] Z. Younossi, Q.M. Anstee, M. Marietti, T. Hardy, L. Henry, M. Eslam, J. George, E. Bugianesi, *Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention*, Nature Reviews Gastroenterology & Hepatology, 15 **(2018)** 11-20

[6] N. Pydyn, K. Miękus, J. Jura, J. Kotlinowski, *New therapeutic strategies in nonalcoholic fatty liver disease: a focus on promising drugs for nonalcoholic steatohepatitis*, Pharmacological Reports, 72 **(2020)** 1-12

[7] H. Nordli, *Towards method miniaturization for determination of sterols in liver organoids: an investigation of nano-liquid chromatography-mass spectrometry (nLC-MS)*, **(2021)**, Master's thesis

[8] R. Mittal, F.W. Woo, C.S. Castro, M.A. Cohen, J. Karanxha, J. Mittal, T. Chhibber, V.M. Jhaveri, *Organ-on-chip models: Implications in drug discovery and clinical applications*, Journal of Cellular Physiology, 234 **(2019)** 8352-8380

[9] X. Yin, Benjamin E. Mead, H. Safaee, R. Langer, Jeffrey M. Karp, O. Levy, *Engineering Stem Cell Organoids*, Cell Stem Cell, 18 **(2016)** 25-38

[10] M. Li, J.C. Izpisua Belmonte, *Organoids — Preclinical Models of Human Disease*, New England Journal of Medicine, 380 **(2019)** 569-579

[11] J.A. Davies, *Chapter 1 - Organoids and mini-organs: Introduction, history, and potential*, in: Organoids and Mini-Organs, Academic Press, (**2018**), pp. 3-23

[12] A. Marsee, F.J.M. Roos, M.M.A. Verstegen, A. Marsee, F. Roos, M. Verstegen, H. Clevers, L. Vallier, T. Takebe, M. Huch, W.C. Peng, S. Forbes, F. Lemaigre, E. de Koning, H. Gehart, L. van der Laan, B. Spee, S. Boj, P. Baptista,… B. Spee, *Building consensus on definition and nomenclature of hepatic, pancreatic, and biliary organoids*, Cell Stem Cell, 28 **(2021)** 816-832

[13] C. Tian, Q. Tu, W. Liu, J. Wang, *Recent advances in microfluidic technologies for organ-on-a-chip*, TrAC Trends in Analytical Chemistry, 117 **(2019)** 146-156

[14] C.M. Leung, P. de Haan, K. Ronaldson-Bouchard, G.-A. Kim, J. Ko, H.S. Rho, Z. Chen, P. Habibovic, N.L. Jeon, S. Takayama, M.L. Shuler, G. Vunjak-Novakovic, O. Frey, E. Verpoorte, Y.-C. Toh, *A guide to the organ-on-a-chip*, Nature Reviews Methods Primers, 2 **(2022)** 33

[15] S. Kogler, K.S. Kømurcu, C. Olsen, J.-y. Shoji, F.S. Skottvoll, S. Krauss, S.R. Wilson, H. Røberg-Larsen, *Organoids, organ-on-a-chip, separation science and mass spectrometry: An update*, TrAC Trends in Analytical Chemistry, 161 **(2023)** 116996

[16] *Method of the Year 2017: Organoids*, Nature Methods, 15 **(2018)** 1-1

[17] Z. Zhao, X. Chen, A.M. Dowbaj, A. Sljukic, K. Bratlie, L. Lin, E.L.S. Fong, G.M. Balachander, Z. Chen, A. Soragni, M. Huch, Y.A. Zeng, Q. Wang, H. Yu, *Organoids*, Nature Reviews Methods Primers, 2 **(2022)** 94
[18] D.L. Nelson, *Structure and Catalysis* in: A.L. Lehninger, M.M. Cox (Eds.) Lehninger principles of biochemistry, W.H. Freeman, New York, (**2013)**, pp. 45-493
[19] N.A. Campbell, J.B. Reece, L.A. Urry, M.L. Cain, S.A. Wasserman, P.V. Minorsky, *Biological Macromolecules and Lipids*, in: Biology: a global approach, Pearson, Boston, (**2015)**, pp. 116-141
[20] R.L. Miesfeld, M.M. McEvoy, R.L. Miesfeld, *Protein Function* in: Biochemistry, W. W. Norton & Company, New York, (**2021)**, pp. 250-307
[21] G.A. Petsko, D. Ringe, *From Sequence to Structure* in: Protein structure and function, New Science Press, London, (**2004)**, pp. 1-47
[22] R.L. Miesfeld, M.M. McEvoy, R.L. Miesfeld, *Physical Biochemistry: Energy Conversion, Water, and Membranes*, in: Biochemistry, W. W. Norton & Company, New York, (**2021)**, pp. 38-89
[23] N.A. Campbell, J.B. Reece, L.A. Urry, M.L. Cain, S.A. Wasserman, P.V. Minorsky, *Biology and its themes*, in: Biology: a global approach, Pearson, Boston, (**2015)**, pp. 48-77
[24] N.C. Mishra, *Historical perspectives*, in: Introduction to proteomics: principles and applications, Wiley, Hoboken, N.J, (**2010)**, pp. 1-38
[25] F.E. Lottspeich, Joachim, *Proteome Analysis* in: Bioanalytics - Analytical Methods and Concepts in Biochemistry and Molecular Biology Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim (**2018)**, pp. 977-1022
[26] C. Arias-Hidalgo, P. Juanes-Velasco, A. Landeira-Viñuela, M.L. García-Vaquero, E. Montalvillo, R. Góngora, Á.-P. Hernández, M. Fuentes, *Single-Cell Proteomics: The Critical Role of Nanotechnology*, International journal of molecular sciences, 23 **(2022)** 6707
[27] *Single-cell proteomics: challenges and prospects*, Nature Methods, 20 **(2023)** 317-318
[28] S.R. Wilson, T. Vehus, H.S. Berg, E. Lundanes, *Nano-LC in proteomics: recent advances and approaches*, Bioanalysis, 7 **(2015)** 1799-1815
[29] Y. Shi, R. Xiang, C. Horváth, J.A. Wilkins, *The role of liquid chromatography in proteomics*, Journal of Chromatography A, 1053 **(2004)** 27-36
[30] N. Iwamoto, T. Shimada, *Recent advances in mass spectrometry-based approaches for proteomics and biologics: Great contribution for developing therapeutic antibodies*, Pharmacology & Therapeutics, 185 **(2018)** 147-154
[31] H.K. Hustoft, L. Reubsaet, T. Greibrokk, E. Lundanes, H. Malerod, *Critical assessment of accelerating trypsination methods*, Journal of Pharmaceutical and Biomedical Analysis, 56 **(2011)** 1069-1078
[32] F.W. McLafferty, K. Breuker, M. Jin, X. Han, G. Infusini, H. Jiang, X. Kong, T.P. Begley, *Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics*, The FEBS Journal, 274 **(2007)** 6256-6268
[33] H. Steen, M. Mann, *The abc's (and xyz's) of peptide sequencing*, Nature Reviews Molecular Cell Biology, 5 **(2004)** 699-711
[34] A.D. Catherman, O.S. Skinner, N.L. Kelleher, *Top Down proteomics: Facts and perspectives*, Biochemical and Biophysical Research Communications, 445 **(2014)** 683-693
[35] P.D. Compton, L. Zamdborg, P.M. Thomas, N.L. Kelleher, *On the scalability and requirements of whole protein mass spectrometry*, Analytical Chemistry, 83 **(2011)** 6868-6874
[36] F.E. Lottspeich, Joachim, *Protein Analytics* in: Bioanalytics - Analytical Methods and Concepts in Biochemistry and Molecular Biology, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, (**2018)**, pp. 3-410

[37] M. Shehadul Islam, A. Aryasomayajula, P.R. Selvaganapathy, *A Review on Macroscale and Microscale Cell Lysis Methods*, Micromachines, 8 **(2017)** 83
[38] *Traditional Methods of Cell Lysis for Protein Extraction*, https://www.thermofisher.com/no/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/traditional-methods-cell-lysis.html, **13th March 2023**
[39] H. Kolsrud Hustoft, *A Critical Review of Trypsin Digestion for LC-MS Based Proteomics*, in, IntechOpen, (**2012)**,
[40] *EDTA (ethylenediaminetetraaceticacid)*, in: G.P. Rédei (Ed.) Encyclopedia of Genetics, Genomics, Proteomics and Informatics, Springer Netherlands, Dordrecht, (**2008)**, pp. 582-582
[41] D.C. Harris, C.A. Lucy, D.C. Harris, *Fundamentals of Spectrophotometry* in: Quantitative chemical analysis, Macmillan International Higher Education, New York, NY, (**2016)**, pp. 432-460
[42] N.J. Anthis, G.M. Clore, *Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm*, Protein Science, 22 **(2013)** 851-858
[43] Thermo Fisher Scientific, *Protein A280 Thermo Scientific NanoDrop Spectrophotometers*, **(2022)**
[44] Thermo Fisher Scientific, *Protocol for BCA Protein Assay using a Nanodrop 2000/2000c Spectrophotometer*, **(2022)**
[45] Thermo Fisher Scientific, *Bradford Protein Assay*, **(2023)**
[46] J.M. Manns, *SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Proteins*, Current Protocols in Microbiology, 22 **(2011)** A.3M.1-A.3M.13
[47] A. Shevchenko, H. Tomas, J. Havli, J.V. Olsen, M. Mann, *In-gel digestion for mass spectrometric characterization of proteins and proteomes*, Nature Protocols, 1 **(2006)** 2856-2860
[48] J.R. Wiśniewski, A. Zougman, N. Nagaraj, M. Mann, *Universal sample preparation method for proteome analysis*, Nature Methods, 6 **(2009)** 359-362
[49] Y. Ren, Z. Shi, C. Zhang, Y. Han, S. Liu, P. Hao, *Evaluation and minimization of over-alkylation in proteomic sample preparation*, International Journal of Mass Spectrometry, 481 **(2022)** 116919
[50] C.A. Evans, *Reducing Complexity? Cysteine Reduction and S-Alkylation in Proteomic Workflows: Practical Considerations*, in: C.A. Evans, P.C. Wright, J. Noirel (Eds.) Mass Spectrometry of Proteins: Methods and Protocols, Springer New York, New York, NY, (**2019)**, pp. 83-97
[51] E. Vandermarliere, M. Mueller, L. Martens, *Getting intimate with trypsin, the leading protease in proteomics*, Mass Spectrometry Reviews, 32 **(2013)** 453-465
[52] K. Hodge, S.T. Have, L. Hutton, A.I. Lamond, *Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS*, Journal of Proteomics, 88 **(2013)** 92-103
[53] J. Doellinger, A. Schneider, M. Hoeller, P. Lasch, *Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction*, Molecular & Cellular Proteomics, 19 **(2020)** 209-222
[54] H.E. Berg, *Exploring miniaturized chromatography formats and sample preparation methods for shotgun proteomics in limited biological samples*, **(2022)**, Ph.D. Thesis
[55] R.M. Tubaon, P.R. Haddad, J.P. Quirino, *Sample Clean-up Strategies for ESI Mass Spectrometry Applications in Bottom-up Proteomics: Trends from 2012 to 2016*, Proteomics, 17 **(2017)** 1700011
[56] E. Lundanes, L. Reubsaet, T. Greibrokk, *Sample Preparation*, in: Chromatography: basic principles, sample preparations and related methods, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, (**2014)**, pp. 161-188

[57] A.L. Richards, A.E. Merrill, J.J. Coon, *Proteome sequencing goes deep*, Current Opinion in Chemical Biology, 24 **(2015)** 11-17
[58] R. Wu, S. Xing, M. Badv, T.F. Didar, Y. Lu, *Step-Wise Assessment and Optimization of Sample Handling Recovery Yield for Nanoproteomic Analysis of 1000 Mammalian Cells*, Analytical Chemistry, 91 **(2019)** 10395-10400
[59] R.T. Kelly, *Single-cell Proteomics: Progress and Prospects*, Molecular & Cellular Proteomics, 19 **(2020)** 1739-1748
[60] N. Slavov, *Single-cell protein analysis by mass spectrometry*, Current Opinion in Chemical Biology, 60 **(2021)** 1-9
[61] J.F. Leary, *Cytometry of Single-Cells for Biology and Biomedicine*, in: F.-G. Tseng, T.S. Santra (Eds.) Essentials of Single-Cell Analysis: Concepts, Applications and Future Prospects, Springer Berlin Heidelberg, Berlin, Heidelberg, (**2016)**, pp. 235-255
[62] M. Glynn, D. King, J. Ducrée, *Systems Biology in Single Cells*, in: F.-G. Tseng, T.S. Santra (Eds.) Essentials of Single-Cell Analysis: Concepts, Applications and Future Prospects, Springer Berlin Heidelberg, Berlin, Heidelberg, (**2016)**, pp. 31-53
[63] C. Ctortecka, K. Mechtler, *The rise of single-cell proteomics*, Analytical Science Advances, 2 **(2021)** 84-94
[64] G. Kaur, J.M. Dufour, *Cell lines - Valuble tools or useless artifacts*, Spermatogenesis, 2 **(2012)** 1-5
[65] R. Krithika, R. Mohankumar, R.J. Verma, P.S. Shrivastav, I.L. Mohamad, P. Gunasekaran, S. Narasimhan, *Isolation, characterization and antioxidative effect of phyllanthin against CCl4-induced toxicity in HepG2 cell line*, Chemico-Biological Interactions, 181 **(2009)** 351-358
[66] V.A. Arzumanian, O.I. Kiseleva, E.V. Poverennaya, *The Curious Case of the HepG2 Cell Line: 40 Years of Expertise*, International journal of molecular sciences, 22 **(2021)** 13135
[67] Z. Zhou, M.-J. Xu, B. Gao, *Hepatocytes: a key cell type for innate immunity*, Cellular & Molecular Immunology, 13 **(2016)** 301-315
[68] J.R. Masters, *HeLa cells 50 years on: the good, the bad and the ugly*, Nature Reviews Cancer, 2 **(2002)** 315-319
[69] *Pierce™ HeLa Protein Digest Standard*, https://www.thermofisher.com/order/catalog/product/88329, **23rd of May 2023**
[70] T. Köcher, P. Pichler, R. Swart, K. Mechtler, *Quality control in LC-MS/MS*, Proteomics, 11 **(2011)** 1026-1030
[71] G. Varnavides, M. Madern, D. Anrather, N. Hartl, W. Reiter, M. Hartl, *In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods*, Journal of proteome research, 21 **(2022)** 2397-2411
[72] H.E. Johnston, K. Yadav, J.M. Kirkpatrick, G.S. Biggs, D. Oxley, H.B. Kramer, R.S. Samant, *Solvent Precipitation SP3 (SP4) Enhances Recovery for Proteomics Sample Preparation without Magnetic Beads*, Analytical Chemistry, 94 **(2022)** 10320-10328
[73] C.S. Hughes, S. Moggridge, T. Müller, P.H. Sorensen, G.B. Morin, J. Krijgsveld, *Single-pot, solid-phase-enhanced sample preparation for proteomics experiments*, Nature Protocols, 14 **(2019)** 68-85
[74] M. Sielaff, J. Kuharev, T. Bohn, J. Hahlbrock, T. Bopp, S. Tenzer, U. Distler, *Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range*, Journal of proteome research, 16 **(2017)** 4060-4072
[75] H. Røberg-Larsen, E. Lundanes, T.A. Nyman, F.S. Berven, S.R. Wilson, *Liquid chromatography, a key tool for the advancement of single-cell omics analysis*, Analytica Chimica Acta, 1178 **(2021)** 338551

[76] E. Lundanes, L. Reubsaet, T. Greibrokk, *High-Performance Liquid Chromatography (HPLC)* in: Chromatography: basic principles, sample preparations and related methods, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, (**2014)**, pp. 47-104
[77] S.R. Wilson, C. Olsen, E. Lundanes, *Nano liquid chromatography columns*, Analyst, 144 **(2019)** 7090-7104
[78] J.P. Chervet, M. Ursem, J.P. Salzmann, *Instrumental Requirements for Nanoscale Liquid Chromatography*, Analytical Chemistry, 68 **(1996)** 1507-1512
[79] E. Lundanes, L. Reubsaet, T. Greibrokk, *General Concepts* in: Chromatography: basic principles, sample preparations and related methods, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, (**2014)**, pp. 1-16
[80] J. Šesták, D. Moravcová, V. Kahle, *Instrument platforms for nano liquid chromatography*, Journal of Chromatography A, 1421 **(2015)** 2-17
[81] G. Yue, Q. Luo, J. Zhang, S.L. Wu, B.L. Karger, *Ultratrace LC/MS proteomic analysis using 10-μm-i.d. porous layer open tubular poly(styrene-divinylbenzene) capillary columns*, Analytical Chemistry, 79 **(2007)** 938-946
[82] M. Callewaert, J.O. De Beeck, K. Maeno, S. Sukas, H. Thienpont, H. Ottevaere, H. Gardeniers, G. Desmet, W. De Malsche, *Integration of uniform porous shell layers in very long pillar array columns using electrochemical anodization for liquid chromatography*, Analyst, 139 **(2014)** 618-625
[83] W. De Malsche, H. Eghbali, D. Clicq, J. Vangelooven, H. Gardeniers, G. Desmet, *Pressure-Driven Reverse-Phase Liquid Chromatography Separations in Ordered Nonporous Pillar Array Columns*, Analytical Chemistry, 79 **(2007)** 5915-5926
[84] W. De Malsche, J. Op De Beeck, S. De Bruyne, H. Gardeniers, G. Desmet, *Realization of 1 × 106 Theoretical Plates in Liquid Chromatography Using Very Long Pillar Array Columns*, Analytical Chemistry, 84 **(2012)** 1214-1219
[85] J. Op De Beeck, M. Callewaert, H. Ottevaere, H. Gardeniers, G. Desmet, W. De Malsche, *Suppression of the sidewall effect in pillar array columns with radially elongated pillars*, Journal of Chromatography A, 1367 **(2014)** 118-122
[86] P. Gzil, N. Vervoort, G.V. Baron, G. Desmet, *Advantages of Perfectly Ordered 2-D Porous Pillar Arrays over Packed Bed Columns for LC Separations: A Theoretical Analysis*, Analytical Chemistry, 75 **(2003)** 6244-6250
[87] R. Aebersold, M. Mann, *Mass spectrometry-based proteomics*, Nature, 422 **(2003)** 198-207
[88] M.A. Baldwin, *Protein Identification by Mass Spectrometry: Issues to be Considered**, Molecular & Cellular Proteomics, 3 **(2004)** 1-9
[89] J. Greaves, J. Roboz, *Analytical techniques and strategies* in: Mass spectrometry for the novice, CRC Press, Boca Raton, (**2014)**, pp. 135-157
[90] J. Greaves, J. Roboz, *Proteomics*, in: Mass spectrometry for the novice, CRC Press, Boca Raton, (**2014)**, pp. 163-187
[91] J. Greaves, J. Roboz, *Measures of instrument performance* in: Mass spectrometry for the novice, CRC Press, Boca Raton, (**2014)**, pp. 111-122
[92] A.G. Marshall, C.L. Hendrickson, *High-Resolution Mass Spectrometers*, Annual Review of Analytical Chemistry, 1 **(2008)** 579-599
[93] C. Ctortecka, K. Mechtler, *The rise of single-cell proteomics*, Analytical Science Advances, 2 **(2021)** 84-94
[94] S.H. Hansen, S. Pedersen-Bjergaard, *High-Performance Liquid Chromatography (HPLC) and High-Performance Liquid Chromatography - Mass Spectrometry (LC-MS)* in: Bioanalysis of pharmaceuticalstimstof : sample preparation, separation techniques, and mass spectrometry, Wiley, West Sussex, England, (**2015)**, pp. 123-172

[95] M. Wilm, M. Mann, *Analytical Properties of the Nanoelectrospray Ion Source*, Analytical Chemistry, 68 **(1996)** 1-8
[96] *Trapped Ion Mobility Spectrometry (TIMS)*, https://www.bruker.com/en/products-and-solutions/mass-spectrometry/timstof.html, **7th december2022**
[97] F. Meier, M.A. Park, M. Mann, *Trapped Ion Mobility Spectrometry and Parallel Accumulation–Serial Fragmentation in Proteomics*, Molecular & Cellular Proteomics, 20 **(2021)** 100138
[98] F. Meier, S. Beck, N. Grassl, M. Lubeck, M.A. Park, O. Raether, M. Mann, *Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device*, Journal of proteome research, 14 **(2015)** 5378-5387
[99] J.H. Gross, *Linear Quadrupole Ion Traps*, in: Mass Spectrometry : A Textbook Springer International Publishing : Imprint: Springer, Cham, (**2017)**, pp. 201-225
[100] J.H. Gross, *Linear Quadrupole Instruments* in: Mass Spectrometry: A Textbook, Springer International Publishing : Imprint: Springer, Cham, (**2017)**, pp. 190-200
[101] P. Ek, *Personal communication by email: timsTOF effective path length* **(2023)**
[102] J.H. Gross, *Time-of-Flight Instruments*, in: Mass Spectrometry: A Textbook, Springer International Publishing : Imprint: Springer, Cham, (**2017)**, pp. 155-177
[103] *timsTOF pro* https://www.napi.uio.no/ms-instruments/university-of-oslo/timstof-pro.html, **1st of April 2023**
[104] J.A. Silveira, M.E. Ridgeway, F.H. Laukien, M. Mann, M.A. Park, *Parallel accumulation for 100% duty cycle trapped ion mobility-mass spectrometry*, International Journal of Mass Spectrometry, 413 **(2017)** 168-175
[105] J. Guo, T. Huan, *Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography–Mass Spectrometry Based Untargeted Metabolomics*, Analytical Chemistry, 92 **(2020)** 8072-8080
[106] L.C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, *Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis**, Molecular & Cellular Proteomics, 11 **(2012)** O111.016717
[107] S. Tyanova, T. Temu, J. Cox, *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics*, Nature Protocols, 11 **(2016)** 2301-2319
[108] B.C. Orsburn, *Proteome Discoverer—A Community Enhanced Data Processing Suite for Protein Informatics*, Proteomes, 9 **(2021)** 15
[109] B.C. Searle, *Scaffold: A bioinformatic tool for validating MS/MS-based proteomic studies*, Proteomics, 10 **(2010)** 1265-1269
[110] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, G. Lajoie, *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry*, Rapid Communications in Mass Spectrometry, 17 **(2003)** 2337-2342
[111] Proteome Software Inc., *Scaffold Version 5.0 User's Manual* **(2021)**
[112] M. Brosch, L. Yu, T. Hubbard, J. Choudhary, *Accurate and Sensitive Peptide Identification with Mascot Percolator*, Journal of proteome research, 8 **(2009)** 3176-3181
[113] The Gene Ontology Consortium, *The Gene Ontology Resource: 20 years and still GOing strong*, Nucleic Acids Research, 47 **(2018)** D330-D338
[114] R. Moulder, S.D. Bhosale, D.R. Goodlett, R. Lahesmaa, *Analysis of the plasma proteome using iTRAQ and TMT-based Isobaric labeling*, Mass Spectrom Rev, 37 **(2018)** 583-606
[115] P.L. Ross, Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, D.J. Pappin, *Multiplexed Protein Quantitation in Saccharomyces*

cerevisiae Using Amine-reactive Isobaric Tagging Reagents, Molecular & Cellular Proteomics, 3 **(2004)** 1154-1169
[116] A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, C. Hamon, *Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS*, Analytical Chemistry, 75 **(2003)** 1895-1904
[117] Y. Zhang, B.R. Fonslow, B. Shan, M.-C. Baek, J.R. Yates, III, *Protein Analysis by Shotgun/Bottom-up Proteomics*, Chemical Reviews, 113 **(2013)** 2343-2394
[118] A.A. Altomare, G. Aiello, J.L. Garcia, G. Garrone, B. Zoanni, M. Carini, G. Aldini, A. D'Amato, *Protein Profiling of a Cellular Model of NAFLD by Advanced Bioanalytical Approaches*, International journal of molecular sciences, 23 **(2022)** 9025
[119] D.A. Megger, T. Bracht, H.E. Meyer, B. Sitek, *Label-free quantification in clinical proteomics*, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1834 **(2013)** 1581-1590
[120] M. Schüller, *A targeted proteomics approach for the investigation of drug-induced liver injury in hepatic organoids using nano liquid chromatography mass spectrometry*, **(2020)**, Master's thesis
[121] F.S. Skottvoll, *Critical evaluation of isolation and characterization techniques of breast cancer exosomes*, **(2017)**, Master's thesis
[122] H.E. Berg, *Optimization of in-house packing of nano liquid chromatography columns*, **(2016)**, Master's thesis
[123] Thermo Fisher Scientific, *Acetone precipitation of proteins*, **(2022)**
[124] V. LaLone, A. Aizenshtadt, J. Goertz, F.S. Skottvoll, M.B. Mota, J. You, X. Zhao, H.E. Berg, J. Stokowiec, M. Yu, A. Schwendeman, H. Scholz, S.R. Wilson, S. Krauss, M.M. Stevens, *Quantitative chemometric phenotyping of three-dimensional liver organoids by Raman spectral imaging*, Cell Reports Methods, **(2023)** 100440
[125] A. Aizenshtadt, S. Liu, C. Wang, A.D. Fernández, J. Stokowiec, F.S. Skottvoll, M. Busek, S. Rayner, H. Scholz, S.R. Wilson, S. Krauss, *In vitro zonitation of human iPSC-derived 3D liver organoids* Hepatology, 76 **(2022)** S1-S1564
[126] Thermo Fisher Scientific, *Instructions Pierce BCA Protein Assay Kit - Reducing Agent Compatible*
[127] N.G. Theodorakis, *Personal communications by email: Pierce BCA Assay kit/compability reagent*, **(2022)**
[128] J. Doellinger, *Personal communication by email: Questions regarding protein quantitation using SPEED*, **(2023)**
[129] B. Budnik, E. Levy, G. Harmange, N. Slavov, *SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation*, Genome Biology, 19 **(2018)** 161-161
[130] Y. Zhu, P.D. Piehowski, R. Zhao, J. Chen, Y. Shen, R.J. Moore, A.K. Shukla, V.A. Petyuk, M. Campbell-Thompson, C.E. Mathews, R.D. Smith, W.-J. Qian, R.T. Kelly, *Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells*, Nature Communications, 9 **(2018)** 882-810
[131] B. Cañas, C. Piñeiro, E. Calvo, D. López-Ferrer, J.M. Gallardo, *Trends in sample preparation for classical and second generation proteomics*, Journal of Chromatography A, 1153 **(2007)** 235-258
[132] X. Duan, R. Young, R.M. Straubinger, B. Page, J. Cao, H. Wang, H. Yu, J.M. Canty, Jr., J. Qu, *A Straightforward and Highly Efficient Precipitation/On-Pellet Digestion Procedure Coupled with a Long Gradient Nano-LC Separation and Orbitrap Mass Spectrometry for Label-Free Expression Profiling of the Swine Heart Mitochondrial Proteome*, Journal of proteome research, 8 **(2009)** 2838-2850

[133] X. Tian, M.P. de Vries, H.P. Permentier, R. Bischoff, *A Versatile Isobaric Tag Enables Proteome Quantification in Data-Dependent and Data-Independent Acquisition Modes*, Analytical Chemistry, 92 **(2020)** 16149-16157
[134] Thermo Fisher Scientific, *TMT Mass Tagging Kits and Reagents*, **(2023)**
[135] J. Cortés-Ríos, A.M. Zárate, J.D. Figueroa, J. Medina, E. Fuentes-Lemus, M. Rodríguez-Fernández, M. Aliaga, C. López-Alarcón, *Protein quantification by bicinchoninic acid (BCA) assay follows complex kinetics and can be performed at short incubation times*, Analytical Biochemistry, 608 **(2020)** 113904
[136] P.D. Crescenzio, *Pierce BCA Assay Kit*, **(2022)**
[137] M.M. Bradford, *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*, Analytical Biochemistry, 72 **(1976)** 248-254
[138] A.C. Lucarini, B.V. Kilikian, *Comparative study of Lowry and Bradford methods: interfering substances*, Biotechnology Techniques, 13 **(1999)** 149-154
[139] Thermo Fisher Scientific, *Pierce Detergent Compatible Bradford Assay Kit*, **(2023)**
[140] J. Doellinger, A. Schneider, M. Hoeller, P. Lasch, *Supporting information: Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction*, Molecular & Cellular Proteomics, 19 **(2020)** 209-222
[141] *Fluoreporter™ FITC Protein Labeling Kit* https://www.thermofisher.com/order/catalog/product/F6434?SID=srch-srp-F6434, **31st of May 2023**
[142] J.J. Jasper, H.L. Wedlick, *Effect of Temperature on the Surface Tension and Density of Trifluoroacetic Acid*, J. Chem. Eng. Data, 9 **(1964)** 446-447
[143] A.M. Merlot, D.S. Kalinowski, D.R. Richardson, *Unraveling the mysteries of serum albumin—more than just a serum protein*, Frontiers in Physiology, 5 **(2014)**
[144] R. Harpio, *Personal communication by email: Shelf life and problem solving regarding Sera-Mag*, **(2023)**
[145] *Reprodicule protein and peptide cleanup for mass spectrometry with Sera-Mag Carboxylate SpeedBeads by SP3*, **1st of May 2023**
[146] H.E. Berg, S. Halldórsson, E.A. Bakketeig, B. Thiede, C.J. Sandberg, E. Lundanes, E. Vik-Mo, S.R. Wilson, *Micro-pillar array columns (μPAC): An efficient tool for comparing tissue and cultured cells of glioblastoma*, Journal of Chromatography Open, 2 **(2022)** 100047
[147] T. Vehus, K.E. Seterdal, S. Krauss, E. Lundanes, S.R. Wilson, *Comparison of commercial nanoliquid chromatography columns for fast, targeted mass spectrometry-based proteomics*, Future Science OA, 2 **(2016)** null
[148] E.J. Hsieh, M.S. Bereman, S. Durand, G.A. Valaskovic, M.J. MacCoss, *Effects of Column and Gradient Lengths on Peak Capacity and Peptide Identification in Nanoflow LC-MS/MS of Complex Proteomic Samples*, Journal of the American Society for Mass Spectrometry, 24 **(2013)** 148-153
[149] A.B.T. Ghisaidoobe, S.J. Chung, *Intrinsic Tryptophan Fluorescence in the Detection and Analysis of Proteins: A Focus on Förster Resonance Energy Transfer Techniques*, International journal of molecular sciences, 15 **(2014)** 22518-22538
[150] A. Zapata, S. Ramirez-Arcos, *A Comparative Study of McFarland Turbidity Standards and the Densimat Photometer to Determine Bacterial Cell Density*, Current Microbiology, 70 **(2015)** 907-909
[151] I.M. Aune, *Investigation the utilization of nano liquid chromatography mass spectrometry for targeted peptide analysis of four blood glucose regulating peptide hormones: insulin, somatostatin-14, glucagon and urocortin 3*, **(2022)**, Master's thesis

[152] J. Greaves, J. Roboz, *Components of instruments and their function*, in: Mass spectrometry for the novice, CRC Press, Boca Raton, (**2014)**, pp. 11-23

7 Appendix

7.1 Other techniques for measuring protein concentration

7.1.1 Tryptophan fluorescents

Tryptophan is amino acid with most dominant fluorescents [36, 149]. Tryptophan fluorescence is based on measuring the fluorescence of tryptophan at the excitation wavelength 295 nm and emission wavelength 350 nm [53], and is used for protein studies and protein quantification due to its high sensitivity and robustness [149]. The total protein concentration is estimated by using a calibration solution containing tryptophan to generate the calibration curve. Assuming that the tryptophan content in complex samples is 1.3%, the protein concentration can be estimated [53, 140].

7.1.2 Turbidity measurement

Some protocols develop alternative approaches for protein quantification. Sample preparation by easy extraction and digestion (SPEED) uses TFA and Trisbase, which are incompatible with BCA assay as the salt concentration is too high and will interfere with the assay. Since proteins aggregate to form small particles in the TFA-Trisbase solution, a turbidity measurement can be performed at 360 nm where biological compounds usually do not absorb at [53]. Since the turbidity is instrument dependent, a standard solution with suspension need to be used to determine the coefficient (**Equation 7**) needed to calculate the protein content in the sample (**Equation 6**). This kind of suspension standard is commonly used in microbiology labs regarding the measurement of bacterial suspension [150]. Turbidity measurement provides rapid quantification, and the sample aliquot can be collected for downstream sample preparation [53].

The calculation behind protein quantification using turbidity measurement for cell lysate is as follows [140]:

$$C_{protein}\left[\frac{\mu g}{\mu L}\right] = A360_{sample} \times coefficient \qquad \textbf{Equation 6}$$

Where $C_{protein}$ is the protein concentration and $A360_{sample}$ is the measured absorbance of the sample at 360 nm. Depending on the concentration of the standard, the coefficient is [140]:

$$coefficient = \frac{\left[Protein\ in\ suspension\ standard,\ \frac{\mu g}{\mu L}\right]}{A360_{supension\ standard}} \qquad \textbf{Equation 7}$$

7.2 Supplementary information for experimental sections

7.2.1 Measures to limit contamination introduced during the sample preparation

- Wear gloves and a laboratory coat when working with proteomic samples. Keratins are structural proteins found in high abundance in skin, hair, and nails. If in high abundance in samples, keratins will suppress the ions originating from lower abundant proteins. This is especially the case when using DDA [52].

- Rinse the gloves as they accumulate static charge and attract dust [47].

- Work in a fume hood. Proper ventilation will provide a near to dust-free environment [47].

- Use protein low-binding tubes as proteins tend to stick to surfaces [52].

- Use only high-purity grade (HPLC, LC-MS grade) chemicals (ACN, TFA, FA, water, etc.) because the chemicals can introduce metal ions that form adducts with other compounds or cause ion suppression [52].

- Use a dedicated box for pipette tips that are kept close when not in use [52].

7.2.2 Cultivation of HepG2 cells

Ph.D. candidate Ingrid Wilhelmsen used the following protocol for the cultivation of HepG2 cells.

Information

Name: HepG2, derivative of ATCC #: HB-8065

Path: 11

Tissue: liver

Product Format: frozen

Morphology: epithelial

Culture Properties: adherent

Biosafety Level: 1

Disease: hepatocellular carcinoma; 15 years adolescent Caucasian male

Materials List

DMEM (Life Technologies; Cat#11965)

Heat-Inactivated Fetal Bovine Serum (Life Technologies Cat# 10082147)

Penicillin-Streptomycin 10,000U/mL (Life Technologies; Cat#15140))

Phosphate Buffered Saline (1X PBS) w/o Ca2+, Mg2+ (CORNING Cellgro; Cat# 21-040-CM)

TrypLE Express (Life Technologies; Cat#12604)

T75, T525 culture flasks

Graduated pipets (1, 5, 10, 25, 50 mL)

Freezing medium (growth medium containing 95%; DMSO, 5%)

DMSO (Fisher; Cat#BP-231-100)

Cryovials (Sarstedt; Cat #72-694-006)

TC20 cell counter (Bio-Rad)

Counting Slides (Bio-Rad; Cat 145-0011)

Microscope

Growth Medium for HepG2

DMEM
10% FBS
Pen-Strep (1X)

Procedure

A. Receipt of Frozen Cells and Starting Cell Culture

1) Immediately place frozen cells in liquid nitrogen freezer storage until ready to culture.

2) When ready to start cell culture, quickly thaw ampoule in a 37°C water bath.

3) As soon as ice crystals disappear, swab outside surface of the ampoule with 70% ethanol, then dispense contents of ampoule into a T75 flask with 20 ml of warm (37°C) growth media.

4) Allow cells to recover overnight in 37°C, 5% CO_2 humidified incubator.

5) The next morning, the diluted DMSO-containing shipping/cryopreservation medium is aspirated from the cell layer and replaced with fresh medium.

B. Sub-culture

Volumes used in this protocol are for 75 cm^2 and/or 525 cm^2 flask; proportionally reduce or increase amount of dissociation medium for culture vessels of the other size.
1) Propagate cells until density reaches 70-80% confluence.

2) Aspirate medium.

3) Wash cells with room temperature (or warm) PBS (1X).

4) Add 3 mL (T-75) or 30 mL (T525) of TrypLE and return to incubator for 3 minutes, or until cells detach.

5) Add 7 ml (T-75) or 30 mL (T525) of complete medium and aspirate the cells by gentle pipetting.

6) Perform 1:4 to 1:6 cell split as needed.

7) Incubate cultures at 37°C, 5% CO_2 humidified incubator.

8) Change Medium twice per week.

9) Record each subculture event as a passage.

C. Maintenance and Generation of Seed Stocks

1) Change media the day after seeding and every 3-4 days thereafter. Use 100 mL of growth medium per T525 flask.

2) Following first or second passage after receipt of cells and with sufficient number of cells to continue maintenance and expansion, the major portion of the flasks should be sub-cultured using TrypLE as above under “Sub-culture” and a small portion should be set aside as a seed stock. The cell pellet for the seed stock should be resuspended in freezing medium.

3) Cells in freezing medium are dispensed into cryovials (1-2 million cells per 1 mL aliquot) and frozen in a -80°C cryo-freezing container overnight.

4) Cryovials are transferred the next day to liquid nitrogen freezer for long-term storage.

D. Harvest

1) Passage cells until the desired number of cells is reached.

2) Remove cells from flasks as described above under “Sub-culture”.

3) Examine viability using Trypan blue staining.

7.2.3 SP3

The added volumes of reagents for the SP3 protocol are presented in **Table 13**.

Table 13. The added volume of reagents for the SP3 method.

Sample	Organoids	Initial screening (each sample)	10 000	25 000	50 000	100 000
Lysis buffer (µL)	60	300	150			
Fraction lysis buffer (µL)	47	290	8.1	20.5	40.5	81
IAM (conc. (M), volume (µL)	0.5, 2.0	0.2, 29.5	0.2, 0.8	0.2, 2.0	0.2, 4.1	0.2, 8.1
DTT (conc. (M), volume (µL)	0.25, 1.0	0.5, 3.2	0.5, 0.5	0.5, 1.1	0.5, 2.2	0.5, 4.5
20 µg/µL, beads	4	8.1	1.3	1.3	1.3	2.3
Water (µL)	0	0	39.5	25	2.0	0.0
0.005 µg/µL trypsin (µL)	100	100*	100	100	100	100

** from 0.01 µg/µL trypsin*

7.2.4 SPEED

The added volumes of reagents for the SPEED protocol are presented in **Table 14**.

Table 14. The added volume of reagents for the SPEED method.

Sample	Organoids	<100 000 cells	150 000 cells	200 000 cells
Sample volume (µL)	1	1	2	3
TFA (µL)	4	4	8	12
2 M Trisbase (µL)	40	40	80	120
0.5 M DTT (µL)	0.9	0.9	1.8	2.7
1 M IAM (µL)	0.9	0.9	1.8	2.8
0.5 M DTT (µL)	0.5	0.5	0.9	1.4
Water (µL)	210	210	423	560
1 µg/µL trypsin (µL)	0.5	0.5*	1	1
TFA (µL)	5	5	10	14

* 1 µL in the initial screening

7.2.5 Calibration solution for BCA assay for estimation of protein concentration

How to prepare the calibration solution for BCA assay containing PBS and BSA is found in **Table 15**.

Table 15. The calibration solutions for BCA assay.

Vial	PBS (µL)	Source (X - µL)	Volume of source (µL)	Conc. (µg/mL)
A	0	Stock	100	2 000
B	25	Stock	75	1 500
C	40	Stock	40	1 000
D	40	B	40	750
E	40	C	40	500
F	40	E	40	250
G	40	F	40	125
H	100	-	0	0

7.2.6 Parameters of the mass spectrometer

The parameters were set by the core facility. The capillary voltage was set to 1.5 kV. MS acquisition was performed in DDA-PASEF mode, with a mass range of 100 to 1700 m/z. The number of PASEF ranges was set to 20 with a total cycle time of 1.16 s, charge up to 5, target intensity of 20 000, intensity threshold of 1750, and active exclusion with release after 0.4 min. An inversed reduced TIMS mobility ($1/k_0$) of 0.85-1.40 Vs/cm^2 was used with a range time of 100 ms, an accumulation time of 100 ms, a duty cycle of 100%, and a ramp rate of 9.51 Hz. Precursors for DDA were fragmented with an ion mobility-dependent collision energy, which was linearly increased from 20 to 59 eV.

7.3 Initial nanoLC-ESI-MS plattform

The nanoLC system consisted of different Agilent modules (**Figure 28**). The system was able to deliver a stable nanoflow of 300 nL/min for hours, however, as with nanoLC in general, the system was often clogged, especially in the unit where the flow was split from the primary flow of 200-500 µL/min to 300 nL/min and in the 10 port valve (**Figure 29**). The SPE column and the analytical column were coupled with a 10-port valve (used as a 6-port), where a loading pump delivered the sample at a flow rate of 5 µL/min. Because of how the 10 port was set up, there were sources of dead volumes with additional capillary tubings, especially between the SPE and analytical column. The peptides will refocus on the inlet of the analytical column, however, the focusing properties of the SPE were not utilized at its fullest.

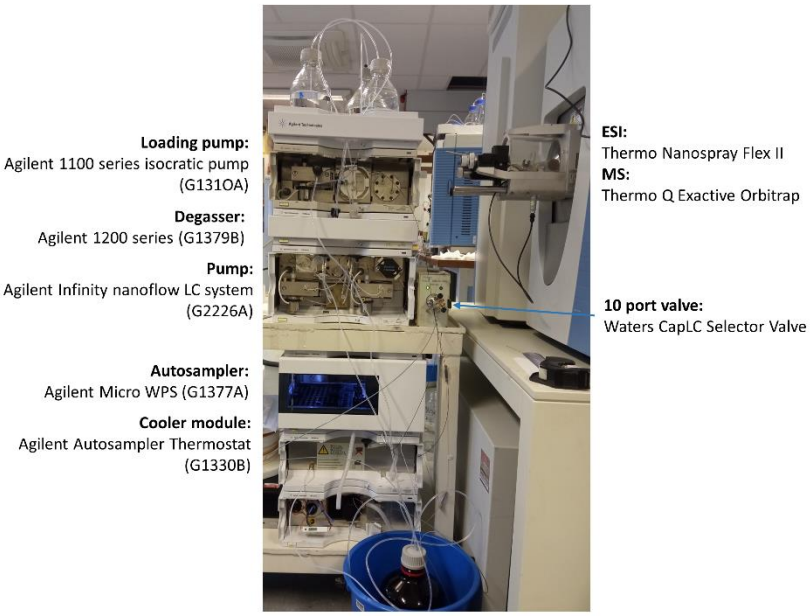

Figure 28. The setup for the nanoLC system with the Q Exactive Orbitrap system is at the right.

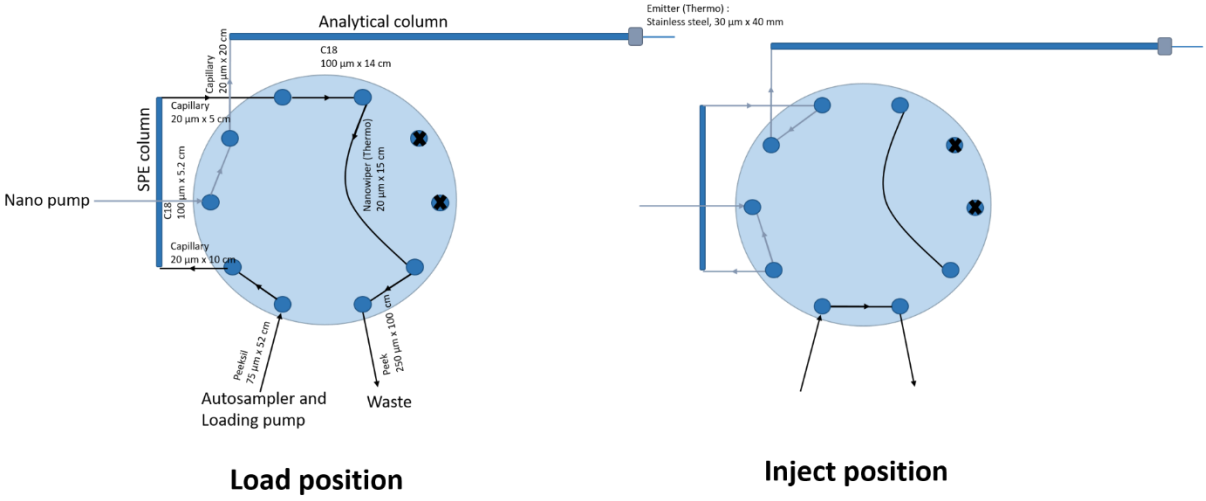

Figure 29. How the 10 port was set up, in theory, was used as a 6 port because of the two stoppers (marked with X) and an extra set of tubing (nanowiper). The coupling between the columns and tubings was made using Valco Union (for 1/16" tubings).

In the hopes of less technical difficulties by eliminating a 10 port valve between the SPE and analytical column and a flow split system, a nanoLC-MS system with Thermo EASY nLC1000 pump was used, which was successfully employed for targeted proteomics by M.Sc. Inga Mork Aune [151]. The limitation in this case was caused by the Q-Exactive Orbitrap where the blades

of the turbo molecular pump (TMP) broke off the axis (**Figure 30B**). This was probably due to the limited maintenance of the oil cartridge, which was completely charred (**Figure 30C**) causing the axis to be off-centered.

The role of the TMP is to provide the MS system with a high-vacuum system to prevent the loss of ions before detection. Different pressure is maintained at different compartments of the instruments. Depending on the type of mass analyzer the pressure varies from 10^{-3} torr (ion traps), 10^{-5} torr (quadrupole) to 10^{-10} torr (Fourier transformation-based mass analyzers). TMP and backing pumps establish the vacuum in MS instruments. A high vacuum is provided by the TMP that draws air molecules out of the instrument. The backing pump removes exhaust from the TMP, in addition to providing an optimum initial vacuum for the TMP to work at [152].

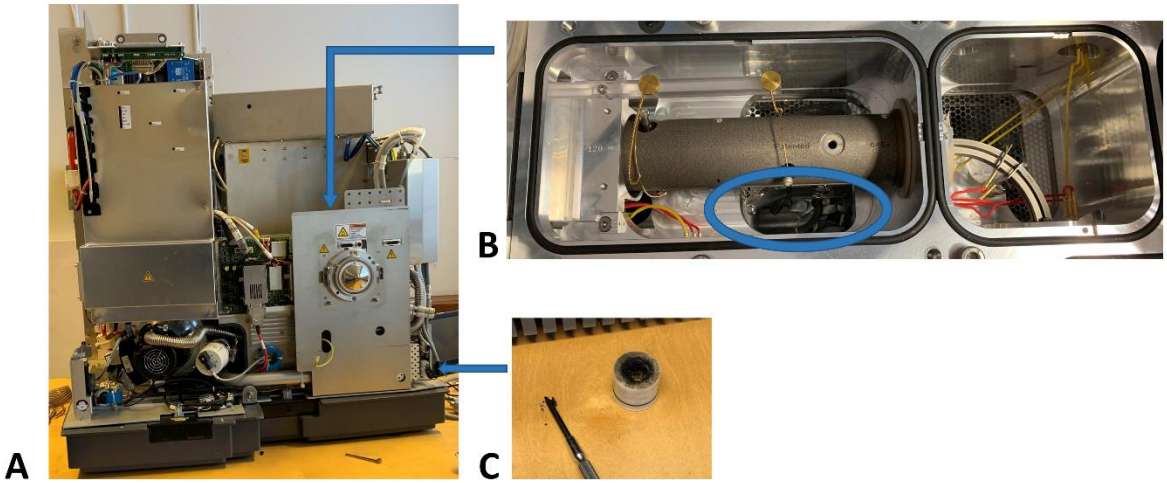

Figure 30. A) The Q-Exactive orbitrap dissembled. B) The interior wall beneath the quadrupole is ruined. Here broken blades of the TMP are circled in blue. C) The charred oil cartidge.

7.4 Evaluation of sample preparation methods

The raw data for the second evaluation (n=3) are presented in **Table 16**, **Table 17,** and **Table 18** for SP3, SPEED, and ISD-Urea, respectively. Representative base peak chromatograms (BPC) and total ion chromatograms (TIC) are found in **Figure 31-Figure 34**. The average number of protein identification with standard deviations is found in **Table 19**, and the number of unique proteins is found in **Figure 35**.

7.4.1 Raw data for SP3-prepared cell samples

Table 16. Protein identifications of samples prepared with SP3-procedure.

# Cells	Replicate number	# Proteins	# Peptides	#Spec	% IDs
0	-	13	115	1478	8
10 000	1	1184	10267	23720	43
	2	827	6335	14257	44
	3	871	7187	16759	43
25 000	1	292	2572	10026	26
	2	824	7564	17399	43
	3	1377	12934	27483	47
50 000	1	1863	19545	38354	51
	2	1716	17541	35217	50
	3	1546	14464	28616	51
100 000	1	1469	17280	37769	46
	2	2505	33637	75983	44
	3	496	4239	18923	22

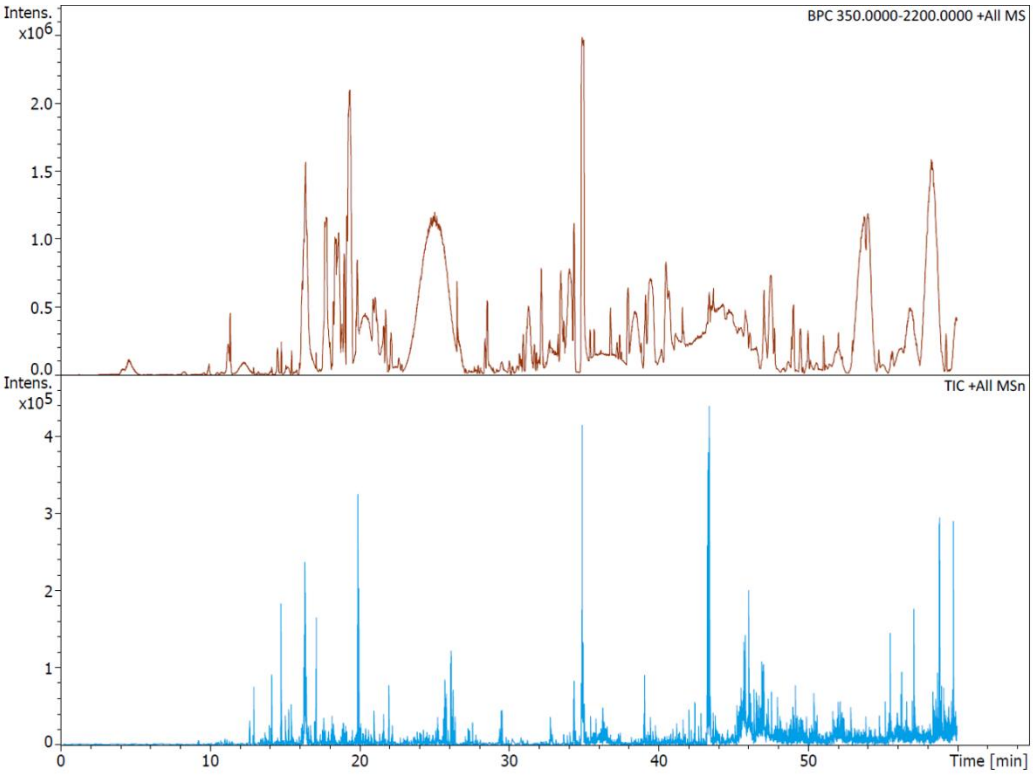

Figure 31. Representative chromatograms, BPC (brown), and total ion TIC (blue), for a sample prepared by SP3 with low protein yield. This chromatogram is the second replicate consisting of 25 000 cells.

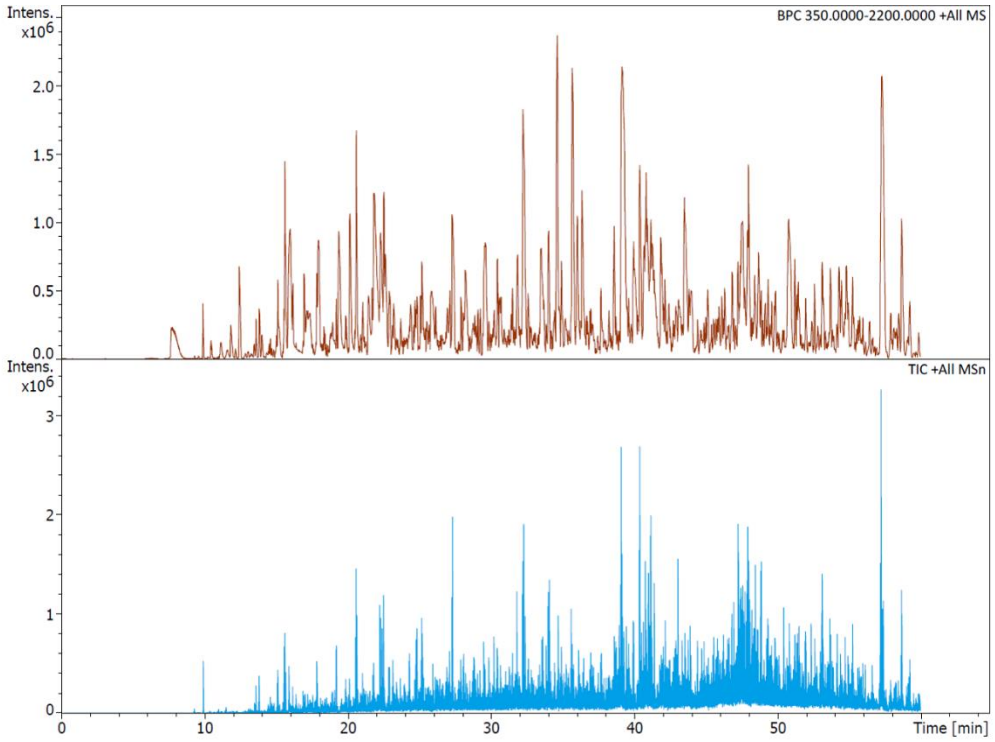

Figure 32. Representative chromatograms, BPC (brown) and total ion TIC (blue), for a sample prepared by SP3 with high protein yield. This chromatogram is the second replicate consisting of 100 000 cells.

7.4.2 Raw data for SPEED-prepared cell samples

Table 17. Protein identifications of cell samples prepared with the SPEED procedure.

# Cells	Replicate number	# Proteins	# Peptides	#Spec	% IDs
0	-	49	338	6011	6
10 000	1	2109	24705	42515	58
	2	1969	22091	37276	59
	3	975	7665	13203	58
25 000	1	2041	23940	38098	63
	2	2421	31738	50396	63
	3	2352	29405	47159	62
50 000	1	2477	34038	52681	65
	2	3036	46090	74368	62
	3	2564	34496	53480	65
100 000	1	3562	63298	109951	58
	2	3581	62430	104253	60
	3	2564	56358	92322	61

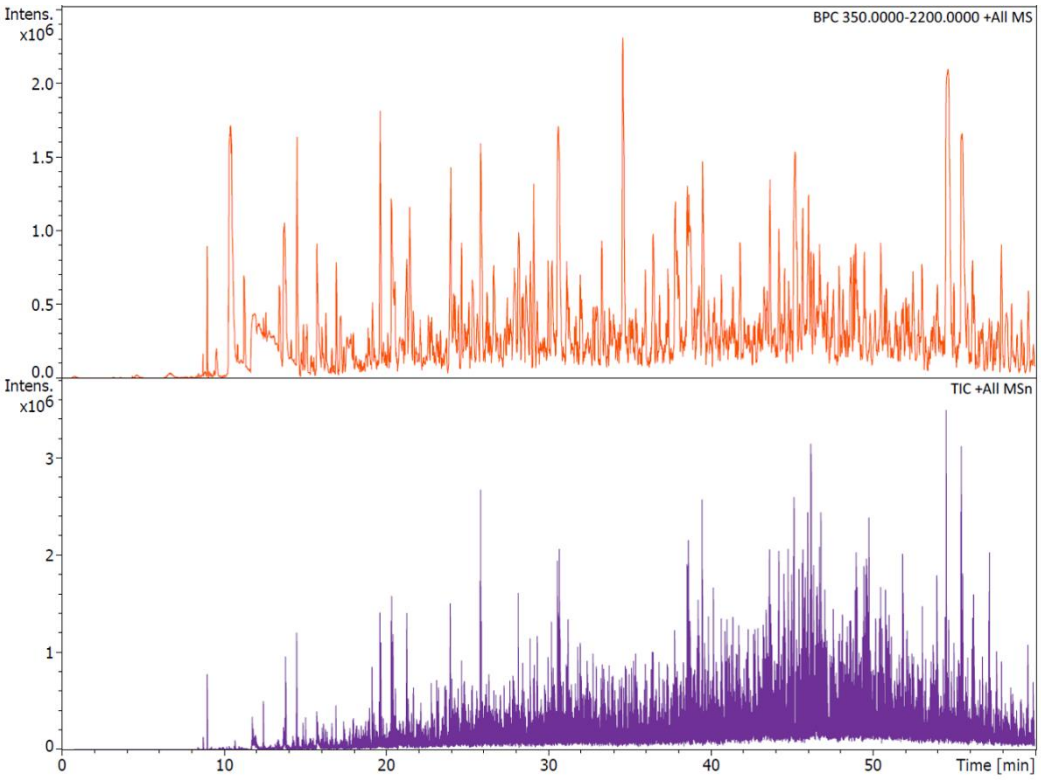

Figure 33. Representative chromatograms, BPC (brown) and total ion TIC (blue), for a sample prepared by SPEED. This is the chromatogram of the second replicate consisting of 100 000 cells.

7.4.3 Raw data for ISD-Urea-prepared cell samples

Table 18. Protein identification using the ISD-Urea method.

# Cells	Replicate number	# Proteins	# Peptides	#Spec	% IDs
0	-	90	562	7202	8
10	1	347	3037	28536	11
	2	212	1829	24765	7
	3	240	2056	25100	8
25	1	568	5361	45277	12
	2	625	6619	41996	16
	3	200	1688	20368	8
50	1	465	4667	30561	15
	2	586	6098	34925	17
	3	406	3793	29297	13
100	1	761	8492	40298	21
	2	373	3639	32168	11
	3	747	7848	46358	17

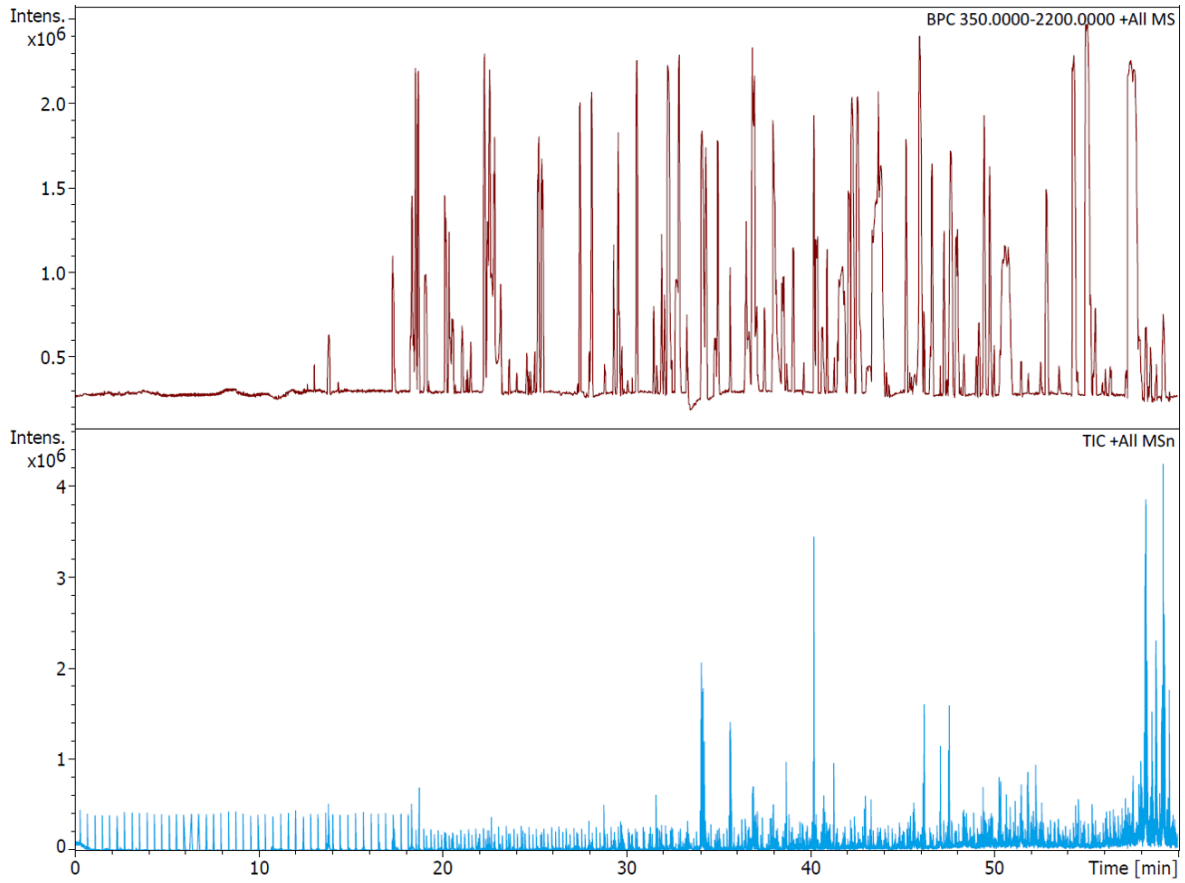

Figure 34. Representative chromatograms, BPC (brown), and total ion TIC (blue), for a sample prepared by ISD-Urea. This is the chromatogram of the second replicate consisting of 100 000 cells.

7.4.4 Average number of proteins and unique proteins for the second evaluation

The average number of protein identification is found in **Table 19**.

Table 19. An average number of cells, the standard deviation of the number of proteins, and relative standard deviations are presented. The calculations are made using the number of identification from the Statistics function in the Scaffold Software.

Method	# Cells	Average (#proteins)	Standard deviation (#proteins)	Relative standard deviation (%)
SP3	10 000	961	195	20
	25 000	831	543	65
	50 000	1708	159	9
	100 000	1490	1005	67
SPEED	10 000	1684	618	37
	25 000	2271	301	9
	50 000	2477	301	12
	100 000	3503	118	3
ISD-Urea	10 000	266	71	27
	25 000	464	231	50
	50 000	486	92	19
	100 000	627	220	35

Unique proteins found in the cell samples of the second evaluation are found in **Figure 35**. The unique proteins were found using the *Quantification* function in Scaffold. The overlapping orange area is the overlapping proteins found in every sample.

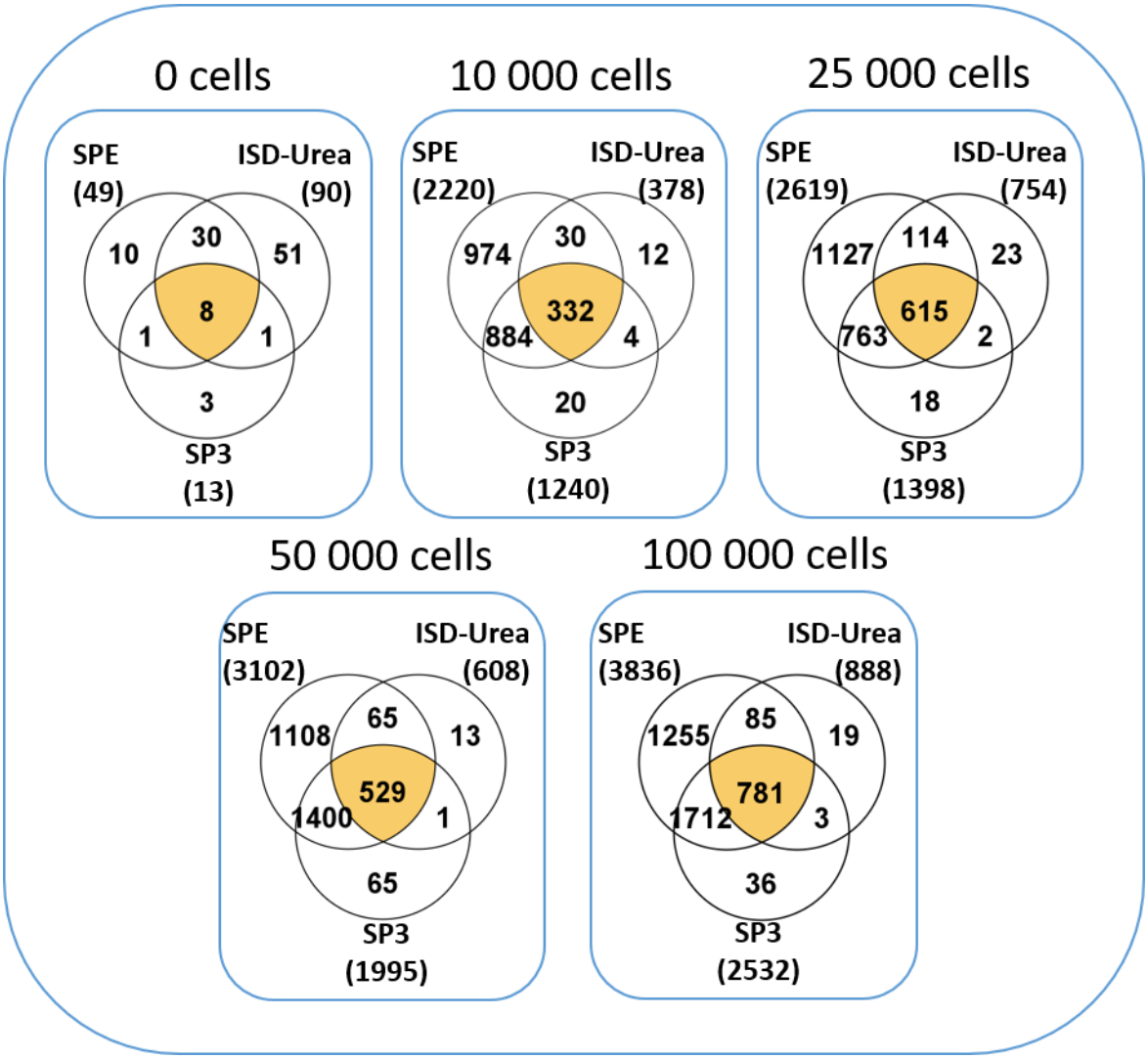

Figure 35. Number of unique proteins found in the cell samples (n=3) using the Quantification function in Scaffold Software. The overlapping area shows the common proteins found in the samples.

7.5 SPEED-prepared liver organoid samples

The number of protein identifications for the gradient optimization using HeLa digest standard is found in **Table 20**. Number of protein identifications for each technical replica is given in **Table 21**. A representative TIC is found in **Figure 36**.

Table 20. The number of protein identifications of 100 ng HeLa Digest standard to optimize gradient.

Time main gradient (min)	120	180	240
# Proteins	2642	2726	2769

Table 21. The number of protein identifications of each technical replicate.

Sample	**Replicate number**	**# Proteins**	**# Peptides**	**#Spec**	**% IDs**
Control	1	2362	48719	135675	36
	2	2066	38564	105277	37
	3	2100	38814	108321	36
	4	2030	36604	101484	36
Protocol 3	1	3115	69951	176590	40
	2	3218	74756	194301	38
	3	3186	73047	189221	39
	4	3292	75734	198113	38
Protocol 2	1	2723	62357	146515	43
	2	2553	55471	128715	43
	3	2753	61667	148957	41
	4	2446	51581	120403	43

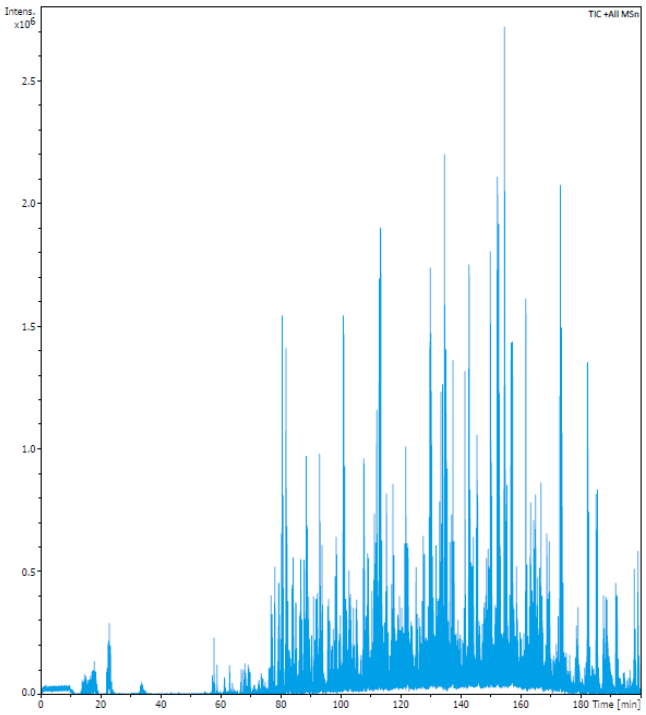

Figure 36. TIC chromatogram of control sample, replicate 4.