I hereby grant a lending permission for my master‘s thesis for educational and scientific purposes and demand for proper evidence of all people this thesis was lend to. All data and information taken from this thesis must be properly cited.

CHARLES UNIVERSITY
FACULTY OF SCIENCE

Study programme: Biology
Branch of study: Genetics, molecular biology and virology

BC. TOMÁŠ OLIVA

PROTEOMIC ANALYSIS OF MYOCARDIAL INTEGRAL MEMBRANE PROTEINS

PROTEOMICKÁ ANALÝZA MEMBRÁNOVÝCH PROTEINŮ MYOKARDU

Master's thesis

Supervisor: doc. RNDr. Jiří Petrák, Ph.D.
Supervisor-consultant: RNDr. Ondřej Vít, Ph.D.

Prague, 2019

Declaration:

I hereby declare that I have written this master's thesis on my own and that I have used only the sources listed in the references. This work or its parts were not used to acquire other or the same academic title.

In Prague, to date …… ……

Signature

Acknowledgement

I would like to thank my supervisor doc. RNDr. Jiří Petrák, Ph.D., for the opportunity to do my master's thesis in the laboratory of clinical proteomics at BIOCEV and for his patience and advices during the work on my thesis. I would also like to thank my consultant RNDr. Ondřej Vít, Ph.D., for his mentorship and invaluable advices throughout the work on my final thesis. I am also grateful to other members of the laboratory of clinical proteomics, namely Mgr. Eliška Jankovská and Mgr. Matěj Běhounek, for their help and support. Finally, I would like to thank Mgr. Pavel Talacko and Mgr. Karel Harant, for providing invaluable insight into mass spectrometry.

Abstract

Cardiovascular diseases (CVD) are the leading cause of morbidity and mortality in Europe. Over 4 million people die from CVDs annually and another 11 million people develops CVDs every year. These numbers show that there is a need for better diagnostic, prognostic and predictive biomarkers and, more importantly, a need for new and more efficient drugs. Integral membrane proteins (IMPs) are ideal candidates for new drug targets.

However, a study of IMPs represents a major challenge in current proteomics. This challenge is associated with the low abundance of IMPs, their low solubility in aqueous solvents and the absence of trypsin cleavage sites in their transmembrane segments. To overcome these issues, methods that selectively target either N-glycosylated extra-membrane segments (CSC, SPEG, N-glyco-FASP) or transmembrane segments (hpTC) were developed. In this thesis we employed a combination of two N-glyco-capture methods (SPEG and N-glyco-FASP) performed on two different samples (membrane-enriched fraction and total tissue lysate) with analysis of membrane-embedded IMP segments by hpTC and with standard non-targeted "detergent+trypsin" approach to analyze rat myocardial membrane proteome. We also performed an evaluation of employed methods for preparation of membrane fraction by western blot analysis of selected IMPs and cytosolic proteins.

This multi-pronged approach led to the identification of over 1000 unique IMPs in the rat myocardial proteome scattered between all major cellular compartments and is up to date one of the most complete analysis of rat myocardial membrane proteome. We showed that a combination of methods targeting different subgroups of IMPs is highly beneficial for membrane proteome coverage as each method provided a number of unique identifications. This complementarity of used methods was well reflected on both protein and peptide level. The data presented in this thesis also showed that a choice of a technique for preparation of membrane-enriched fraction can lead to different results in terms of enrichment of IMPs and removal of major cytoskeletal and soluble cytosolic proteins.

Keywords: heart, myocardium, integral membrane proteins, glycoproteins, mass spectrometry

Abstrakt

Kardiovaskulární onemocnění představují nejvyšší zdroj mortality i morbidity v Evropě. Více než 4 miliony lidí ročně na tato onemocnění zemře a více než 11 milionů nových případů kardiovaskulárních onemocnění se objeví každý rok. Tato čísla jen poukazují na to, že je zde potřeba nejen pro lepší diagnostické, prognostické a prediktivní biomarkery, ale také potřeba pro nové a efektivnější léky. Integrální membránové proteiny (IMP) jsou ideálními kandidáty na nové cíle léčiv.

Analýza IMP stále představuje jeden z hlavních problémů současné proteomiky. Důvodem je nízká úroveň exprese IMP, jejich snížená rozpustnost ve vodě a také nedostatek trypsinových štěpných míst v transmembránových segmentech. Byla vyvinuta celá řada metod, které se snaží tyto problémy obejít. Jedná se zejména o metody, které selektivně cílí na N-glykosylované segmenty IMP (CSC, SPEG, N-glyco-FASP), a nebo metody, které naopak cílí na hydrofobní transmembránové segmenty (hpTC). My jsme se v této práci rozhodli spojit dva přístupy, které cílí na N-glykosylované extramembránové segmenty IMP (SPEG a N-glyco-FASP) provedené na dvou typech vzorku (na celotkáňovém lyzátu a membránové frakci) spolu s hpTC metodou, která naopak cílí na hydrofobní transmembránové segmenty a s necíleným „detergent+trypsin“ přístupem. Tento kombinovaný přístup byl využit k analýze membránového proteomu potkaního myokardu. Dále bylo provedeno srovnání dvou metod přípravy membránové frakce pomocí western blotové analýzy zvolených membránových a cytosolických proteinů.

S využitím tohoto kombinovaného přístupu jsme identifikovali přes 1000 membránových proteinů v potkaním myokardiálním proteomu. Identifikované IMP pocházely ze všech hlavních buněčných organel. Ukázali jsme, že kombinovaný přístup, kde jednotlivé metody cílí na skupiny IMP odlišných chemicko-fyzikálních vlastností, vede k vyššímu celkovému pokrytí membránového proteomu, než kdybychom používali jen jednu ze zvolených metod. Komplementarita zvolených metod se projevila nejen na úrovni proteinů, ale také na úrovni peptidů. Data, prezentovaná v této práci také ukázala, že volba metody přípravy membránové frakce vede nejen k rozdílům v nabohacení membránových proteinů, ale také k rozdílům v depleci cytosolických proteinů.

Klíčová slova: srdce, myokard, integrální membránové proteiny, glykoproteiny, hmotnostní spektrometrie

Abbreviations:

2-ME	2-mercaptoethanol
ACN	acetonitrile
AcOH	acetic acid
ALS	acid labile surfactant
ARVC	arrhythmogenic right ventricular cardiomyopathy
ATP	adenosine triphosphate
BCA	bicinchoninic acid
Bcl-2	B-cell lymphoma 2 protein
BNPS-skatole	3-Bromo-3-methyl-2-(2-nitrophenylthio)-3H-indole
BSA	bovine serum albumin
CD	cluster of differentiation
CHF	chronic heart failure
CNBr	cyanogen bromide
ConA	concanavalin A
CSC	cell surface capture
CVD	cardiovascular diseases
DAVID	Database for Annotation, Visualization and Integrated Discovery
DTT	dithiothreitol
EC	enzyme commission
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ESI	electrospray ionization
FA	formic acid
FASP	filter-aided sample preparation
GA	Golgi apparatus
GAPDH	glyceraldehyde-3-phosphate dehydrogenase

GlcNAc	N-acetyl-glucosamine
GO	gene ontology
GPCR	G-protein-coupled receptor
GPI-anchor	glycosylphosphatidylinositol anchor
GRAVY	grand average of hydropathy
HCD	high-energy collision dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HILIC	hydrophilic interaction chromatography
Hpe	hypertonic
HPLC	high performance liquid chromatography
Hpo	hypotonic
hppK-CNBr	high pH, proteinase K, cyanogen bromide
hpTC	high pH, trypsin, cyanogen bromide
HRP	horseradish peroxidase
IAA	iodoacetamide
IEX-LC	ion-exchange liquid chromatography
IHD	ischaemic heart disease
IMP	integral membrane protein
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	liquid chromatography
MALDI	matrix-assisted laser desorption/ionization
MEP	membrane-embedded proteins and peptides
MF	membrane fraction
MS	mass spectrometry
MS/MS	tandem mass spectrometry
PAP	protease accessible peptide
PBS	phosphate-buffered saline

PBST	phosphate-buffered saline with Tween-20
PDGFRβ	platelet-derived growth factor receptor beta
PE	protein evidence
PMP	peripheral membrane protein
PNGase F	peptide N-glycosidase F
PTM	post-translational modification
PVDF	polyvinylidene difluoride
Q-OT-qIT	quadrupole-Orbitrap-quadrupole ion trap
RCA_{120}	*Ricinus communis* agglutinin
RP-LC	reverse phase liquid chromatography
SDC	sodium deoxycholate
SDS	sodium dodecyl sulfate
SERCA	sarco-endoplasmic reticulum Ca^{2+} transporter
SNAP	synaptosomal nerve-associated protein
SPEG	solid phase extraction of N-glycopeptides
SR	sarcoplasmic reticulum
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
TfR1	transferrin receptor protein 1
TGS	tris-Glycine-SDS buffer
TIM	translocase of the inner membrane
TL	tissue lysate
TMH	transmembrane helix
TMHMM	tied mixture hidden Markov model
TOF	time-of-flight
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet

WGA	wheat germ agglutinin

Contents

1. INTRODUCTION

Cardiovascular diseases (CVD) are the leading cause of death in Europe. It has been estimated that more than 85 million people were living with CVD in 2015 and nearly 4 million of them die from these diseases annually. The incidence in the last 25 years is steadily increasing (Wilkins *et al*., 2017). Such alarming numbers only reflect the fact that there is an urgent need for better diagnostic, prognostic and predictive biomarkers as well as a need for new and efficient drugs.

Integral membrane proteins (IMPs) execute a number of cellular functions such as cell signaling, intercellular communication, transmembrane transport, cell adhesion and many more. Due to the roles they partake in these processes, they are attractive pharmaceutical targets. In fact, up to 60% of currently approved drugs target IMPs (Yildirim *et al*., 2007). For this reason, studies of membrane proteome are of high interest of both basic and applied research.

However, proteomic analysis of IMPs is hindered by several challenges associated with their physico-chemical properties – it is their low expression, low solubility and lack of trypsin cleavage sites (Seddon *et al*., 2004). Over the years, a number of strategies have been developed to cope with or to overcome these issues. The "classical" approach targets IMPs as whole molecules. This strategy, however, has to cope with the challenges associated with IMP analysis by introducing a number of additional steps to aid with the enrichment of membrane material, low solubility and digestion. Alternatively, a "divide & conquer" approach was developed to overcome the challenges by targeting either hydrophilic or hydrophobic segments of IMPs separately (Vít and Petrák, 2017).

A combination of strategies that target different subgroups of IMPs should be highly complementary. The aim of this thesis is to confirm this presumption by combining three "divide & conquer" approaches – two targeting hydrophilic and one targeting hydrophobic segments of IMPs – with one "classical" non-targeted "detergent+trypsin" approach on the analysis of rat myocardial proteome.

2. LITERATURE REVIEW

2.1 PROTEOME AND PROTEOMICS

The word proteome was first used by Marc Wilkins in 1994 as a parallel to the word "genome" (Wasinger *et al*., 1995). In a wider sense, proteome designates the entire complement of proteins in a given cell, tissue, or organism at a certain time and/or under specific conditions. It also includes all proteoforms and all post-translational modifications (PTMs). Proteomics studies the proteome. Proteomic experiments may lead to determination of relative or absolute protein abundance, detection of protein modifications, identification of protein-protein, protein-lipid and other interactions and thus serve as a complement to other OMICS approaches and to contribute to better understanding of biological processes.

2.1.1 TECHNOLOGICAL FOUNDATIONS OF PROTEOMICS

The technological foundation of current proteomics is based on a) methods for effective protein/peptide separation which include electrophoresis and, more importantly, liquid chromatography (LC), and b) mass spectrometry (MS). While electrophoresis was rather important during the early days of proteomic research, currently most of the protocols are centered around the combination of LC with MS.

High performance liquid chromatography (HPLC) is an analytical technique used to separate, identify and quantify analytes within a complex mixture. The basic principle of HPLC relies on high-pressure pumping of a liquid solvent (mobile phase) that contains a sample through a column filled with an adsorbent (stationary phase). Several types of chromatographic separations are employed in proteomics – reversed phase LC (RP-LC), ion-exchange LC (IEX-LC) or hydrophilic interaction chromatography (HILIC). In the RP-LC the mobile phase is polar and the stationary phase is non-polar. When the polar mobile phase containing the sample flows through the non-polar stationary phase, the less polar components of the sample mixture interact with the stationary phase through van der Waals forces. The less polar the separated component is, the longer it remains retained on the column. In the IEX-LC the components of the sample mixture are separated according to their charge. The stationary phase has an opposite charge of the molecules of interest (a negatively charged stationary phase is used for positively charged analytes and *vice versa*). The stronger the charge of the sample is, the stronger will be the attraction to the stationary phase (Shi *et al*., 2004). HILIC is an alternative technique that is used to separate small polar compounds diluted in polar mobile phase on polar stationary

phase (Hemström and Irgum, 2006). Two or more types of chromatographic techniques can be also coupled together to achieve more efficient separation. Such combined chromatographic approach is then called two, three or multi-dimensional chromatography (Shi *et al*., 2004). In proteomics, LC methods are used to separate highly complex mixtures of proteins or peptides in biological samples prior their mass spectrometric analysis.

Mass spectrometry is an analytical technique used to accurately measure molecular mass of different components of the sample. A mass spectrometer consists of three main parts - an ion source, a mass analyzer and a detector. The ion source is responsible for ionization of analyzed material. The two most frequently used ionization techniques are electrospray ionization (ESI, Fenn *et al*., 1989) and matrix-assisted laser desorption/ionization (MALDI, Karas *et al*., 1985). The ions are then separated according to their mass-to-charge ratio (m/z) within an analyzer. The frequently used analyzers in proteomics include time-of-flight analyzer (TOF, Weickhardt *et al*., 1996), ion cyclotron resonance analyzer (Marshall *et al*., 1998), quadrupole ion trap (Paul, 1990) and orbitrap mass analyzer (Hu *et al*., 2005). Finally, the detector registers the amount of ions that reach the detector for m/z values.

Reverse phase liquid chromatography (RP-LC) coupled with tandem mass spectrometry (MS/MS) has become an indispensable technology in current proteomics (Aebersold and Mann, 2003). Principally, the separated peptides are eluted from the LC, ionized in the ion source and analyzed in the analyzer. During the MS/MS two or more MS measurements are performed in consecutive order. Typically, in the first MS, the molecular masses of the intact peptides are measured. From the spectra of these peptides, the precursor peptide ions are selected to undergo fragmentation. The fragmentation product ions are then again separated in the second mass analyzer and their mass (m/z) is determined (Schroeder *et al*., 2004).

The determined masses of peptides and their fragments are then compared with theoretical masses of all possible peptides and fragments acquired by *in silico* digestion and fragmentation of all theoretical peptides present in gene/protein databases. The process of identification is carried by database search engines such as Mascot (Perkins *et al*., 1999) or Sequest (Eng *et al*., 1994).

Although, by the principle, MS is not a quantification technique, several ways of how to relatively quantify proteins using the MS instrument do exist. Label-free quantification and differential isotope labeling are the two major strategies. The latter utilizes the fact that MS is capable to distinguish between peptides that are chemically

identical but differ in isotope compositions. For this reason proteins or peptides tagged with molecules containing stable isotopes (usually ^{13}C, ^{15}N or ^{18}O) can be distinguished from the proteins/peptides containing the standard variant of C, N or O as the presence of the heavy isotope produces a mass shift that allows for relative quantification (Ong *et al*., 2002).

During the label-free quantification two or more MS runs under exactly the same conditions are performed. The MS spectrum of a peptide is then compared to the MS spectrum of the same peptide in a different run and the relative quantity is determined based on the area under curve. Alternatively, a spectral counting technique can be used for relative quantification. The spectral counting is based on the presumption that more abundant proteins are identified by more peptides, or in other words, they produce more spectra (Gstaiger and Aebersold, 2009).

2.1.2 TOP-DOWN OR BOTTOM-UP?

Several ways of how to study proteome had emerged. In MS-based proteomics "top-down" and "bottom-up" are the most prominent (**Figure 2.1**). The former, top-down proteomics, studies proteins in their intact form (Lakshmanan and Loo, 2019). However, the lack of instrumentation that is capable to efficiently handle the complex mixtures of intact proteins had caused the lag of top-down proteomics behind the bottom-up (Gregorich *et al*., 2014; Catherman *et al*., 2014).

The bottom-up setup is a high throughput strategy that allows for effective protein identification. In the bottom-up experiment, the proteins of interest are first digested into peptides. Peptides generated in this manner are then subjected to MS or MS/MS analysis and by using computer algorithms searching through databases the peptides are matched with corresponding proteins (Eng *et al*., 1994; Perkins *et al*., 1999). However, a drawback of bottom-up proteomics is that a loss of information occurs during the sample preparation resulting into only partial sequence coverage during the MS/MS analysis (Gregorich *et al*., 2014).

Although top-down approach has its use in proteomic studies, the bottom-up is often the strategy of choice and for this reason the following chapters will be focused on the bottom-up setup.

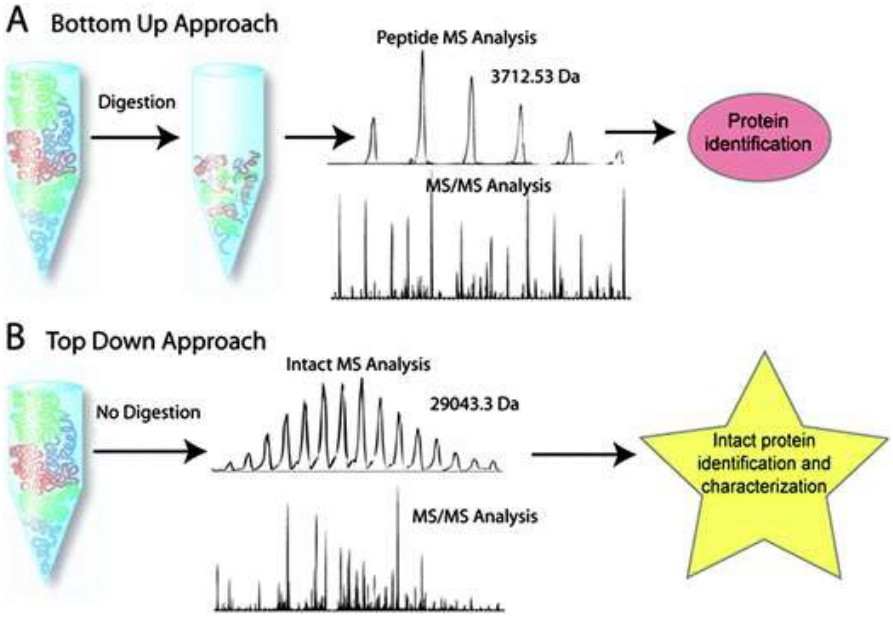

Figure 2.1: Schematic representation of bottom-up and top-down proteomic workflows adopted from Kellie *et al*. (2010)

2.2 GENERAL DESCRIPTION OF BOTTOM-UP PROTEOMIC ANALYSIS

The bottom-up proteomic analysis traditionally follows this experimental design: In the first step, the biological material is obtained and the proteome of interest is isolated from whole tissues, cells or organelles. The protein mixture is subjected to denaturation, reduction and alkylation. Such denatured proteins are then sequence-specifically cleaved into peptides, desalted/purified and separated with LC. After the separation, the peptides enter the mass spectrometer to be analyzed. The simplified schematic representation of the general bottom-up proteomic workflow is showed in **Figure 2.2**.

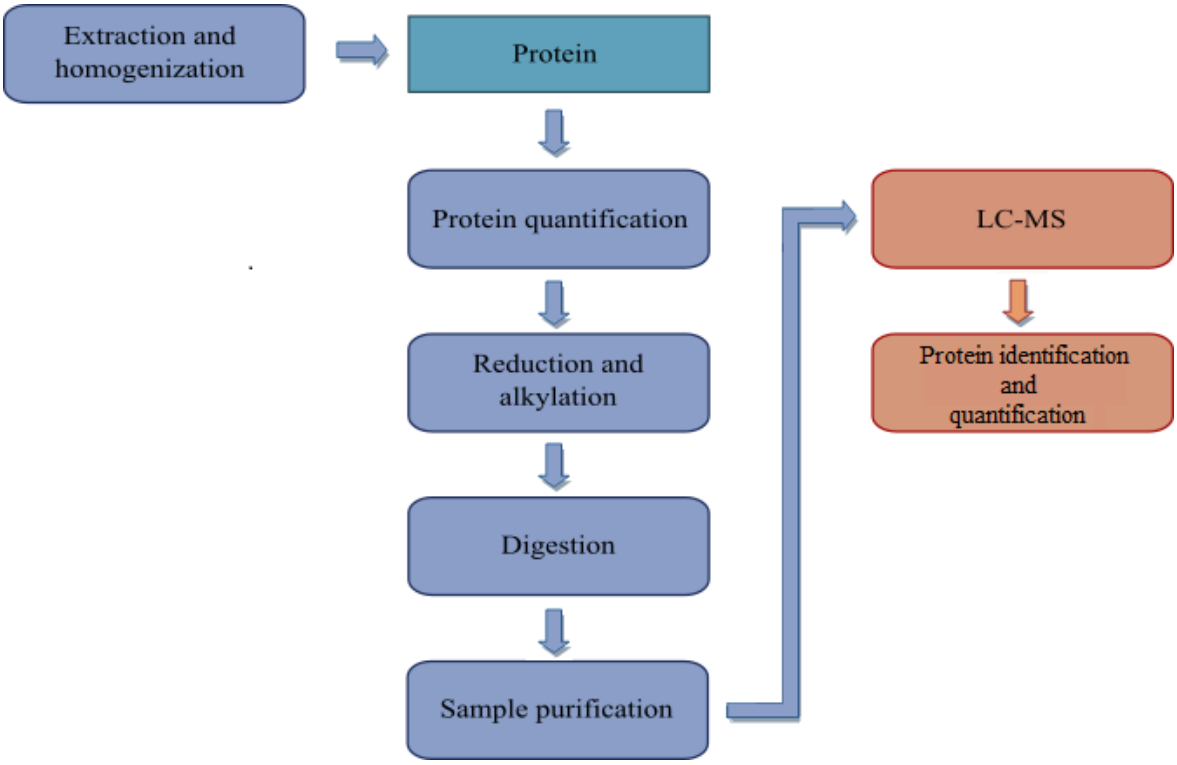

Figure 2.2: Schematic representation of general bottom-up proteomic experiment

2.2.1 EXTRACTION AND HOMOGENIZATION

The extraction and solubilization of proteins is the very first step of every proteomic analysis. The most commonly used methods for tissue disruption are, depending on the tissue character, grinding with different kinds of beads, sonication, liquid nitrogen treatment combined with mechanical homogenization, methods utilizing osmotic shock

and freeze/thaw technique. Usually a combination of several is employed to ensure proper homogenization. Protein solubilization is often facilitated by various detergents. Detergents not only solubilize proteins but also disrupt cell membranes and break down protein-lipid interactions. An effective tissue disruption and cell lysis require a combination of both mechanical and detergent-based approaches (Cañas *et al*., 2007).

Methods for membrane disruption and membrane protein solubilization will be discussed in greater detail in chapter **2.4.1.2.1 Detergents**.

2.2.2 PROTEIN QUANTIFICATION

Being able to accurately measure a total protein concentration within a sample is essential in multitude of proteomic assays.

Several methods for protein quantification have been developed. These include a measurement of UV absorption of tyrosine and tryptophan at 280nm (Warburg and Christian, 1941), Bradford assay (Bradford *et al*., 1976) or Lowry assay (Lowry *et al*., 1951). In proteomics, BCA assay (Smith *et al*., 1985) is often (but not always) the method of choice (and it is also the method employed in this thesis).

The principle of the BCA assay is a reduction of cupric (Cu^{2+}) ions to cuprous (Cu^{+}) under alkaline conditions = biuret reaction. The main contributors to the biuret reaction are three amino acids - cysteine/cystine, tyrosine and tryptophan - and a peptide backbone (Wiechelman *et al*., 1988). The resulting cuprous ions then chelate two molecules of BCA forming a complex 2BCA-Cu^{+} that is characteristic by its deep purple color. This complex exhibits a strong absorbance at 562 nm. The absorbance is directly proportional to protein quantity and can be determined from a standard curve derived from a measurement of defined concentrations of protein standard, such as bovine serum albumin (Smith *et al*., 1985).

A major advantage of the BCA assay compared to the other mentioned protein quantification techniques is that it can be used with detergents, including SDS, present in a sample. However, it is still incompatible with reducing agents (Noble *et al*., 2007; Smith *et al*., 1985).

2.2.3 REDUCTION AND ALKYLATION

The next step of general proteomic analysis is sample reduction. In other words, it is crucial to disrupt disulfide bonds formed between cysteine residues of a protein. The most commonly used agents are dithiothreitol (DTT) (Cleland, 1964), 2-mercaptoethanol (2-ME), and Tris(2-carboxyethyl)phosphine (TCEP) (Burns *et al*., 1991). The disruption of covalent interactions and thus relaxation of tertiary and secondary structure is one of the most critical step in general bottom-up proteomic approach as proteins with relaxed structure are more effectively digested into peptides.

However, the reduction alone is not sufficient. The sulfur atoms in thiol groups may spontaneously re-associate and re-form disulfide bridges. A covalent modification of cysteine residues (alkylation) by strong alkylating agents such as iodoacetamide (IAA), iodoacetic acid or acrylamide prevents re-formation of the disulfide bonds and is therefore vital for sustaining the relaxation of protein secondary and tertiary structure and thus essential for effective protein digestion (Müller and Winter, 2017).

2.2.4 CLEAVAGE STRATEGIES

The digestion of proteins into peptides is an important and indispensable sample-preparation step of every "bottom-up" analysis. This can be carried either enzymatically, involving proteolytic enzymes, non-enzymatically, by employing peptide bond-specific chemicals, or by combination of both. The digestion can be performed in two distinct ways – "in-gel" or "in-solution". Because of the focus of this work, the emphasis will be put solely on the "in-solution" methodology.

2.2.4.1 Enzymatic cleavage

An array of proteolytic enzymes varying in cleavage specificities, efficiencies and optimal working conditions is available. Over the last two decades, trypsin has become the gold standard in bottom-up proteomics. Such position is owed to its sequence specificity for arginine or lysine, its ready availability during early days of protein biochemistry as trypsin was one of the very first proteins to be purified on an industrial level (Neurath, 1994), to its low price and affordability and to the fact that resulting peptides have positive charge (Olsen *et al*., 2004).

Trypsin is a pancreatic serine protease (EC: 3.4.21.4) that was first observed and also named in 1876 by Wilhelm Kühne. Trypsin cleaves proteins into peptides

C-terminally at either arginine or lysine residue, unless followed by proline (Olsen *et al*., 2004). Both arginine and lysine are relatively abundant within mammalian/human proteome (see **Figure 2.3**). Peptides generated by tryptic cleavage are usually 500 - 3,000 Da in length, i.e. the resulting tryptic peptides have an optimal length for chromatographic separation and MS analysis.

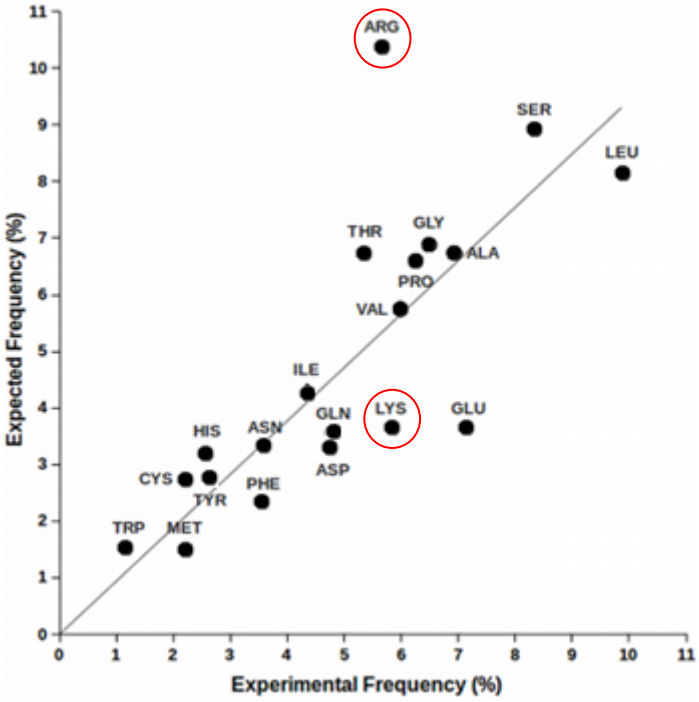

Figure 2.3: Frequency of amino acids within human proteome (Adopted from Gardini *et al*., 2016). The expected frequency stands for values that were predicted from the total of 93,493 human coding DNA sequences. The experimental frequencies stand for values that were calculated from analysis of all known human protein sequences

The trypsin digestion of proteins is performed at 37°C and in optimal pH in the range of 7.5 - 8.5, traditionally in a presence of ammonium bicarbonate. The enzyme to protein ratio used must be sufficient enough to ensure protein digestion, but also not too high to avoid autolysis. The ratio mostly lies within a range of 1:50 to 1:100 (Vandermarliere *et al*., 2013).

2.2.4.1.1 *Alternative proteases in bottom-up proteomic studies*

As an alternative to trypsin, proteases such as Lys-C, Glu-C or chymotrypsin have found its place in proteomic studies.

Lys-C is a bacterial protease that cleaves proteins C-terminally to lysine. Unlike trypsin, Lys-C cleaves lysine residues even when followed by proline. Lys-C also very well tolerates urea and for this reason is often employed together with trypsin as pre-digestion step to decrease missed cleavages (Gauci *et al*., 2009; Tsiatsiani and Heck, 2015). Glu-C, similarly to Lys-C, is also bacterial protease. It cleaves mainly at carboxyl terminus of glutamine. Interestingly, the activity and amino acid preference of Glu-C is dependent on pH of used buffer. At pH 4, Glu-C targets predominantly glutamine. However, at pH 8 it also targets asparagine alongside with glutamine (Drapeau *et al*., 1971). The major advantage of both Lys-C and Glu-C compared to trypsin is that both retain its protease activity in high concentration of urea due to which both proteases have been adopted in many proteomic studies (Giansanti *et al*., 2016). Wiśniewski *et al*. (2009) identified more than 4000 proteins in mouse hippocampus employing Lys-C coupled with trypsin. Bian *et al*. (2014) identified more than 6000 proteins in human liver samples digested by Glu-C in combination with trypsin.

However, it must be noted that high concentration of urea together with temperatures over 37°C lead to extensive protein/peptide carbamoylation. Working in lower temperate is necessary (Betancourt *et al*., 2018).

Chymotrypsin is a serine protease that cleaves C-terminally at large hydrophobic residues such as tyrosine, phenylalanine and tryptophan and to lesser extent at leucine and methionine. Chymotryptic digestion alone produces peptides longer than trypsin that may escape MS detection (Meyer *et al*., 2014). However an *in silico* digestion using a combination of chymotrypsin and trypsin showed a 100-fold decrease in occurrence of peptides larger than 4 kDa (Fischer and Poetsch, 2006).

Other proteases such as elastase, pepsin or proteinase K are sequentially less specific resulting in a mixture of multi-overlapping peptides. The non-specific proteases are generally not suitable for high-throughput proteomic experiments.

2.2.4.2 Chemical cleavage strategies

Non-enzymatic digestion is an alternative to enzymatic cleavage. Various chemical reagents are able to specifically cleave proteins. Such chemicals include formic acid (FA) (Li *et al*., 2001) that cleaves aspartic acid-proline peptide bonds, acetic acid (AcOH)

(Swatkoski *et al*., 2007) targeting aspartic residues, cyanogen bromide (CNBr) (Gross *et al*., 1961) selectively cleaving at methionine or BNPS-skatole that cleaves at tryptophan (Rahali and Gueguen *et al*., 1999; Crimmins *et al*., 2005). However, only CNBr is being more commonly employed in proteomic protocols.

Cyanogen bromide is a peptide bond hydrolyzing reagent. The reaction is carried in acidic environment. The peptide bond is cleaved C-terminally to methionine residues and results in formation of peptidyl homoserine or homoserine lactone (Gross *et al*., 1961). The efficiency of CNBr treatment was reported as excellent, reaching up to 90% of methionine sites being cleaved (Kaiser and Metzka, 1999).

The advantages of CNBr become more evident with its use on membrane proteins as it will be shown in chapter **2.4.2.2.2 Strategies targeting hydrophobic segments**.

Despite the efficiency of CNBr cleavage, the toxicity of CNBr must be always kept in mind. Although the minute amounts needed for the procedures do not pose significant thread, safe working conditions and appropriate waste disposal must be ensured (Vít and Petrák, 2017).

2.2.5 SAMPLE PURIFICATION

Prior to mass spectrometric analysis, the peptide mixture is often pre-purified from used chemicals, detergents or salts to avoid unwanted interference with LC separation and/or MS analysis. Traditionally, simplified low pressure liquid chromatography with reverse phase matrix is used. The purification of peptides using the reverse phase columns results in efficient removal of salts. However, small hydrophilic peptides may get lost during the purification (Hustoft *et al*., 2011).

2.3 BIOLOGICAL MEMBRANES AND PROTEOMICS OF INTEGRAL MEMBRANE PROTEINS

The proteomics has established itself as a fruitful and efficient approach for analysis of complex protein mixtures. Current MS-based approaches are able to identify over 10,000 proteins in a single sample depending on used fractionation (Nagaraj *et al*., 2011). However, some parts of the proteome escape the detection by traditional proteomic strategies and still remain a mystery.

The biological membranes provide physical barrier between the cell, sub-cellular compartments and, of course, the extracellular environment. The building bricks of cellular membranes are lipids. Lipids consist of two parts, a hydrophilic head-group on one side and one or more fatty acid hydrocarbon hydrophobic tails on the opposite side. This amphipathic character results in spontaneous aggregation and self-organization into spherical structures known as micelles or into bilayers. Membranes also contain a large fraction of proteins scattered between or attached to membrane lipids. Proteins usually account for 50% of membrane mass (Dupuy and Engelman, 2008).

Two distinct types of membrane proteins can be distinguished - peripheral and integral. Peripheral membrane proteins (PMPs) are either non-covalently attached to the membrane via Van der Waals forces or electrostatic interactions with the lipid head-groups or other embedded membrane proteins, or, as said above, covalently through a PTM anchor (Khan *et al*., 2016; Monje-Galvan and Klauda, 2016). Integral membrane proteins (IMPs), on the other hand, span the membrane. Two classes based on the structure of IMPs exist - α-helical and β-barrels. The latter represent only a minor subset of all IMPs and have been found only in membranes of Gram-negative bacteria and in outer layer of mitochondria and chloroplasts (Wimley, 2002). The majority of eukaryotic IMPs is of α-helical nature.

The membrane proteins represent up to 26% of all proteins encoded by mammalian genome based on prediction of transmembrane helical segments (TMHs). This number does not include proteins that associate with membrane by other mechanisms such as GPI-anchor, terpene anchor, acylation, myristoylation, or proteins that loosely interact with other membrane proteins already embedded within plasma membrane (Fagerberg *et al*., 2010).

The amphipathic character of the membrane together with the hydrophilic environment surrounding the membrane both from the inner and outer side is the driving

force of amino acid sequence of IMPs. The average length of an α-helix is 26.4 amino acid residues (Bowie, 1997). Transmembrane segments are principally composed by hydrophobic residues (Ulmschneider and Sansom, 2001). A structural analysis of 160 transmembrane helices showed that up to 78.5% of residues are of hydrophobic character among which leucine is the most abundant (Hildebrand *et al*., 2004). The hydrophobic amino acid preference within the membrane determined by Ulmschneider and Sansom (2001) from analysis of 129 transmembrane α-helices is demonstrated in **Figure 2.4**.

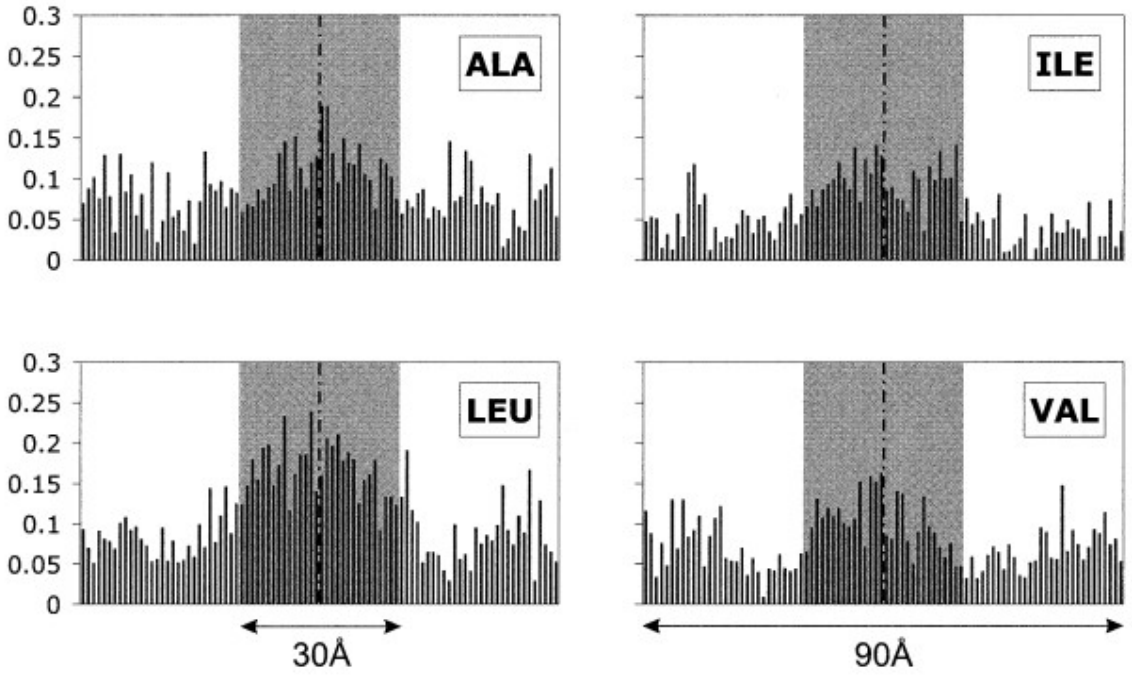

Figure 2.4: Distribution and relative abundance of the four most typical aliphatic hydrophobic residues in IMPs (adopted from Ulmschneider and Sansom, 2001). The gray area represents the 30 Å hydrophobic core of plasma membrane. The 30 Å part of the graph left from the hydrophobic core is always the cytoplasmic side of the membrane.

Aromatic amino acids, such as tyrosine, histidine and tryptophan are generally found at the interface of the membrane where they partake in anchoring the protein to the membrane through interactions with lipid head-groups (Ulmschneider and Sansom, 2001; Yau *et al*., 1998).

Charged residues, such as arginine or lysine (trypsin cleavage sites), are energetically unfavoured within the membrane core resulting in predominant

occupation of the extra-membrane segments of IMPs as seen in **Figure 2.5**. They are also tolerated at the cytoplasmic ends of TMHs (= positive inside rule) (von Heijne, 1989; Granseth *et al*., 2005). Polar chains of residues localized at the termini are also capable of so called snorkeling effect. That is, to orient the polar chain to more hydrophilic parts of the membrane resulting in pulling water into the membrane and thus creating favourable hydrophilic microenvironment for the polar chain. This is explained by the general principle that the hydrophobic segments prefer to be embedded within the hydrophobic environment while polar elements favour polar environment (Granseth *et al*., 2005).

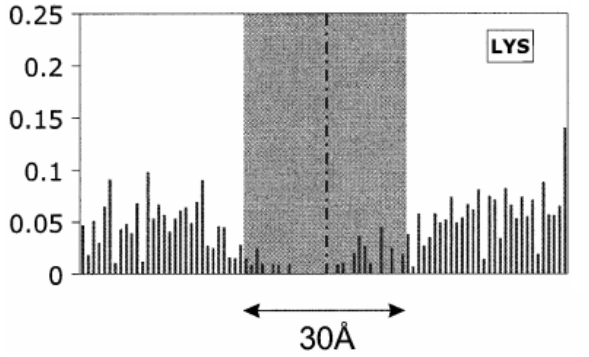

Figure 2.5: Distribution of charged amino acids in α-helices of IMPs shown for lysine (adopted from Ulmschneider and Sansom, 2001)

Polar residues predominantly occupy the extra-membrane regions of IMPs. However, they are also present in low quantities within the membrane core where they utilize the snorkeling effect to produce polar microenvironment for themselves (Illergård *et al*., 2010).

The amino acid composition of extra-membrane segments of IMPs resembles the one of soluble proteins.

IMPs play a pivotal role in many different functions such as selective transport of ions and molecules in and out of the cell and thus aiding in maintaining cellular homeostasis. IMPs are responsible for regulation of vesicular transport, attaching cells to each other, signal transduction. They also contribute to cellular identity and immunity. Due to their important functions IMPs are of paramount interest to both basic and applied research. Profiling of IMPs in a specific cell type during specific state has a great potential in finding novel disease markers and thus aiding the diagnosis, screening or monitoring. Due to their localization and functions, membrane proteins can also serve as therapeutic

targets (Vít and Petrák, 2017). In fact, up to 60% of currently approved drugs targets membrane proteins (Yildirim *et al*., 2007).

However, proteomic profiling of the membrane proteome bears several challenges resulting from the nature of membrane proteins. There are three major reasons rendering proteomic analysis of IMPs problematic:

First, α-helical proteins are amphipathic. They consist of one or more hydrophobic α-helical segments spanning across the membrane and hydrophilic extra-membrane segments. This results in low solubility of IMPs in aqueous solvents and thus, complicates proteomic studies (Vuckovic *et al*., 2013)

The second problem is that trypsin cleavage sites, represented by charged arginine and lysine, are virtually absent within hydrophobic transmembrane segments. A tryptic digestion of an IMP would generate long peptides. Such peptides are not suitable for MS analysis (Fisher and Poetsch, 2006).

Moreover, IMPs are usually expressed at low levels and are often shadowed by more abundant non-membrane proteins during MS analysis (Nagaraj *et al*., 2011).

Current most-widely used proteomic methods do not specifically address the aforementioned problems. Therefore, methods specifically designed to cope with the adverse physico-chemical properties and low abundance of IMPs have to be adopted.

2.4 DESCRIPTION OF MEMBRANE PROTEOMIC ANALYSIS

2.4.1 TOOLS USED IN MEMBRANE PROTEOMICS

Apart from the tools that are common for both general proteomic analysis and membrane protein analysis, membrane proteomics utilizes tools that deal with the specific physico-chemical properties and low abundance of IMPs, namely methods aiding with enrichment of membrane material and with solubilization and digestion of membrane proteins.

2.4.1.1 Enrichment of membrane material

The enrichment of membrane material and subsequent purification of membrane proteins is vital for any membrane proteome analysis. In general, membrane enrichment involves cell lysis that is followed by centrifugation. In this regard, cellular membranes are first disrupted by methods mentioned in chapter **2.2.1 Extraction and homogenization**. Subsequently, the membrane material is usually enriched by variety of centrifugation techniques ranging from simple differential centrifugation to density-gradient ultracentrifugation setups (Speers and Wu, 2007).

However, such "membrane fractions" obtained by sedimentation do not contain only the disrupted membrane fragments but are also heavily contaminated by major cellular proteins, cytoskeletal proteins, mitochondrial proteins, proteins associated with membranes and IMPs, or ribosomes. This is due to both specific and non-specific interactions with components of plasma membrane (Vít *et al*., 2016). Highly abundant proteins can easily shadow less abundant membrane proteins and thus hamper their MS analysis. Therefore, additional purification steps are usually adopted. A popular choice is to wash membrane fractions with aqueous high pH sodium carbonate solution (Fujiki, 1982). The high pH together with high ionic strength of sodium carbonate aids the disruption of membrane material including artificially formed membrane vesicles containing entrapped unwanted proteins. Additionally, the carbonate washes are sometimes coupled with other high salt buffers such as NaCl or KCl to aid the membrane purification (Dormeyer *et al*., 2008).

The membrane enrichment can be further complemented with various capturing methods utilizing specific IMPs characteristics such as protein modifications, the most commonly N-glycosylation, via lectins or hydrazide resin. These techniques will be discussed in detail in chapter **2.4.2.2 "Divide and conquer" strategies**. However,

regardless of the amount of purification steps taken, the contamination by non-membrane proteins is only decreased but never completely eliminated.

2.4.1.2 Protein solubilization

A critical step of the IMPs analysis workflow is the solubilization of the enriched membrane fraction resulting into extraction of protein content into an aqueous solution. The solubilization is particularly important because a part of an IMP molecule is often hidden within a plasma membrane and thus inaccessible to protease activity. The membrane environment therefore prevents digestion of the transmembrane segment, hence the disruption is essential. Various detergents, organic solvents and chaotropes are used in this regard.

2.4.1.2.1 Detergents

Detergents represent the most widely adopted reagents employed in protein solubilization. This is the result of their amphipathic nature that in aqueous solutions leads into mimicking physico-chemical properties of membrane phospholipids, thus enabling either partial or complete membrane disruption followed by IMPs solubilization.

Two major categories of detergents are defined. Ionic detergents contain either positively or negatively charged hydrophilic domain. They are very efficient in membrane disintegration and they also almost always denature the proteins to some extent. Nonionic detergents, on the other hand, embody uncharged hydrophilic head-groups and are considered to be significantly milder than ionic detergents and rather non-denaturing. The choice of the detergent depends on the type of analysis, type of sample, desirability of denaturation or other detergent-specific feature (Seddon *et al*., 2004).

2.4.1.2.1.a Sodium dodecyl sulfate

Sodium dodecyl sulfate (SDS) is a strong anionic detergent frequently used in proteomic applications. It is highly efficient in both membrane solubilization as well as protein denaturation. Due to these properties SDS has become a benchmark of protein solubilization. However, it is noteworthy that SDS (and also many other ionic detergents) is not compatible with several downstream stages of proteomic workflow. First, trypsin proteolytic activity is limited in even 0.1% SDS. Furthermore, SDS interferes with liquid chromatography and reduces its separation capacity. It is also incompatible

with MS detection as it negatively affects the peptide ionization process (Bosserhoff *et al*., 1989; Loo *et al*., 1994)

Several SDS depletion methods had been developed. These include procedures such as precipitation with potassium chloride (KCl) (Zhou *et al*., 2012) or with organic solvents (Cox *et al*., 2008; Doucette *et al*., 2014). Elimination of SDS using strong-cation exchange chromatography was also successfully employed (Sun *et al*., 2012)

Later on, Wisniewski *et al*. (2009) developed an alternative method of proteomic sample preparation suitable for detergent removal. The procedure, known as filter-aided sample preparation (FASP) allows for effective depletion of any low molecular weight compounds, including detergents, by centrifugation through ultrafiltration units. To briefly describe the principle, the membrane enriched fraction is first solubilized in a buffer containing SDS and high concentration of urea. The mixture is then placed into an ultrafiltration cartridge with a filter with defined high-molecular-weight (10-30 kDa) cut-off. The centrifugation removes the SDS by washing with urea. Urea is later removed in the similar manner by a buffer compatible with trypsin digestion. All follow-up steps, including reduction, alkylation and "on-filter" digestion, take place on the ultrafiltration unit before the digested peptides are finally released from the cartridge by centrifugation.

The efficiency of FASP in membrane proteomics was clearly showed in the study that successfully identified more than 1000 membrane proteins from murine hippocampus (Wiśniewski *et al*., 2009). FASP was also successfully applied in the analysis of membrane microdomains of renal cell carcinoma (Raimondo *et al*., 2015). More recently, FASP facilitated identification of more than 400 IMPs from the membrane of red blood cells (Bryk and Wiśniewski, 2017).

Due its high popularity, ready availability and ease of use the method has been heavily used and modified. One of such modifications is "N-glyco-FASP" utilizing lectins to selectively capture N-glycoproteins (Zielinska *et al*., 2010). The details of N-glyco-FASP will be described in chapter **2.4.2.2.1b Lectin affinity capture and N-glyco-FASP**.

2.4.1.2.1.b Sodium deoxycholate

Bile acid salts, such as sodium deoxycholate (SDC), are also anionic detergents. However, they significantly differ from the linear-chain detergents, such as SDS. The molecule is steroidal with polar and non-polar part. The denaturing and solubilization

capabilities are slightly lower compared to SDS (Seddon *et al*., 2004). However, SDC is tolerated by trypsin in up to 5-10% concentration (Masuda *et al*., 2008). The MS analysis, nonetheless, still requires SDC depletion.

A widely adopted SDC removal procedure is a phase transfer of SDC into ethyl acetate, diethyl ether, or other water-immiscible organic solvent (Masuda *et al*., 2008). After the cell lysis and trypsin digestion, the sample is acidified and mixed together with a water-immiscible organic solvent. The low pH of the environment then precipitates SDC into the organic phase while membrane proteins remain within the water phase. The original protocol led to identification of total of 1450 proteins of which 764 (53%) were membrane proteins. It is also noteworthy that unwanted losses of the sample when using SDC together with phase transfer are minimal (Masuda *et al*., 2008).

Several alternative methods for SDC depletion were developed employing acid precipitation (Lin *et al*., 2008), gel centrifugation (Horigome and Sugano, 1984) or oligosaccharide-based beads (Antharavally *et al*., 2011).

2.4.1.2.1.c Acid-labile surfactants

A unique answer to the necessity of laborious and time-consuming detergent removal are acid-labile surfactants (ALS). The molecule of ALS contains an acid-labile functional group that is located in-between hydrophilic and hydrophobic ends of the detergent. When the molecule is subjected to a low pH environment the functional group is cleaved. The hydrophobic part forms a water-immiscible precipitate that can be easily removed by phase transfer or centrifugation (Yu *et al*., 2003).

One of the first commercially available ALS was RapiGest. RapiGest structurally mimics SDS by being consisted of an ionic moiety and a hydrophobic alkyl chain. It is efficient in protein solubilization and does not inhibit trypsin protease activity when present in up to 1% concentrations (Yu *et al*., 2003; Chen *et al*., 2007). Study from 2011 performed on *E. Coli* shows that RapiGest-assisted protein solubilization is an efficient alternative to SDS. The method led to discovery of more than 1600 proteins of which almost 54% were predicted to be IMPs (Wu *et al*., 2011).

On the other hand, it has been reported that the hydrophobic fragment of RapiGest may spontaneously precipitate with the most hydrophobic peptides in the sample mixture and thus hinder the analysis (Yu *et al*., 2003).

As much as RapiGest simplifies the protein sample preparation and decreases the overall time requirements, it must be noted that its high price counterbalances its advantages.

2.4.1.2.2 Organic Solvents

Organic solvents (acetonitrile, methanol) are an alternative to detergents to aid membrane solubilization and facilitate the digestion. Unlike most detergents they are compatible with downstream LC-MS analysis and can be easily removed by evaporation if needed. It is noteworthy, that trypsin maintains only partial protease activity when present in higher concentrations of organic solvents (Simon *et al*., 1998, 2001; Russell *et al*., 2001). Blonder *et al*. (2004) reports that trypsin retains only 20% of its protease activity in 60% methanol compared to aqueous buffer.

Concentrated formic acid (FA) is an effective solubilization reagent but it is incompatible with trypsin (it lowers the pH of the environment bellow trypsin working pH range) and leads into extensive formylation of the sample (Loo and Loo, 2007). Fortunately, formylation can be minimized by working at low temperatures (Doucette *et al*., 2014).

Several membrane proteomic studies have found the use of formic acid to be efficient and yielding satisfying results. The most common application of FA is membrane proteome solubilization followed by CNBr protein digestion (Washburn *et al*., 2001; Da Cruz *et al*., 2003; Blackler *et al*., 2008)

However, recent data clearly shows that novel detergent removal methods, such as FASP, greatly surpass benefits of using organic solvents (Waas *et al*., 2014) in membrane solubilization.

2.4.1.2.3 Chaotropes

Chaotropes, such as urea, are also often used in membrane proteomics. This is due to its ability to disrupt hydrophobic interactions and for its strong denaturing effect. This, however, requires high concentrations of used chaotropic agent (8 M urea) that has to be lowered prior the trypsin digestion (2 M for urea). 2 M urea is sufficient enough to facilitate digestion in otherwise inaccessible cleavage sites (Wiśniewski *et al*., 2009). Similarly to organic solvents, chaotropes do not interfere with LC-MS analysis to such extent as detergents do. A simple desalting procedure is sufficient enough for removal of the chaotrope (Speers and Wu, 2007).

2.4.2 STRATEGIES EMPLOYED IN ANALYSIS OF MEMBRANE PROTEOME

Generally, two distinct strategies of membrane proteomics exist. The "classical" approach targets membrane proteins as whole molecules. Alternatively, a "divide & conquer" strategy can be used to aim separately at either hydrophilic (extra-membrane) or hydrophobic (transmembrane) segments of IMPs.

2.4.2.1 The "classical" approach

The "classical" approach workflow usually follows the pattern of general proteomic analysis described in chapter **2.2** but it also utilizes membrane proteomic tools mentioned in chapter **2.4.1**, namely the membrane material enrichment and solubilization using detergents.

Three discoveries increased efficiency of the "classical" approach – introduction of FASP as a cheap and efficient detergent removal technique (Wisniewski *et al*., 2009), employment of SDC and its subsequent removal by phase transfer (Masuda *et al*., 2008) and introduction of pre-digestion by Lys-C before final digestion with trypsin (Wisniewski *et al*., 2009).

However, the "classical" approach has its drawbacks. These drawbacks are associated with the specific properties of IMPs – their amphipathy, lack of trypsin cleavage sites and, in general, low abundance of IMPs (chapter **2.3 Biological membranes and proteomics of integral membrane proteins**).

2.4.2.2 "Divide & conquer" strategies

While "classical" approach has to cope with the challenges associated with analysis of intact IMPs, the "divide & conquer" strategy more or less circumvents these complications by specifically aiming at either hydrophilic extra-membrane segments or hydrophobic transmembrane helices of IMPs.

2.4.2.2.1 Strategies aimed for hydrophilic segment analysis

Several strategies targeting extra-membrane segments of IMPs have been developed. N-glycoprotein capture methods are the most notable. N-glycosylation is a prominent protein post-translational modification. It has been estimated that up to 50% of all proteins are glycosylated (Apweiler *et al*., 1999) with vast majority of membrane (and in general surface proteins) proteins being glycoproteins (Chandler

and Costello, 2016). N-glycosylation plays role in a vast amount of cellular functions including cell-cell interactions, cell signaling and communication, immune response or apoptosis (Fernandes *et al*., 2000; Kaszuba *et al*., 2015). The N-glycosylation occurs at asparagine residue in a sequence motif of Asn-X-Ser/Thr where X denotes any amino acid except proline (Tian *et al*., 2007).

The methods aiming at hydrophilic segments utilize the fact that majority of membrane proteins are glycosylated and specifically capture these glycosylated IMPs or their glycopeptides.

2.4.2.2.1a Solid-phase extraction of N-glycopeptides

One of the methods of choice for the analysis of glycosylated hydrophilic IMPs segments is solid-phase extraction of N-glycopeptides (SPEG; Zhang *et al*., 2003; Tian *et al*., 2007). SPEG utilizes affinity capture of N-glycopeptides onto hydrazide-coated beads followed by removal of unbound peptides by extensive washing and subsequent release of the bound N-glycopeptides using PNGase F.

In more detail, the enriched fraction of membrane proteins is first digested with a protease resulting in a mixture of both glycosylated and non-glycosylated peptides. This mixture is then subjected to oxidation with a strong oxidizing agent such as sodium periodate. The oxidative reaction converts carbohydrate cis-diol groups of glycopeptides into aldehydes that can further form hydrazone bonds with hydrazide-coated beads (**Figure 2.6**) while non-glycosylated peptides remain unbound and can be freely washed away. The washing results in up to 90% enrichment of glycosylated peptides (Tian *et al*., 2007). After several steps of extensive washing, the beads are treated with PNGase F. PNGase cleaves the linkage between the N-glycosylated peptide and the oligosaccharide chain releasing the peptide from the hydrazide support (Zhang *et al*., 2004; Tian *et al*., 2007). This cleavage results in conversion of asparagine (Asn) to aspartate (Asp) with a characteristic decrease in molecular weight of 0.98 Da (Yang and Zubarev, 2010). This mass change can serve as a marker of former N-glycosylation (Tian *et al*., 2007).

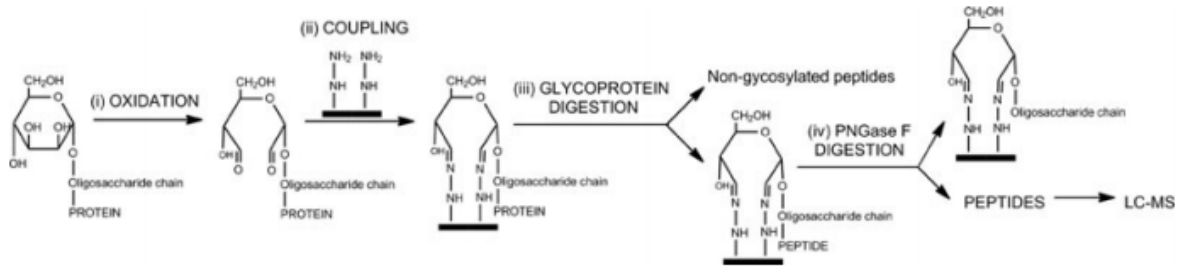

Figure 2.6: The basic principle of hydrazide chemistry used in SPEG protocol. Adopted from Ongay *et al*. (2012)

SPEG was successfully used in the proteome analysis of cardiac tissue and resulted in identification of 694 glycoproteins of which 534 (77%) were IMPs (Tian *et al*., 2014). Another study of skin carcinomas led to identification of 318 glycoproteins (Tian *et al*., 2008).

2.4.2.2.1b Lectin affinity capture and N-glyco-FASP

An alternative approach to selectively study glycosylated extra-membrane segments of IMPs is to use lectin affinity capture technology. Lectins are carbohydrate-binding proteins that specifically bind to various glycan groups like mannose, N-acetyl-glucosamine (GlcNAc), sialic acid or galactose, hence allowing capture of glycosylated proteins from complex protein mixtures. For instance, Concanavalin A (ConA), a *Canavalia ensiformis* lectin, predominantly binds to mannose (Becker *et al*., 1974). Wheat germ agglutinin (WGA) interacts with GlcNAc and sialic acid (Bakry *et al*., 1991). *Ricinus communis* agglutinin (RCA_{120}) recognizes galactose (Bakry *et al*., 1991). A single lectin or multiple lectin affinity capture protocols can be used to isolate glycoproteins from a complex biological sample (Bunkenborg *et al*., 2004; Drake *et al*., 2011)

The benefits of lectin affinity capture coupled with benefits of FASP were utilized in glyco-FASP protocol introduced by Zielinska *et al*. (2010). In this method, membrane proteins are first solubilized in SDS and then reduced, alkylated and digested on the ultrafiltration device (FASP). The digested peptides are then mixed with lectin solution on the ultrafiltration cartridge. The N-glycosylated peptides bind to relevant lectins while unbound non-glycosylated peptides are washed away by centrifugation. PNGase F is then used to release N-glycopeptides from the lectins.

The original protocol led to identification of more than 2300 of glycosylated proteins while, according to gene ontology, 31% of those proteins were localized within

plasma membrane (Zielinska *et al*., 2010). Later, a study of B-cell lymphoma led to identification of over 900 IMPs (Deeb *et al*., 2013).

2.4.2.2.1c Cell surface capture

Glycosylated hydrophilic IMP segments can also be targeted by cell surface capture (CSC) method (Wollscheid *et al*., 2009), by tagging the proteins on the surface of intact cells with biotin-based tag and subsequent capture of these proteins after cell lysis on streptavidin-coated beads.

In greater detail, the cells are first harvested from the cell culture. Surface glycoproteins are then oxidized with an oxidizing agent such as sodium periodate which again leads into formation of aldehydic groups from cis-diol groups. The N-glycoproteins are tagged with biocytin hydrazide that interacts with newly-formed aldehydic groups. Cells are lysed and proteins are digested into peptides by combination of Lys-C and trypsin. The N-glycopeptides tagged with biocytin hydrazide are subsequently captured on streptavidin-coated beads, washed, eluted using PNGase F and subjected to LC-MS analysis.

Although CSC was successfully employed in several membrane proteomic studies (Wollscheid *et al*., 2009; Hofmann *et al*., 2010; Moest *et al*., 2013), the laboriousness and requirement for living cells prevents the wide adoption of the method.

2.4.2.2.2 Strategies targeting hydrophobic segments

In contrast to hydrophilic, easily accessible extra-membrane segments if IMPs, the hydrophobic transmembrane alpha helical segments have been neglected by the proteomics community, mostly due to their adverse physico-chemical properties and the lack of trypsin cleavage sites.

In 2008, Blackler *et al*. presented a method (hppK-CNBr = high pH, proteinase K, CNBr) that partially overcame these issues. In their procedure, the membrane samples were first washed with sodium carbonate solution at high pH and low temperature, resulting in opening membrane vesicles and forming so called membrane "sheets". These membrane sheets were then treated with proteinase K which non-specifically cleaves the hydrophilic segments of IMPs as well as all contaminating non-membrane protein. This resulted in "shaved" membrane sheets containing only hydrophobic segments. The sheets were then solubilized in 90% formic acid and re-digested with CNBr prior the LC-MS identification (**Figure 2.7**).

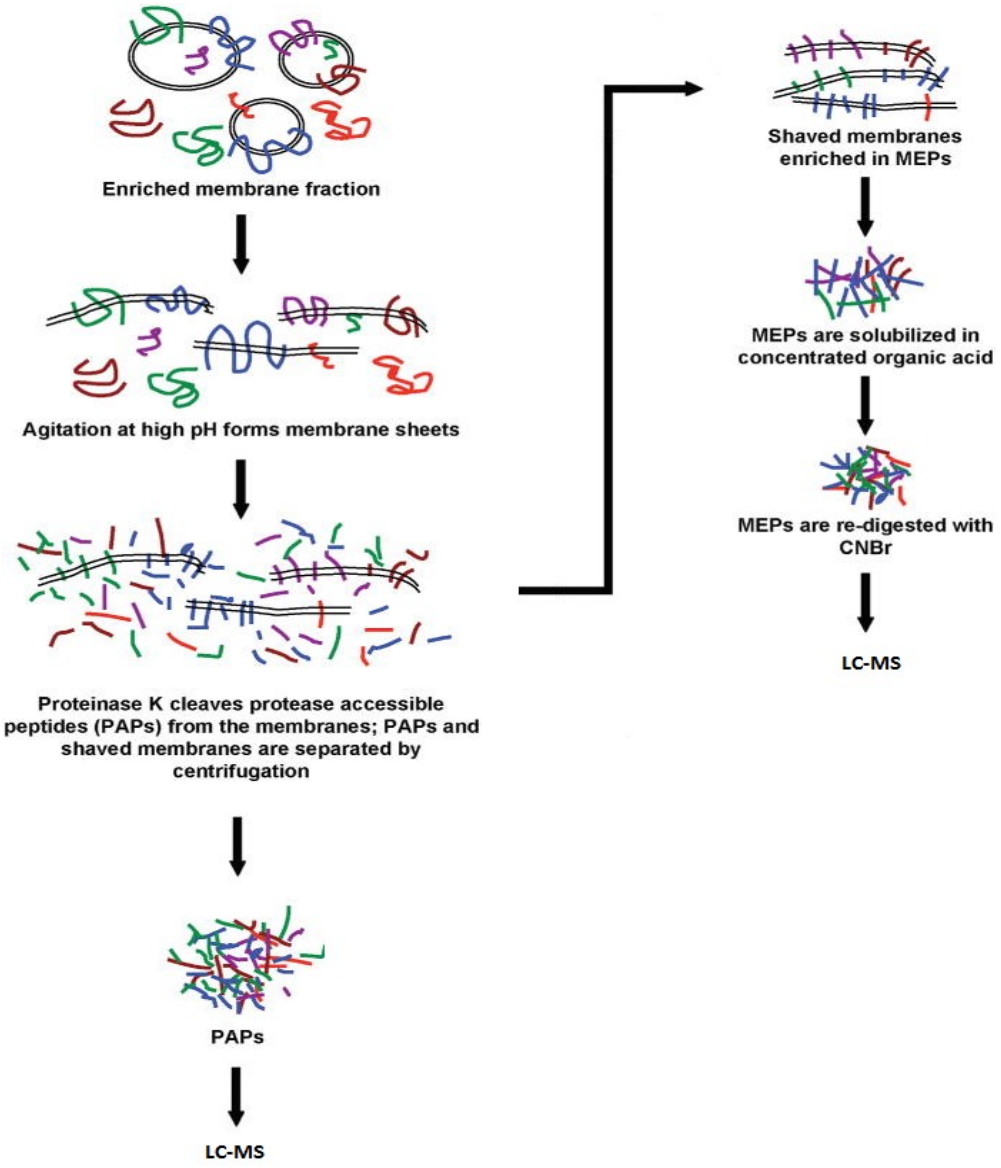

Figure 2.7: Schematic representation of the hppK-CNBr protocol (adapted from Blackler *et al*., 2008)

The digestion with non-specific proteinase K, however, results in formation of numerous overlapping peptides. Such peptides, first, increase the sample complexity rendering the IMP identification less efficient. Furthermore, peptides generated in this manner hinder quantitative analysis. To overcome these issues, our laboratory has modified the original protocol by replacing proteinase K with trypsin and by delipidation of CNBr-cleaved peptides using dichloromethane prior MS analysis. The modified and simplified protocol – named hpTC (= high pH, trypsin, CNBr) enabled identification of 1,224 proteins in human lymphoma cells of which 802 (65.5%) were IMPs (Vit *et al*., 2016).

The establishment of FASP as a cheap and efficient method for detergent removal, introduction of SDC and its removal by phase transfer and usage of Lys-C together with trypsin significantly improved the membrane proteome analysis in the "classical" approach. However, the problems associated with analysis of IMPs (low abundance, low solubility and lack of trypsin cleavage sites) in the "classical" manner led to the development of the "divide & conquer" strategies that overcome these issues by selectively targeting either hydrophilic segments (N-glyco-capture methods) or hydrophobic segments (hppK-CNBr, hpTC). Moreover, very recent work of our laboratory shows that a combination of both "classical" and "divide & conquer" approaches into one big "divide, conquer & combine" strategy provides even better insight into the membrane proteome (Vít *et al*., 2019). We therefore decided to use this combined strategy for analysis of myocardial membrane proteome to get as detailed description of membrane proteome as possible. Detailed knowledge of cardiac membrane proteome is essential for the description of molecular mechanisms governing the heart physiology and pathology.

2.5 CARDIAC TISSUE AND CARDIOMYOCYTES

Heart is a highly organized tissue that consists mainly of cardiomyocytes, pacemaker cells, Purkinje fibers, supportive cardiac fibroblast cells and adjacent vasculature and functions as a muscular pump that a) collect blood from all tissues and transports it to lungs and b) collects blood from lungs and transports it to all tissues of a body. The pumping mechanism is a result of contraction and relaxation of individual cardiomyocytes that is being regulated by cyclic increase and decrease in levels of intracellular Ca^{2+} by number of ion channels and transporters that control import and export of Ca^{2+} in and out of the cell and sarcoplasmic reticulum (Williams *et al*., 1992; Marks, 2003).

Cardiomyocytes account for approximately 35% of the cells in the heart and constitute for up to 70% of its volume (Nag and Zak, 1979). Cardiomyocytes are relatively small, 10-25 μm in diameter and 50-100 μm in length, elongated cylindrical cells containing rod-like structures called myofibrils that are responsible for cardiac muscle contraction. Myofibrils are composed of actin filaments, myosin filaments and associated proteins such as titin, tropomyosin and troponins that aid the contraction. Each myofibril is surrounded by network of tubules of sarcoplasmic reticulum (SR). SR is in close interaction with plasma membrane so that it can react to depolarization of the membrane by release of Ca^{2+} ions into the cytoplasm and triggering myofibril contraction (Williams *et al*., 1992; Marks, 2003). The contraction is then relaxed by re-uptake of Ca^{2+} from the cytosol by sarco-endoplasmic reticulum Ca^{2+} transporter (SERCA) or by removal by Na+/Ca2+ exchanger (Bers *et al*., 2003; Hilgemann, 2004).

In order for the contraction to keep happening, cardiomyocytes have to possess enough energy. The energy is generated by large amount of mitochondria present within each cell. Mitochondria, in fact, comprise for over 30% of the cardiomyocyte volume (Piquereau *et al*., 2013). Mitochondria are localized into three main regions – lined around myofibrils, close to plasma membrane and around the nucleus. This is an ideal localization for the distribution of energy for contraction (myofibrillar mitochondria), transcription and translation (perinuclear mitochondria) and for ion homeostasis and signaling pathways (mitochondria localized close to plasma membrane) (Piquereau *et al*., 2013).

Cardiomyocytes are joined together by structures called intercalated discs located at blunt ends of the cell, forming cardiac muscle fibers. The intercalated discs comprise of three distinct cell junction structures, each having their own functions.

The *fascia adherens* and desmosomes are responsible for connecting the cardiomyocytes to each other through interaction with either cytoskeletal proteins in case of desmosomes or with myofibrillar proteins in case of *fascia adherens*. The gap junctions form clusters of channels spanning the entirety of the membrane enabling effective conduction of action potentials (Estigoy *et al*., 2009). The connection between each cardiomyocyte results in coordinated and synchronous spread of activation and contractile force from one cell to another, leading to a heartbeat.

The complex and fully-functional transfer of information, action potential and contractive force between each cardiomyocyte of the cardiac muscle fiber is essential for a heart contraction. Therefore, disruption of any mechanism playing a role in this machinery can potentially lead to pathophysiological states.

2.6 CARDIOVASCULAR DISEASES

Cardiovascular diseases (CVD) are the leading cause of death in Europe and account for 45% of all deaths. It has been estimated that over 3.9 million people die of CVDs annually in Europe. The ischaemic heart disease (IHD) is the main form of CVDs and alone accounts for more than 1.73 million deaths per year in Europe. Other forms of CVD include hypertensive heart disease, rheumatic heart disease, congenital heart disease and cardiomyopathies. The 2015 incidence data report that 11.3 million of new cases of CVD appeared in Europe of which over 5.7 million were IHD. The data mapping the incidence between years 1990 and 2015 also showed that the incidence is increasing. The prevalence data showed that over 85 million people across Europe were suffering from CVD in 2015. The major contributors to this number were IHD (30 million cases) and peripheral heart disease (over 36 million cases) (Wilkins *et al*., 2017).

In addition to CVDs, chronic heart failure (CHF) is another major cause of death worldwide with increasing prevalence that ranges between 1-2% in adult population and is rising to over 8% among people with >75 years of age (Mosterd and Hoes, 2007). The prognosis of patients with CHF is unfavourable and more than 40% of them die within 2.5 years of the initial diagnosis despite therapeutic intervention (Pocock *et al*., 2012).

These alarming numbers show that there is an urgent need for better diagnostic, prognostic and predictive markers as well as a need for new and efficient drugs. For this reason mapping of cardiac membrane proteome is of high interest as IMPs are in general considered to be attractive drug targets. This is best demonstrated on the fact that up to 60% of currently approved drugs target IMPs (Yildirim *et al*., 2007).

3. AIMS OF THE THESIS

The goals of this thesis were:

- **To uncover as large as possible portion of the rat myocardial membrane proteome employing four different approaches targeting IMPs of different physico-chemical properties, using:**
  - The "classical" approach, represented by the standard "detergent+trypsin" workflow
  - The hpTC method that targets hydrophobic segments of IMPs by removal of their trypsin-cleavable portions and subsequent re-digestion of the membrane-protected parts of IMPs with cyanogen bromide
  - Two N-glycopeptide capture methods that should predominantly target the surface and luminal N-glycosylated IMPs – solid phase extraction of N-linked glycopeptides (SPEG) and N-glyco-FASP
  - To perform bioinformatics assessment of the results
  - To determine if the used methods are complementary to each other

- **To evaluate the impact of used starting material – a total tissue lysate or membrane-enriched fraction – on the total number of identified IMPs by SPEG and N-glyco-FASP.**

- **To compare two different methods of preparation of membrane-enriched fraction by western blot analysis of:**
  - Membrane proteins
  - Cytoskeletal proteins
  - Soluble cytosolic proteins

4. MATERIAL AND METHODS

4.1 MATERIAL AND EQUIPMENT

4.1.1 BIOLOGICAL MATERIAL

- samples of right ventricle from male Wistar rats (*Rattus Norvegicus*)

4.1.2 CHEMICALS

- all used chemicals were purchased at Sigma unless specified
- 2-mercaptoethanol
- 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)
- acetonitrile (ACN)
- ammonium bicarbonate $(NH_4)HCO_3$
- bicinchoninic acid
- bovine serum albumin (BSA)
- bromphenol blue (Bio-rad)
- calcium chloride ($CaCl_2$)
- copper(II) sulfate pentahydrate ($CuSO_4 \cdot 5H_2O$)
- cyanogen bromide (CNBr)
- deionized water
- dithiothreitol (DTT; Promega)
- ethylacetate
- ethylenediaminetetraacetic acid (EDTA)
- formic acid (FA)
- glycine
- iodoacetamide (IAA; Bio-rad)
- isopropyl alcohol
- liquid nitrogen (Siad)
- magnesium chloride ($MgCl_2$)
- manganese(II) chloride ($MnCl_2$)
- methanol
- non-fat dried milk powder (Laktino)
- sodium carbonate (Na_2CO_3)
- sodium chloride
- sodium deoxycholate (SDC)

- sodium periodate ($NaIO_4$)
- sodium thiosulfate ($Na_2S_2O_3$)
- sodium-dodecyl sulfate (SDS)
- trifluoroacetic acid (TFA)
- tris(hydroxymethyl)aminomethane (Tris)
- tris(hydroxymethyl)aminomethane hydrochloride (Tris HCl)
- Tween® 20

4.1.3 SOLUTIONS

- 10x TGS buffer (Bio-rad)
- 5% (w/v) SDC solution in 100 mM $(NH_4)HCO_3$
- 500 mM DTT in 100 mM $(NH_4)HCO_3$
- 500 mM IAA in 100 mM $(NH_4)HCO_3$
- 5x Laemmli buffer (10% SDS, 50% glycerol, 0.3 M tris-HCl, pH 6.8, bromophenol blue)
- 80% ACN in 0.1% TFA
- aqueous solution of 100 mM Na_2CO_3 and 1 mM EDTA
- binding buffer (1 mM $CaCl_2$, 1 mM $MnCl_2$, 0.5 M NaCl in 20 mM tris-HCl, pH 7.6)
- homogenization buffer (2 M NaCl, 1 mM EDTA, 10 mM HEPES-NaOH, pH 7.4)
- HPLC mobile phase buffer A (2% ACN, 0.1% FA in water)
- HPLC mobile phase buffer B (80% ACN, 0.1% FA in water)
- hypotonic lysis buffer (10 mM HEPES-NaOH pH 7.4, 10 mM NaCl, 2 mM $MgCl_2$)
- PBS buffer pH 7.4
- PBST buffer (PBS buffer, 0.1% Tween 20)
- solvent A (1% aqueous formic acid)
- solvent B (80% ACN, 10% isopropyl alcohol, 1% FA)
- Towbin buffer (25 mM tris, 192 mM glycine, 10% methanol, pH 8.3)

4.1.4 ENZYMES

- pancreatic bovine deoxyribonuclease I
- PNGase F (Roche)
- sequencing grade modified trypsin (Promega)

4.1.5 COMMERCIAL SOLUTIONS AND KITS

- Affi-Gel Hz 10x Coupling Buffer Concentrate (Bio-rad)
- Affi-Gel Hz Hydrazide Gel (Bio-rad)
- ImmunoCruz Luminol reagent (Santa Cruz)
- Pierce Quantitative Colorimetric Peptide Assay kit (Thermo Fisher Scientific)
- Mini-PROTEAN TGX stain-free precast gels (Bio-rad)

4.1.6 LECTINS

- Wheat Germ Agglutinin (WGA)
- Concanavalin A (ConA) from *Canavalia ensiformis*
- RCA_{120} from *Ricinus communis*

4.1.7 ANTIBODIES

4.1.7.1 Primary antibodies

- rabbit anti-Transferrin receptor (ab84036, Abcam)
- mouse anti-Caveolin (610057, BD Transduction Laboratiories)
- rabbit anti-Connexin 43 (C6219, Sigma)
- rabbit anti-Ferritin (ab75973, Abcam)
- goat anti-Actin (sc-1616, Santa Cruz)
- mouse anti-α-Tubulin (F2168, Sigma)
- rabbit GAPDH anti-GAPDH (G9545, Sigma)

4.1.7.2 Secondary antibodies

- donkey anti-goat HRP (sc-2020, Santa Cruz)
- goat anti-mouse HRP (sc-2005, Santa Cruz)
- donkey anti-rabbit HRP (711-035-752, Jackson Immuno Research)

4.1.8 PROTEIN STANDARD

- Precision Plus protein unstained standards (Bio-rad; **Figure 4.1**)

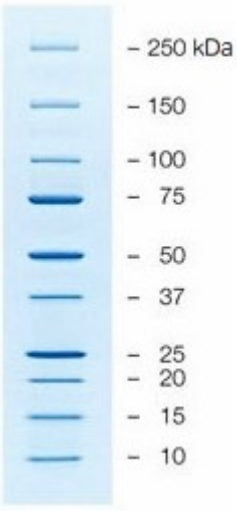

Figure 4.1: Bio-rad protein standard with range of 10-250 kDa and with three reference bands at 25, 50 and 75 kDa

4.1.9 EQUIPMENT

- 3D Rotating Mixer (Miulab)
- centrifuge Eppendorf 5810 R
- ChemiDoc MP Imaging system with Image Lab 5 software (Bio-rad)
- Concentrator Plus (Eppendorf)
- gel blotting paper (Whatman)
- HPLC-MS with reversed phase nano column (EASY-Spray column, 50 cm length, 75 µm internal diameter, PepMap C18, 2 µm particle size, 100 Å pore size) and trap column (Acclaim PepMap300 C18, 5 mm length, 300 µm internal diameter, 5 µm particle size, 300 Å pore size)
- hypodermic needles (26G×1 - 0.45×25 mm; 20G×1½ - 0.90×40 mm)
- Imunno-Blot PVDF membrane (Bio-Rad)
- incubator MCO-170 AICUV-PE (Panasonic)
- Mini-PROTEAN tetra cell system (Bio-rad)
- Nanodrop 2000 (Thermo Scientific) and associated software
- Opti-Trap Column Cartidges with manual holder (Optimize Technologies)
- Orbitrap Fusion mass spectrometer (Q-OT-qIT, Thermo)
- PANPEHA pH indicator strips (Whatman)

- plastic bag heat sealer (ETA)
- PowerPac HC power supply (Bio-rad)
- probe sonicator Q125 (QSonica)
- shaker MR-12 (Biosan)
- shaker Multi Bio 3D (Biosan)
- thermoblock (Bioer)
- Trans-Blot Turbo (Bio-rad)
- Ultracel YM-10 filtration unit (Merck)

4.2 METHODS

4.2.1 SAMPLE PREPARATION

4.2.1.1 Model organism

Male Wistar rats (*Rattus norvegicus)* were chosen for analysis. The animal model was humanely killed and the heart was surgically removed. Both right and left ventricle were separated from the rest of the heart tissue and were stored in -80 °C.

4.2.1.2 Tissue homogenization

The frozen right ventricle tissue sample was mechanically processed into fine powder in a cold mortar with pestle while maintaining the frozen state of the tissue with liquid nitrogen. A portion of the heart tissue powder was transferred to a clean microtube and used for preparation of membrane fraction (**4.2.1.3**). Another part of the tissue powder was immediately lysed with sodium deoxycholate (SDC) (**4.2.1.4**). The amount of the heart powder used was ranging between 0.03 - 0.1 g. The remaining heart powder was stored in -80 °C.

4.2.1.3 Isolation of crude membrane fraction

The crude membrane fraction was isolated according to protocol developed by Nielsen *et al*. (2005).

The sample (ca 100 mg of the tissue homogenate per replicate) was resuspended with 1 ml of homogenization buffer (2 M NaCl, 1 mM EDTA, 10 mM HEPES-NaOH, pH 7.4), then briefly homogenized by passing 20 times through a gauge of a hypodermic needle followed by thorough homogenization with probe sonicator (30-35% amplitude, 6 cycles of 15 s pulses and 15 s pauses).

The sample was then centrifuged in a benchtop centrifuge at 20,000 × g at 4 °C for 30 minutes. The pellet was resuspended in an aqueous solution of 100 mM Na_2CO_3 and 1 mM EDTA and agitated on ice for 30 minutes. The sample was again centrifuged, the supernatant was discarded and the washing with 100 mM Na_2CO_3 and 1 mM EDTA was repeated one more time.

4.2.1.4 Tissue lysis in sodium deoxycholate

Lysis in SDC was done according to the original method developed by Masuda *et al*. (2008).

The membrane pellet was lysed with 0.5 ml of 5% SDC in 100 mM $(NH_4)HCO_3$ for 10 minutes at room temperature without agitation. The sample was then sonicated (amplitude 20-25%) and centrifuged at 15.000 × g at 4 °C for 15 minutes. The resulting supernatant was transferred to a new microtube and protein concentration was measured.

4.2.1.5 Protein concentration measurement

The protein concentration was determined using standard BCA assay. The working reagent containing bicinchoninic acid and $CuSO_4 . 5H_2O$ was prepared in ratio 50:1.

18 µl of working reagent was mixed with 2 µl of BSA in 0.2, 0.4, 0.6, 0.8, 1 mg/ml concentrations in triplicates. Standards were incubated at 37 °C for 30 minutes. The absorbance was measured at 562 nm on Nanodrop 2000 and standard curve was generated.

18 µl of the working reagent was mixed with 2 µl of the sample (non-diluted or 10x diluted). Samples were incubated at 37 °C for 30 minutes. The absorbance was measured at 562 nm on Nanodrop 2000. The concentration was determined from the calibration curve. The BCA assay was performed in triplicates.

4.2.1.6 Protein reduction, alkylation and digestion

An amount of samples corresponding to 0.5 – 1 mg of protein was used. The reduction was performed by addition of 500 mM DTT in 100 mM $(NH_4)HCO_3$ to a final concentration of 10 mM. The samples were incubated in thermoblock at 37 °C for 30 minutes with agitation.

The samples were treated with 500 mM IAA in 100 mM $(NH_4)HCO_3$ to a final concentration of 25 mM. The incubation was carried at room temperature in the dark for 30 minutes with agitation.

The samples were then diluted with 100 mM $(NH_4)HCO_3$ to a final concentration of 1% (w/v) SDC. The sequencing grade modified trypsin was added to the samples in 1:50 enzyme to protein ratio. The incubation was carried overnight at 37 °C.

4.2.1.7 Phase transfer – detergent removal

After the incubation, the samples were acidified with 10% TFA to a pH <3. The samples were mixed with ethyl acetate in 1:1 ratio, vigorously agitated for 1 minute and then centrifuged at benchtop centrifuge at 15.000 × g for 2 minutes to divide the sample into the aqueous and the ethyl acetate phase. The ethyl acetate phase was removed while the aqueous phase remained in the tube. The pH was measured again, adjusted to <3, the process was repeated two more times.

4.2.1.8 Peptide desalting

The peptide samples were purified using OptiTrap desalting column according to manufacturer instructions. The column was first washed with 500 µl of 80% ACN in 0.1% TFA solution and then equilibrated in 500 µl of aqueous 0.1% TFA. The sample was acidified with 10% TFA to pH <3, loaded into OptiTrap column. The OptiTrap was washed with 500 µl of 0.1% TFA. The sample was then eluted to a new microtube with 200 µl of 80% ACN in 0.1% TFA and concentrated using a vacuum concentrator.

4.2.1.9 Peptide concentration measurement

The peptide concentration was determined with Pierce Quantitative Colorimetric Peptide Assay Kit following the manufacturer instructions.

The working reagent was prepared by mixing together 50 parts of Reagent A, 48 parts of Reagent B and 2 parts of reagent C. The peptide standard (BCA tryptic peptide mixture from the kit) was diluted to contain 0.2, 0.4, 0.6, 0.8, 1 mg/ml of peptides in triplicates. 18 µl of the working reagent and 2 µl of the standard were mixed in a fresh tube. The absorbance was measured at 480 nm and the calibration curve was generated.

18 µl of the working reagent and 2 µl of the sample were mixed in a fresh tube. The absorbance was measured at 480 nm. The standard curve was used to determine the peptide concentration of each sample. The assay was performed in duplicates.

4.2.2 N-GLYCOPROTEIN ENRICHMENT

Two distinct methods for N-linked glycoprotein isolation were used – modified solid-phase extraction of N-glycopeptides protocol based on Tian *et al*. (2007) and modified N-glyco-FASP protocol based on Zielinska *et al*. (2010).

4.2.2.1 Solid-phase extraction of N-glycopeptides

The Affi-Gel Hz Hydrazide Gel and Affi-Gel Hz 10x Coupling Buffer Concentrate (Bio-rad) were used in this step. First, the peptide samples (ca 100 µg) were diluted in Affi-Gel Hz Coupling Buffer and briefly sonicated in a bath sonicator. The peptide samples were then treated with 100 mM sodium periodate to a final concentration of 10 mM and incubated at room temperature in the dark for 1 hour with agitation. The reaction was stopped with 200 mM sodium thiosulfate added to a final concentration of 20 mM.

Subsequently, 25 µl of hydrazide resin (Affi-Gel Hz Hydrazide Gel, 50 µl of 50% slurry) per sample was prepared. The resin was briefly centrifuged (4000 × g, 45 s) and supernatant was removed. The hydrazide was washed once with 250 µl of deionized water and two times in Affi-Gel Coupling Buffer. The buffer was removed and the samples were pipetted onto the hydrazide support. The incubation was left overnight at room temperature with mild agitation.

The resin was washed two times with 250 µl of 100 mM $(NH_4)HCO_3$, once with 1.5 M NaCl and once with 80% ACN. The resin was resuspended in 25 µl of 100mM $(NH_4)HCO_3$, and 3 units of PNGase F were added to the mixture (1 U/µl). The de-glycosylation was left overnight at 37 °C with mild agitation.

After the incubation, the resin was centrifuged (4000 × g, 45 s) and supernatant containing N-glycopeptides was collected to a fresh tube. The hydrazide was washed again two times with 100 µl of 100 mM $(NH_4)HCO_3$. The washes were pooled with previously collected supernatant.

The mixture was desalted on OptiTrap (**4.2.1.8**) and concentrated within the Speedvac. The peptide concentration was determined with Pierce Quantitative Colorimetric Peptide Assay Kit (**4.2.1.9**).

The samples were dried until complete dryness and subsequently analyzed with LC-MS.

4.2.2.2 N-glyco-FASP

The peptide samples (approximately 100 µg each) were solubilized in 100 µl of binding buffer (1 mM $CaCl_2$, 1 mM $MnCl_2$, 0,5 M NaCl in 20 mM Tris HCl). Lectins were then added to the samples in following fashion: 100 µg WGA, 100 µg Concanavalin A and 80 µg of RCA_{120}. The mixture of the samples and lectins was

then transferred to YM-10 filtration cartridge and incubated for 1 hour at room temperature.

The unbound peptides were removed by centrifugation (14.000 × g, 10 min). The samples were washed four times with 200 µl of binding buffer. The buffer was added to the filtration unit, the units were centrifuged (14.000 × g, 15-30 min) and the filtrate was discarded. The samples were then washed two times with 200 µl of 40 mM $(NH_4)HCO_3$ in the same fashion.

The de-glycosylation was performed by addition of 40 µl of 40 mM $(NH_4)HCO_3$ and 3 µl of PNGase F onto the filter. The reaction was left overnight at 37 °C.

The samples were eluted into new microtubes via centrifugation (14.000 × g, 18 °C, 10-15 min). The filtration units were washed two times with 50 µl of 40 mM $(NH_4)HCO_3$ and once with 40 µl of 0.5 M NaCl in the manner mentioned above. The washes were pooled with previously collected supernatant.

The samples were acidified to pH <3, adjusted to 500 µl volume with 40 mM $(NH_4)HCO_3$ and briefly sonicated in a bath sonicator.

The samples were desalted on OptiTrap, concentrated and the peptide concentration was determined. The samples were then dried and analyzed with LC-MS.

4.2.3 HPTC

Approximately 0.05 g of previously prepared heart tissue powder was used per replicate. The heart powder was treated with 0.5 ml of hypotonic lysis buffer (10 mM HEPES pH 7.4, 10 mM NaCl, 2 mM $MgCl_2$). The lysis was carried on ice for 15 minutes with agitation. The samples were homogenized by passing through gauge of hypodermic needle followed by probe sonicator homogenization (amplitude 20-25%). The samples were centrifuged (500 × g, 4 °C, 5 min) and supernatant was collected.

The supernatant was treated with 120 Kunitz units of bovine deoxyribonuclease I and with $MgCl_2$ added to a final concentration of 25 mM and $CaCl_2$ added to a final concentration of 5 mM. The samples were incubated in 37 °C for 30 minutes with agitation.

The samples were centrifuged (20.000 × g, 4 °C, 30 min). The pellet was solubilized in ice-cold 100 mM Na_2CO_3, agitated for 30 minutes on ice and centrifuged again (20.000 × g, 4 °C, 30 min). This step was performed two times.

The supernatant was removed and resulting pellet was resuspended in 250 µl of 50 mM $(NH_4)HCO_3$ with sequencing grade modified trypsin and incubated at 37°C overnight.

After the digestion the mixture was centrifuged (20.000 × g, 4 °C, 30 min). The pellet was resuspended in freshly prepared ice-cold 100 mM Na_2CO_3, agitated on ice for 30 minutes, then repeatedly frozen and thawed and finally centrifuged at 20.000 × g.

The trypsin-digested membranes were resuspended in 50 µl of deionized water and sonicated in bath sonicator for 5 minutes. The suspension was mixed with concentrated TFA to a final concentration of 70% and with CNBr to a final concentration of 20 mg/ml. The digestion was performed overnight in the dark at room temperature with agitation.

The digested peptides were dried until complete dryness within a speedvac with cold trap and then twice solubilized with 70% methanol and dried again to ensure the complete CNBr removal.

The dried samples were resuspended with Solvent B (80% ACN, 10% isopropyl alcohol, 1% formic acid) and further formic acid was added to 5% concentration. Samples were sonicated and diluted with Solvent A (1% aqueous formic acid) to a final <10% ACN concentration.

The samples were then delipidated with OptiTrap. The column was equilibrated with 500 µl of Solvent A, the samples were loaded and the column was washed again with 500 µl of Solvent A. The delipidation was performed with 2500 µl solution of dichloromethane with 1% formic acid. The column was washed again with Solvent A. The peptides were eluted with 250 µl of Solvent B to a fresh tube and concentrated to 20-30 µl within a SpeedVac. The peptide concentration was determined using Pierce Quantitative Colorimetric Peptide Assay Kit. The samples were then dried until complete dryness, stored in -20°C and later analyzed by LC-MS.

4.2.4 LC-MS

The samples were resuspended in buffer A (2% ACN, 0.1% FA in water) and separated on an EASY-Spray reversed phase nano column (internal diameter 75 µm, length 50 cm, particle size 2 µm, pore size 100 Å). Two mobile phase buffers were used - buffer A and buffer B (80% ACN, 0.1% FA in water).

Samples were loaded onto an Acclaim PepMap300 C18 trap column (internal diameter 300 µm, length 5 mm, particle size 5 µm, pore size 300 Å) for 4 min at a flow rate 15 µl/min. The loading solution was composed of water, 2% ACN and 0.1% TFA.

For all samples except hpTC, the valve was opened after 4 minutes and mobile phase B flow was increased from 4% to 35% in 120 minutes at a flow rate of 300 nl/min, followed by a wash with 75% B for 5 minutes at flow rate increased to 400 nl/min, and then 4% B for 5 minutes until the end.

For the hpTC samples, the mobile phase B was increased from 4% to 50% in 120 min at a flow rate of 300 nl/min, followed by a wash with 75% B for 5 minutes at a flow rate of 400 nl/min and then with 4% B for 5 minutes until the end.

Separated peptides were ionized by electrospray ionization and analyzed on Orbitrap Fusion mass spectrometer (Q-OT-qIT, Thermo Scientific). Precursor ions with charge state 2–6 were selected for MS^2. The scanning of precursors was performed at m/z ranging from 350 to 1400 at a resolution of 120,000 (at 200 m/z) with a 5×10^5 ion count target at a precursor selection window of 1.5 Th. High-energy collision dissociation (HCD) fragmentation was used with normalized collision energy of 30 eV with rapid scan in the ion trap.

Additional settings: The MS^2 ion count target was 10^4. The maximum injection time was set to 35 ms. Dynamic exclusion duration was 45 s with 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The top speed mode with 2 seconds cycles was used.

4.2.5 PROTEIN IDENTIFICATION

Raw MS data were analyzed with MaxQuant v1.6.0.1 (Cox and Mann, 2008). The data were searched against UniProt *Rattus norvegicus* protein database, a combination of reviewed Swiss-Prot database (8,063 entries) and un-reviewed TrEMBL database (28,047 entries). The settings were set to maximum of two missed cleavage sites for both trypsin and CNBr. For SPEG and N-glyco-FASP samples, the deamidation of asparagine to aspartate was included in the settings. In hpTC samples the exchange of C-terminally located methionine to homoserine and homoserine lacton was set. Additionally, the oxidation of methionine and N-terminal protein acetylation were set as variable modifications. Carbamidomethylation of cysteine was used as a fixed modification

for all samples except hpTC. False discovery rate was set to 0.01 for both proteins and peptides.

4.2.6 BIOINFORMATIC ANALYSIS

Acquired lists of protein IDs were converted to FASTA format and processed with TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) to predict a presence and number of transmembrane helices (TMHs). Entries with at least one TMH were filtered out of the dataset and considered as IMPs.

The TMHMM prediction algorithm was also used for prediction of TMHs in rat proteome. The FASTA sequence of the complete *Rattus norvegicus* proteome was obtained from UniProt database of proteins (Proteome ID: UP000002494) containing 8.068 reviewed entries from Swiss-Prot database and 21.876 un-reviewed entries from TrEMBL database. The complete proteome (29.944 sequences) was processed with TMHMM to obtain the list of predicted rat IMPs.

The grand average of hydropathy (GRAVY) score was calculated with GRAVY calculator (http://www.gravy-calculator.de/). The statistic evaluation of GRAVY scores was done with two-sample Kolmogorov-Smirnov test (two-sided).

4.2.7 POLYACRYLAMIDE GEL ELECTROPHORESIS

First, the protein concentration of the myocardial SDC-lysate samples was determined. The electrophoresis samples were prepared so that the protein concentration did not exceed 5 mg/ml. The samples were mixed with 5x Leammli buffer and with 2-mercaptoethanol to a final concentration of 5% and boiled at 100 °C for 5 minutes.

Mini-PROTEAN TGX stain-free precast gels from Bio-rad were used. The gels were mounted into the electrophoresis chamber. The chamber was filled with TGS buffer.

In total, an amount of the sample corresponding to 25 µg of proteins from each sample was pipetted onto the gel wells alongside with 4 µl of protein standard.

The electrophoretic conditions were as follows: For the first 30 minutes the voltage was set to 50 V then the voltage was increased to 200 V.

4.2.8 WESTERN BLOTTING

The gels were washed in distilled water for 5 minutes then the stain-free fluorophore was activated in ChemiDoc. The gels were equilibrated in Towbin buffer (25 mM Tris, 192 mM glycine, 10% methanol, pH 8.3).

The PVDF membranes were activated by a sequence of washes in 100% methanol (5 s), distilled water (10 min) and Towbin buffer (15 minutes).

The sandwich, consisted of 2 layers of gel blotting paper on anode side followed by activated PVDF membrane, gel and 2 layers of gel blotting paper on cathode side, was prepared. The blotting was performed in Trans-Blot Turbo by Bio-rad using the pre-set “turbo” mode.

The membranes were washed in PBST buffer (PBS buffer, 0.1% Tween 20) for 5 minutes and then blocked by washing for 30 minutes in 5% non-fat milk in PBST. The milk was removed by a sequence of three washes in PBST for 5 minutes.

The membranes were sealed with primary antibodies diluted in PBST in a plastic bag and incubated for 90 minutes (or overnight) at room temperature with agitation.

The membranes were washed three times in PBST and then incubated with secondary antibodies conjugated with horseradish peroxidase, diluted 1:10.000 in PBST. The incubation was done in Petri dishes. After the incubation the membranes were again washed three times in PBST.

The immunodetected proteins on the membranes were visualized using ImmunoCruz Luminol reagent. The chemiluminiscent signals were detected in ChemiDoc Imaging system using Image Lab 5 software.

5. RESULTS

5.1 PROFILING OF RAT CARDIOMYOCYTE PROTEOME USING „PITCHFORK“ APPROACH

To analyze rat myocardial membrane proteome we used four different methods – hpTC (Vít *et al*., 2016) to target hydrophobic (transmembrane) segments of IMPs, N-glyco-FASP (Zielinska *et al*., 2010) and solid phase extraction of N-glycopeptides (SPEG; Tian *et al*., 2007) to target the hydrophilic (extra-membrane, extracellular and lumenal) segments of IMPs, and SDC-trypsin (Masuda *et al*., 2008) to target the remaining hydrophilic segments of IMPs. We chose these four protocols because a combination of methods targeting IMPs and peptides of different characteristics should lead to increased membrane proteome coverage as shown in the latest publication of our laboratory (Vít *et al*., 2019).

In the original N-glyco-FASP protocol, samples are first fully processed and digested using the FASP protocol (Wiśniewski *et al*., 2009) which utilizes an ultrafilter centrifugation device. Then the N-glycopeptides are captured using lectins retained on the same type of filters. However, we found out that the FASP workflow is prone to major sample loss (due to the adhesion of proteins and peptides to the filters) and needs to be done in multiple parallel replicates to gain 100 µg of peptide digest that is needed for the following steps of the method. Therefore, we adopted an alternative technique of sample digestion (avoiding filters during the digestion) - SDC-aided in-solution digestion followed by detergent removal by phase transfer (Masuda *et al*., 2008). This allowed us to prepare a sufficient amount of peptide material per sample replicate without the need to combine multiple digests prepared with traditional FASP.

In the original SPEG protocol (Tian *et al*., 2007) samples are digested in the presence of 2 M urea at 37 °C. This, however, is known to lead to carbamoylation of peptides (Kollipara and Zahedi, 2013). Moreover, urea does not solubilize membranes, which may not be optimal for analysis of IMPs. We therefore again employed the in-solution digestion according to Masuda *et al*. (2008) and used SDC in the initial step instead of urea.

We were also interested to see to what extent does the use of membrane-enriched fraction (MF) as a starting material in SPEG and N-glyco-FASP impact the number of identified IMPs compared to the total tissue lysate (TL). The reasoning behind this was to test whether preparation of MF is even necessary for identification of significantly

higher numbers of IMPs since preparation of MF comes at the cost of significantly larger amount of initial sample. This is especially critical in case of tissues with limited availability. The total tissue lysate was prepared following the protocol by Masuda *et al*. (2008), the crude membrane-enriched fraction was prepared according to the standard protocol by Nielsen *et al*. (2005). These two different starting materials were used with both N-glyco-FASP and SPEG. The original N-glyco-FASP and SPEG protocols did not employ such enrichment.

Both N-glyco-capture strategies allow for selective enrichment of hydrophilic segments of IMPs. However, the hydrophobic transmembrane helices are completely neglected in these workflows. The hpTC protocol specifically targets these hydrophobic segments and allows for their identification by targeting and cleaving methionine residues that are highly frequent in membrane-embedded parts of IMPs by cleavage of the residues with CNBr. This method was used as described by Vít *et al*. (2016). The starting material for hpTC was a crude membrane-enriched fraction isolated and purified after lysis in hypotonic buffer.

Both N-glyco-capture methods aiming at glycosylated extra-membrane segments together with hpTC targeting the transmembrane segments should be complementary and should provide decent membrane proteome coverage. However, some of the proteins that do not fall into these categories, such as non-glycosylated (or poorly glycosylated) IMPs, might be missed during the detection. To capture these proteins or peptides, we also performed an analysis of non-targeted standard total lysate trypsin digest (in the presence of SDC) according to Masuda *et al*. (2008).

Methods were performed in technical duplicates. The LC/MS analysis of SPEG, N-glyco-FASP and SDC-trypsin was identical to hpTC. The only exception was a steeper gradient of organic solvent (4-50%) in the separation of hpTC samples instead of 4-35% that was used for the analysis of remaining samples. This was done to ensure the elution of the most hydrophobic peptides. Presented results were gained by combined data analysis of the duplicates at false discovery rate set to 0.01 for both proteins and peptides, and proteins confidently identified with at least one peptide.

5.1.1 IDENTIFICATION OF IMPS IN EACH DATASET

An IMP is characterized as a protein with at least one transmembrane alpha helix (TMH). The prediction of TMHs was based on Tied Mixture Hidden Markov Model

(TMHMM, Krogh *et al*., 2001) using web-based application available at www.cbs.dtu.dk/services/TMHMM/.

The overall numbers of identified proteins and IMPs are summarized in **Table 5.1**. The highest number of identified IMPs yielded SPEG protocol from membrane-enriched fraction (483) followed by SDC-trypsin (374) and SPEG from total tissue lysate (349). By combination of all six datasets, we identified 4494 unique proteins of which 1006 (22%) were IMPs. The complete list of all identified IMPs is available for download at https://biocev.lf1.cuni.cz/membrane-proteome, or is provided as **Supplementary Tables 1-7** on the attached CD.

Dataset	Number of identified proteins	Number of identified IMPs	Relative proportion of identified IMPs
hpTC	886	308	35%
N-glyco-FASP MF	1613	314	19%
N-glyco-FASP TL	913	273	30%
SPEG MF	1571	483	31%
SPEG TL	1577	349	22%
SDC-trypsin	2585	374	14%

Table 5.1: Total number of identified IMPs in all six data sets

The Venn diagram in **Figure 5.1A** shows the overlap in identified IMPs between SDC-trypsin, hpTC and both N-glyco-capture methods of both total tissue lysate (TL) and membrane-enriched fraction (MF) combined. We combined the results of all N-glyco-capture datasets into one set because they all target peptides of the similar physico-chemical properties. The overlap of the identified IMPs in all six datasets is rather small (102 IMPs out of 1006). The Venn diagram **Figure 5.1B** shows detailed results of the two N-glyco-capture methods applied to two different starting materials and their overlaps with SDC-trypsin and hpTC. This data demonstrate that all N-glyco-capture methods but SPEG MF contributed almost evenly to the total number of identified IMPs. The SPEG MF provided the highest number of unique identifications (218) and also identified the most IMPs out of all six datasets.

This data suggests that each protocol favors a different subset of IMPs and also shows that these methods are complementary to each other.

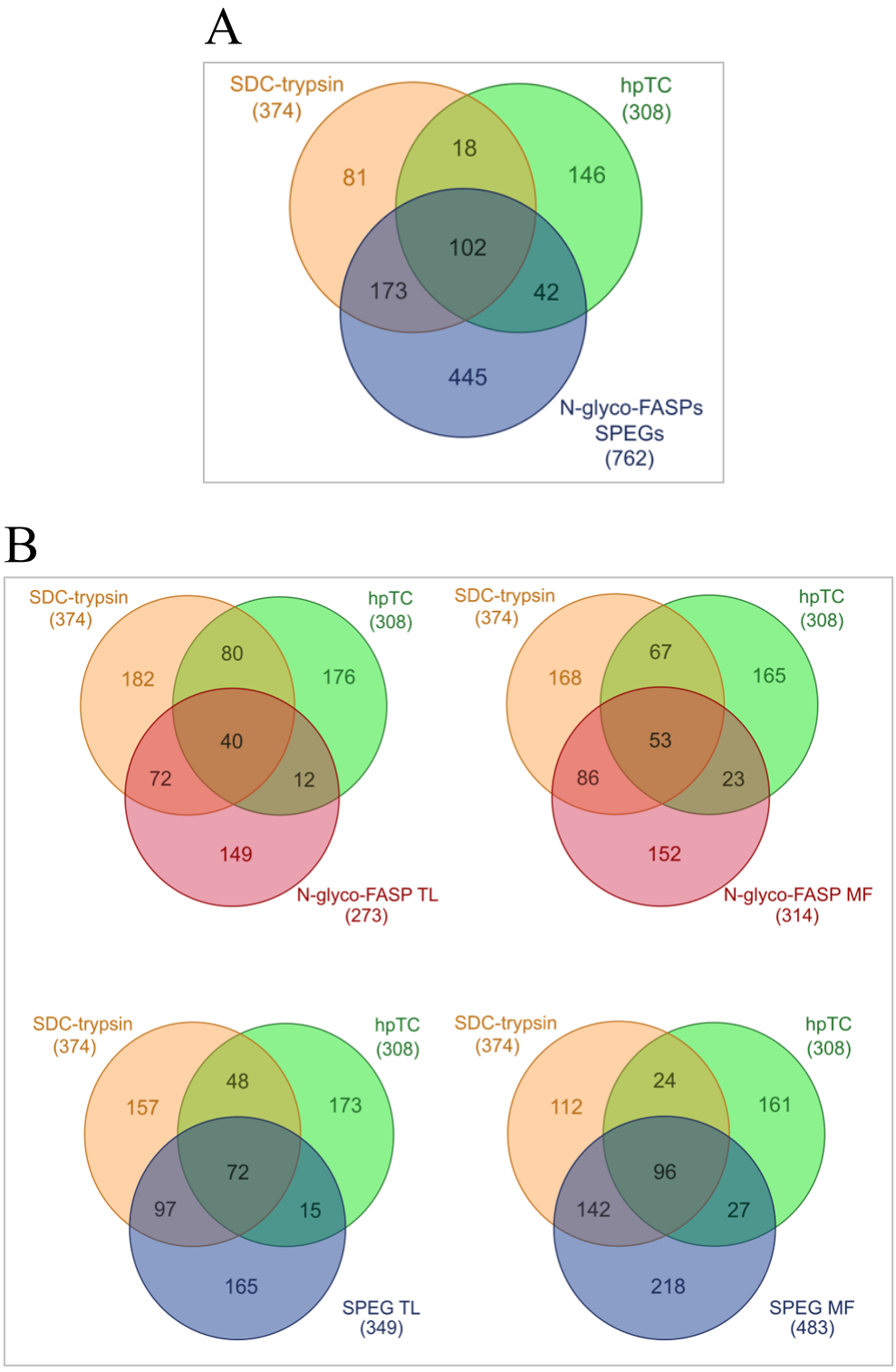

Figure 5.1: Number of identified IMPs in each data set. Each method was performed in technical duplicate and the data was combined. **A)** Comparison of SDC-trypsin, hpTC and N-glyco-FASP, SPEG MF and TL combined. **B)** Differences among the four N-glyco-capture workflows. Comparison of SDC-trypsin, hpTC and each N-glycoprotein capture method separately.

The comparison of the effect of different starting material in **Figure 5.2** shows that in both N-glyco-FASP and SPEG the use of membrane-enriched samples (MF) resulted in significantly higher numbers of identified IMPs (15% higher in N-glyco-FASP, 38% higher in SPEG) as well as higher number of unique identifications (36% higher in case of N-glyco-FASP, 104% higher in SPEG compared to the total lysate). This clearly demonstrates that the pre-enrichment of the starting material positively impacts the number of identified IMPs.

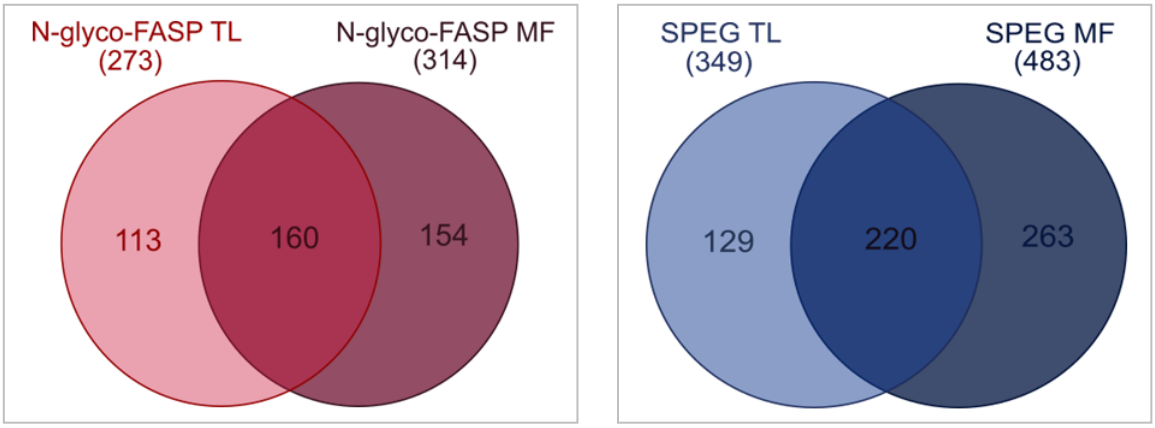

Figure 5.2: Difference between used starting material (TL or MF) in both N-glyco-capture methods.

The identified IMPs in each protocol contained 1-33 transmembrane helices as shown in **Figure 5.3**. The comparison of TMHs distribution in each dataset with a rat membrane proteome prediction (see chapter **4.2.6 Bioinformatic analysis** in Methods) was similar with the exception of 1-span and 7-span IMPs. The data also show that SPEG, N-glyco-FASP and SDC-trypsin most effectively identify more hydrophilic single-spanning IMPs. The hpTC protocol resulted in twice as lower share of single-spanning IMPs but in the largest relative share of multi-spanning IMPs compared to the remaining datasets. This preference of hpTC for multi-span IMPs is likely caused by the fact that the method targets the membrane-protected segments of IMPs, i.e. TMHs. The method therefore favors IMPs with a higher number of TMHs. The hpTC data roughly corresponds to the predicted rat proteome except for 7-span IMPs. This is likely caused by the fact that majority of 7-span IMPs are G-protein coupled receptors (GPCR) which expression is tissue specific (Regard *et al*., 2008). For this reason we cannot observe the majority of GPCRs in a single and highly specialized tissue such as myocardium.

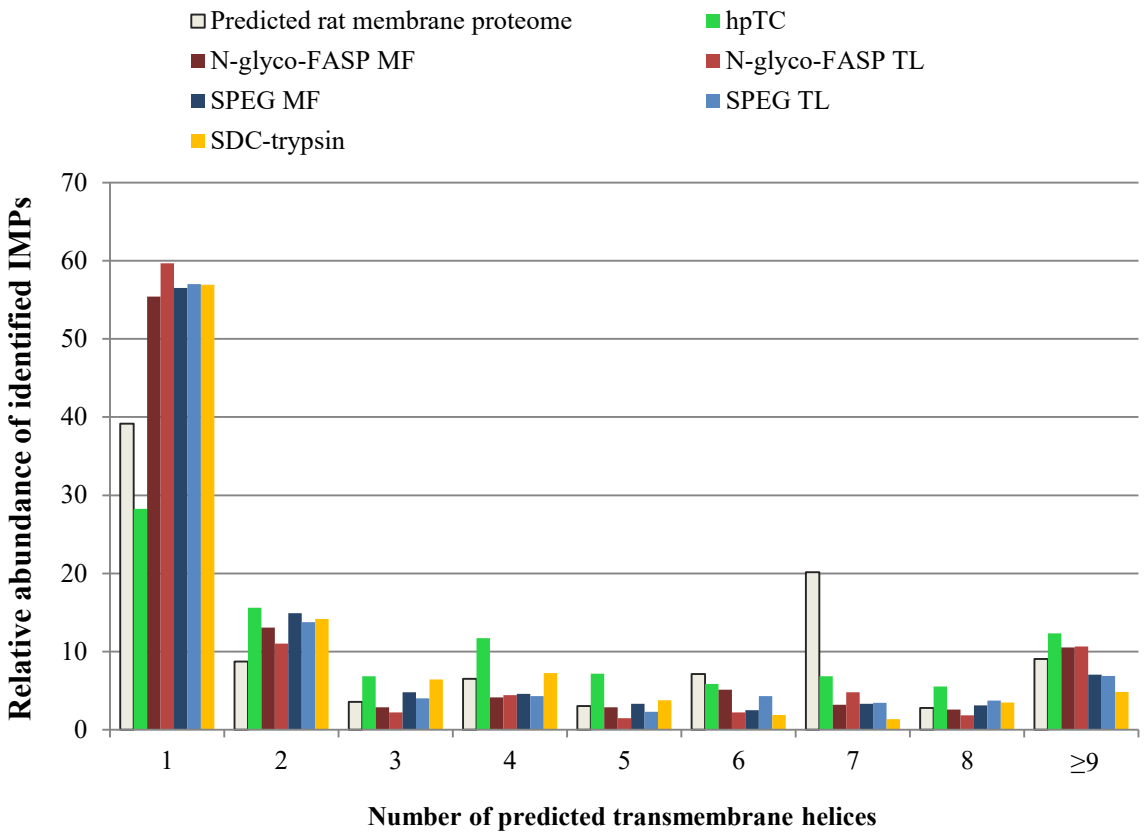

Figure 5.3: Relative abundance of predicted TMHs in all 6 datasets compared to the relative abundance of predicted TMHs in rat proteome.

5.1.2 DISTRIBUTION OF PROTEIN AND PEPTIDE HYDROPATHY

To test whether each method targets a different subset of proteins we calculated the grand average hydropathy (GRAVY) score (Kyte & Doolittle, 1982) of identified IMPs. The results in **Figure 5.4** show that hydrophobicity of IMPs identified by hpTC method is significantly shifted towards more hydrophobic GRAVY values compared to all four N-glyco-capture datasets and SDC-trypsin. This difference is statistically significant (**Table 5.2).** This demonstrates that hpTC targets different subsets of membrane proteome compared to both N-glyco-capture methods and SDC-trypsin.

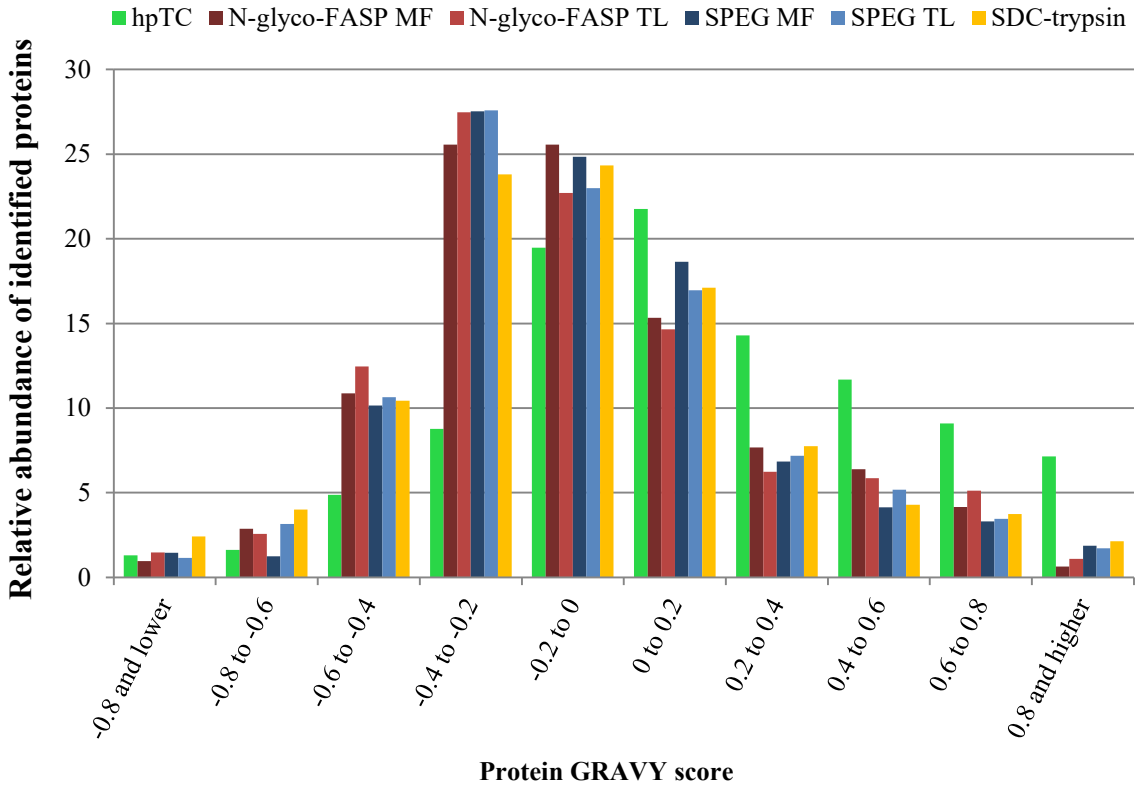

Figure 5.4: Comparison of GRAVY scores of identified IMPs in all six datasets

N-glyco-FASP TL	D = 0.0430 p = 0.946				
SPEG MF	D = 0.0426 p = 0.875	D = 0.0475 p = 0.818			
SPEG TL	D = 0.0327 p = 0.994	D = 0.0342 p = 0.993	D = 0.0311 p = 0.988		
hpTC	D = 0.2846 p = 0.000	D = 0.3050 p = 0.000	D = 0.2739 p = 0.001	D = 0.2864 p = 0.001	
SDC-trypsin	D = 0.0358 p = 0.979	D = 0.0584 p = 0.641	D = 0.0466 p = 0.739	D = 0.0366 p = 0.966	D = 0.2746 p = 0.001
	N-glyco-FASP MF	**N-glyco-FASP TL**	**SPEG MF**	**SPEG TL**	**hpTC**

Table 5.2: The statistical analysis of differences between the distributions of GRAVY values of identified IMPs in all six datasets performed with two-sample Kolmogorov-Smirnov test (two-sided). The datasets that differed statistically significantly are highlighted.

To confirm that the individual methods also target peptides of distinct physico-chemical properties we also calculated the GRAVY scores of peptides belonging to identified IMPs in all six datasets (**Figure 5.5**). The GRAVY values of the hpTC show the same shift towards positive (more hydrophobic) values compared to other 5 datasets. This difference is again statistically significant (see **Table 5.3** for statistical data). Moreover, there is a visible and statistically significant shift towards more hydrophilic peptides in case of N-glyco-FASP TL dataset compared to all remaining datasets, except SDC-trypsin, where the difference between the distributions was not statistically significant.

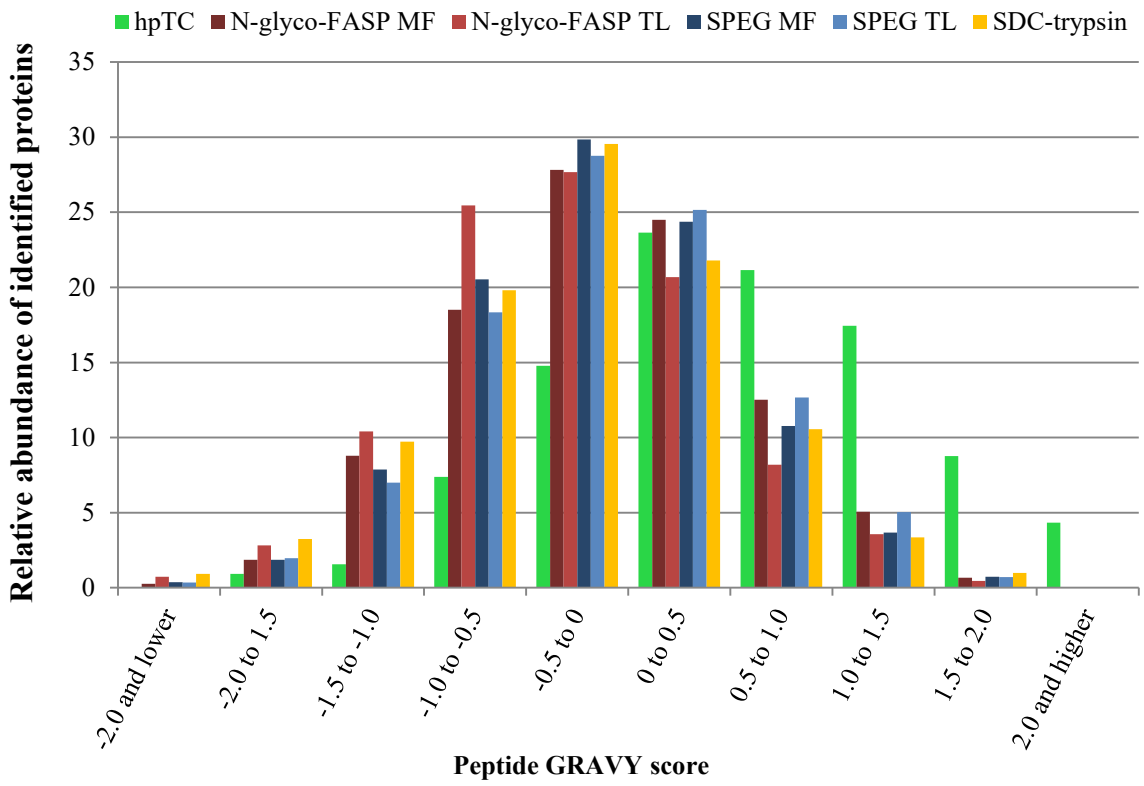

Figure 5.5: Comparison of peptide GRAVY score in each dataset

N-glyco-FASP TL	D = 0.1199 p = 0.001				
SPEG MF	D = 0.0398 p = 0.492	D = 0.0997 p = 0.001			
SPEG TL	D = 0.0285 p = 0.870	D = 0.1237 p = 0.001	D = 0.0495 p = 0.160		
hpTC	D = 0.3535 p = 0.001	D = 0.4357 p = 0.001	D = 0.3754 p = 0.001	D = 0.3414 p = 0.001	
SDC-trypsin	D = 0.0672 p = 0.039	D = 0.0647 p = 0.065	D = 0.0432 p = 0.294	D = 0.0696 p = 0.014	D = 0.3944 p = 0.001
	N-glyco-FASP MF	**N-glyco-FASP TL**	**SPEG MF**	**SPEG TL**	**hpTC**

Table 5.3: The statistical analysis of differences between the distributions of GRAVY scores of peptides belonging to identified IMPs in all six datasets performed with two-sample Kolmogorov-Smirnov test (two-sided). The datasets that differed statistically significantly are highlighted.

This data confirms our findings that hpTC targets IMPs and peptides of different physico-chemical properties compared to the remaining methods and that the methods are complementary.

5.1.3 DEAMIDATION OF ASPARAGINE TO ASPARTATE

Glycosylation is a prominent post-translational modification among membrane proteins (Chandler and Costello, 2016). To distinguish which of the identified peptides in the N-glyco-FASP and SPEG datasets were glycosylated, we analyzed the proportion of peptides carrying the deamidation of Asn to Asp in both N-glycopeptides capture datasets. The deamidation is a result of cleavage of the bond between the N-glycosylated peptide and the oligosaccharide chain by PNGase F in both N-glyco-capture methods and is characterized by mass change of 0.98 Da (Yang and Zubarev, 2010). The mass change is detected during MS analysis and serves as a marker of glycosylation (Tian *et al*., 2007).

The data in **Table 5.4** show that the deamidation of peptides from IMPs was prominent in the N-glyco-FASP TL dataset where 64% of all identified IMP peptides carried the modification and in N-glyco-FASP MF dataset where the modification was carried in 41% of peptides from IMPs. The deamidation in SPEG datasets was present in-between 26-31% of identified IMP peptides. This data suggests that N-glyco-FASP

is more specific towards N-glycosylated IMPs compared to SPEG. However, total numbers of identified IMPs were higher in SPEG datasets.

The result of both SPEG datasets and N-glyco-FASP MF dataset also suggests that majority of the identified IMP peptides were bound non-specifically to used consumables (filtration devices, lectins, agarose beads, plasticware) and other peptides and were not removed even with the extensive washing that was employed in both protocols. In N-glyco-FASP TL dataset, the non-specific interaction accounted for only 36% of identified IMP peptides.

The deamidation was observed in 50% and 47% of IMPs in N-glyco-FASP MF and TL datasets, respectively. In the SPEG samples the modification was present in 38% IMPs in MF dataset and in 45% IMPs in TL dataset (see **Table 5.4**). This suggests that 38-50% of identified IMPs (depending on the dataset) were identified thanks to the former presence of N-glycans.

Dataset	**Total number of identified IMP peptides**	**Number of identified IMP peptides with deamidation**	**Percentage**
N-glyco-FASP MF	751	310	41%
N-glyco-FASP TL	672	433	64%
SPEG MF	1933	501	26%
SPEG TL	1113	349	31%
Dataset	**Total number of identified IMPs**	**Number of identified IMPs with deamidation**	**Percentage**
N-glyco-FASP MF	314	157	50%
N-glyco-FASP TL	273	129	47%
SPEG MF	483	182	38%
SPEG TL	349	157	45%

Table 5.4: Number of IMPs with detected deamidation in each dataset and the deamidation of identified peptides related to IMPs

5.1.4 CELLULAR LOCALIZATION OF IDENTIFIED IMPS

To evaluate whether this multi-pronged approach is able to access all cellular compartments we searched the list of the 1006 identified IMPs against the rat Uniprot database and based on gene ontology (GO) we matched identified IMPs with their respective cellular localization.

The plasma membrane protein in this context was defined by at least one of the plasma membrane GO annotation in **Table 5.5** and at the same time not being characterized by any other GO annotation of any other compartment. This way we were able to identify IMPs that are associated solely with plasma membrane and not with any other cellular compartment as the overlap between compartments was extensive. This overlap is a result of the dynamic nature of proteins with respect to their synthesis, post-translational modification, transport and final localization within a membrane which reflects into multiple shared GO annotations.

We used the same approach to filter out proteins associated with mitochondria, endoplasmic reticulum, Golgi apparatus, nucleus and vesicles.

The localization of identified IMPs is shown in **Figure 5.6**. Among the identified proteins we found 167 IMPs associated solely with plasma membrane IMPs and 64 IMPs related to plasma membrane-associated vesicles. We also identified 122 IMPs associated with either endoplasmic reticulum (ER) or Golgi apparatus (GA) and also 14 vesicular proteins associated with either of these compartments. 113 IMPs were characterized as mitochondrial. Only 14 IMPs were localized within nuclear envelope. Proteins corresponding to other organelles such as endosomes (7), lysosomes (6) and peroxisomes (1) were also identified.

Among the identified IMPs the most prominent were plasma membrane IMPs (16.6%) mitochondrial IMPs (11.2%) and IMPs associated with either ER or GA (12.1%). This suggests that a combination of the methods targets membranes of all major cellular compartments.

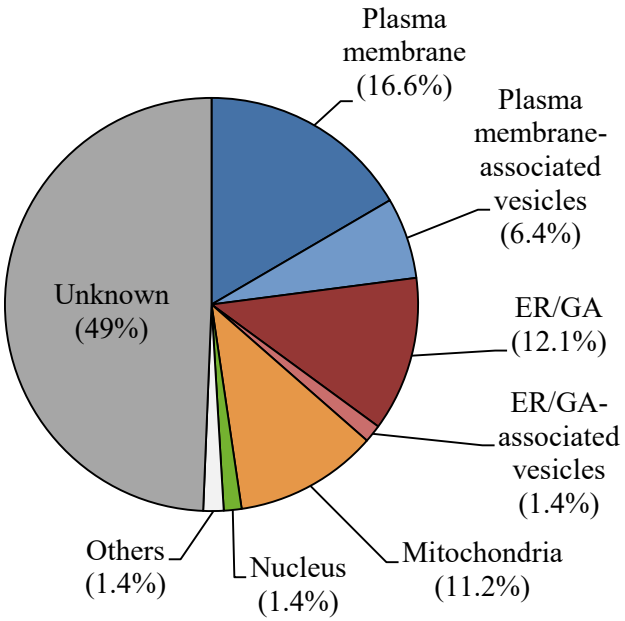

Figure 5.6: Cellular localization of identified IMPs. "Others" stands for IMPs associated with endosomes, lysosomes, peroxisomes and vacuole. "Unknown" in this context represents IMPs that were either annotated by more than one GO tag assigning them to more than one compartment, or IMPs that were annotated with neither of GO annotation tags listed in Table 5.5.

Cellular compartment	Gene ontology annotation
Plasma membrane	integral component of plasma membrane [GO:0005887], plasma membrane [GO:0005886], cell surface [GO:0009986], membrane raft [GO:0045121], plasma membrane raft [GO:0044853]
Mitochondrion	mitochondrion [GO:0005739], mitochondrial inner membrane [GO:0005743], integral component of mitochondrial inner membrane [GO:0031305], mitochondrial outer membrane [GO:0005741], integral component of mitochondrial outer membrane [GO:0031307]; respirasome [GO:0070469], mitochondrial respiratory chain complex I-IV [GO:0005747] [GO:0005749] [GO:0005750] [GO:0005751]; mitochondrial proton-transporting ATP synthase complex [GO:0005753], cation-transporting ATPase complex [GO:0090533], proton-transporting ATP synthase complex, coupling factor F(o) [GO:0045263], calcium ion-transporting ATPase complex [GO:0090534]
Endoplasmic reticulum	endoplasmic reticulum [GO:0005783], endoplasmic reticulum membrane [GO:0005789], integral component of endoplasmic reticulum membrane [GO:0030176], smooth endoplasmic reticulum [GO:0005790], rough endoplasmic reticulum [GO:0005791], sarcoplasmic reticulum membrane [GO:0033017], sarcoplasmic reticulum [GO:0016529]
Golgi apparatus	endoplasmic reticulum-Golgi intermediate compartment [GO:0005793], Golgi apparatus [GO:0005794], Golgi membrane [GO:0000139], cis-Golgi network [GO:0005801], trans-Golgi network [GO:0005802]
Nucleus	nucleus [GO:0005634], nuclear membrane [GO:0031965], nuclear inner membrane [GO:0005637], nuclear envelope [GO:0005635]
ER/GA-associated vesicles	COPI-coated vesicle membrane [GO:0030663], COPII-coated ER to Golgi transport vesicle [GO:0030134], trans-Golgi network transport vesicle [GO:0030140], ER to Golgi transport vesicle membrane [GO:0012507]
Plasma membrane-associated vesicles	vesicle [GO:0031982], clathrin-coated vesicle [GO:0030136], cytoplasmic vesicle membrane [GO:0030659], transport vesicle [GO:0030133], transport vesicle membrane [GO:0030658], synaptic vesicle [GO:0008021], synaptic vesicle membrane [GO:0030672]
Others	endosome [GO:0005768], endosome membrane [GO:0010008], lysosome [GO:0005764], lysosomal membrane [GO:0005765], peroxisome [GO:0005777] , peroxisomal membrane [GO:0005778]

Table 5.5: List of used GO annotations

We also asked whether the relative representation of the cellular compartments differs between the individual methods. Since plasma membrane, mitochondrial and ER/GA IMPs were the most abundant among the identified IMPs, we looked more closely at the IMP count associated with these compartments in each dataset (**Table 5.6)**. We used the same approach and the same GO annotations described in the previous section to filter out plasma membrane, mitochondrial and ER/GA-associated IMPs.

Dataset	**Number of identified IMPs**	**Number of plasma membrane IMPs**	**Proportion of plasma membrane IMPs**
hpTC	308	30	10%
N-glyco-FASP MF	314	54	17%
N-glyco-FASP TL	273	73	27%
SPEG MF	483	82	17%
SPEG TL	349	53	15%
SDC-trypsin	374	39	10%
	Number of identified IMPs	**Number of mitochondrial IMPs**	**Proportion of mitochondrial IMPs**
hpTC	308	64	21%
N-glyco-FASP MF	314	43	14%
N-glyco-FASP TL	273	33	12%
SPEG MF	483	72	15%
SPEG TL	349	51	15%
SDC-trypsin	374	69	18%
	Number of identified IMPs	**Number of ER/GA IMPs**	**Proportion of ER/GA IMPs**
hpTC	308	48	16%
N-glyco-FASP MF	314	41	13%
N-glyco-FASP TL	273	29	11%
SPEG MF	483	54	11%
SPEG TL	349	39	11%
SDC-trypsin	374	48	13%

Table 5.6: Percentage of identified plasma membrane, mitochondrial and ER/GA-associated IMPs in each dataset

The data show that each individual method is able to access IMPs associated with membranes of all major cellular compartments. It showed that hpTC identified higher proportion of mitochondrial IMPs compared to plasma membrane IMPs (21% mitochondrial IMPs compared to 10% plasma membrane IMPs) which may be attributed to the fact that the surface of mitochondrial membranes is significantly higher compared to the surface of plasma membrane due to the fact that mitochondria account for over 30% of cardiomyocyte cell volume (Piquereau *et al*., 2013). For this reason the cells contain a higher absolute amount of copies of hydrophobic mitochondrial IMPs than copies of hydrophobic plasma membrane IMPs, explaining more mitochondrial IMPs identified by hpTC compared to plasma membrane IMPs. The N-glyco-capture methods identified higher share of plasma membrane IMPs (15-27% of all identified IMPs depending on the dataset) which corresponds to the design of the methods which predominantly target the N-glycosylated proteins that frequently reside on the cell surface.

This data confirm that the combination of all six methods is beneficial for accessing membrane proteomes of all major cellular compartments and that the methods are complementary.

5.2 EFFICIENCY OF DIFFERENT METHODS FOR PREPARATION OF MEMBRANE FRACTION

Effective enrichment of membrane material is an essential prerequisite of every membrane proteomic study. It produces a sample where an amount of non-membrane material (soluble cytosolic proteins, cytoskeletal proteins, proteins attached to the membrane and other molecules) is significantly reduced. Therefore, the identification of (generally less abundant) IMPs is less hindered by the presence of non-membrane components during LC-MS.

To analyze the efficiency of IMPs enrichment in the tested methods, we verified the relative enrichment of 7 different proteins corresponding to different cellular localizations. Three proteins were membrane proteins – caveolin 1, connexin 43 and transferrin receptor protein 1 (TfR1). Four proteins were cytosolic of which 2 were cytoskeletal - α-tubulin and actin, and the other 2 were soluble cytosolic proteins - ferritin and GAPDH.

Three types of samples corresponding to the procedures used in the proteomic analysis were prepared and analyzed by western blot. The first was a total tissue lysate that was prepared according to Masuda *et al*. (2008, section **4.2.1.4** in **Methods**) without any subcellular fractionation. This method was used for preparation of the SDC-trypsin sample as well as for preparation of starting material for both N-glyco-capture methods. The second sample was the crude membrane fraction isolated after lysis in hypotonic buffer with carbonate washes (section **4.2.3** in **Methods**) and was used in hpTC (Vít *et al*., 2016). The third was the membrane fraction sample, isolated after lysis in hypertonic buffer with carbonate washes, prepared according to Nielsen at al. (2005, section **4.2.1.3** in **Methods**). This method of isolation of crude membrane fraction was also used for preparation of starting material for both N-glyco-capture methods in their variant starting with membrane enrichment.

The data of relative enrichment are based on optical density of proteins visualized by western blotting, normalized against the optical density of total protein load and visualized using the stain-free gel technology. Our initial idea was to compare both membrane enrichment strategies to the tissue lysate sample. However, the selected membrane proteins were in all cases under the detection limit in the tissue lysate sample, thus such comparison was not possible. Therefore, we only compared both membrane enrichment techniques against each other.

The data (**Figure 5.7**) shows that all three membrane proteins were enriched in both hypotonic and hypertonic membrane fractions. The relative enrichment of TfR1 in both fractions was comparable. However, the enrichment among the two membrane fractions differed more significantly for caveolin-1 and connexin 43. The relative enrichment in the hypertonic fraction was 36.5% higher in case of caveolin-1 and 25.3% higher for connexin 43 compared to hypotonic membrane fraction.

Interestingly, cytoskeletal proteins were greatly enriched in hypertonic fraction. The relative enrichment of α-tubulin and actin was 36.4% and 176.2%, respectively, when comparing hypotonic and hypertonic fractions with each other.

The analysis of ferritin shows that both methods for preparation of membrane fractions are effective in reducing the amount of cytosolic proteins. The reduction was 57.8% in hypotonic and 68.5% in hypertonic fraction compared to total tissue lysate. GAPDH, on the other hand, was only partially removed from the hypotonic membrane fraction but remained present in the hypertonic fraction.

These results show that both methods for preparation of crude membrane fraction are efficient in the enrichment of IMPs but samples prepared in this manner are still contaminated by cytoskeletal proteins and some of the most abundant cytosolic proteins.

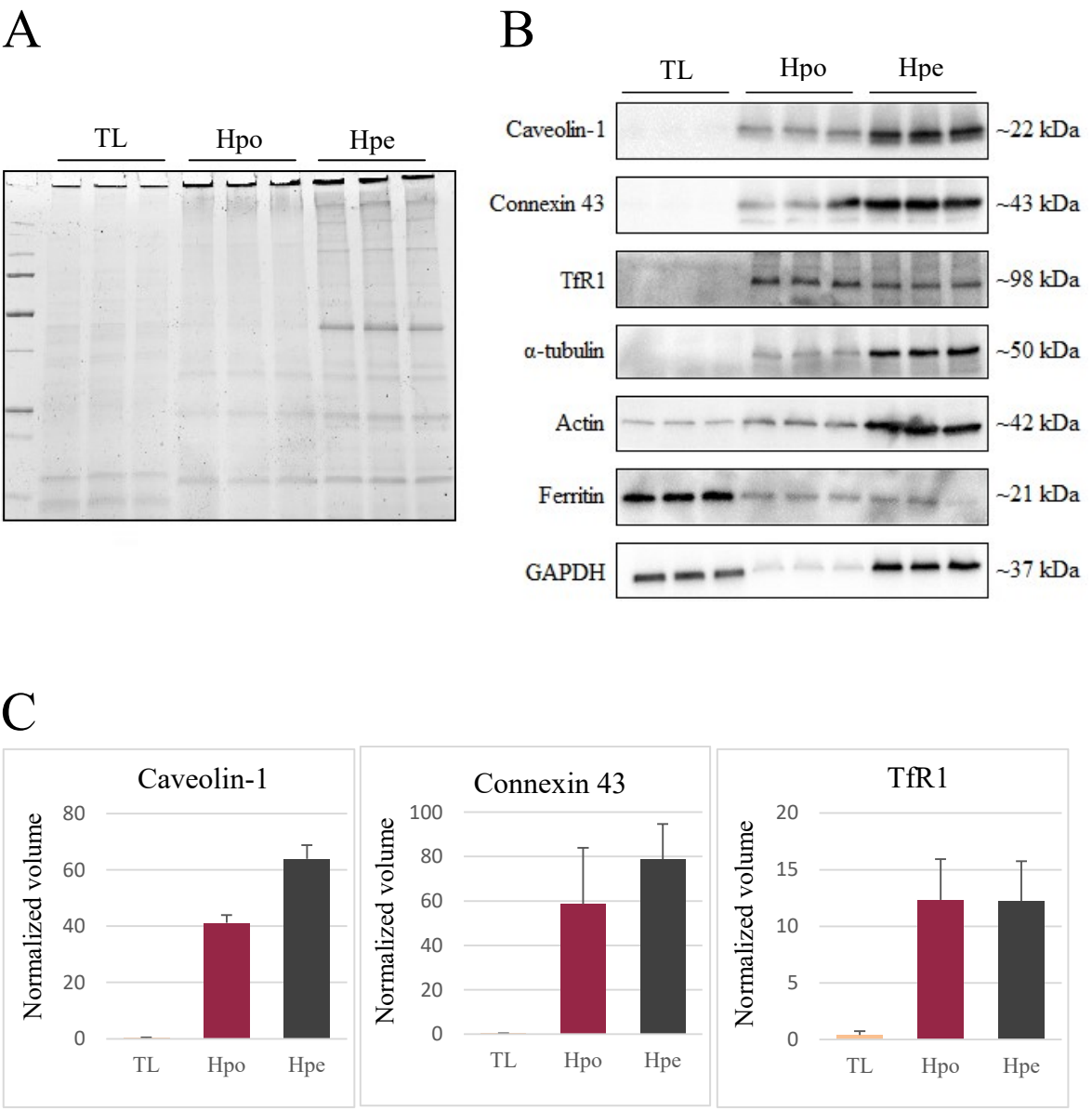

Figure 5.7: Western blot verification of the relative enrichment of selected proteins in total tissue lysate (TL) prepared according to Masuda *et al*. (2008), hypotonic (Hpo) membrane fraction prepared according to Vít *et al*. (2016) and hypertonic (Hpe) membrane fraction prepared according to Nielsen at al. (2005) **A)** A picture of gel used for normalization **B)** Western blot analysis of selected proteins **C)** The chart representation of the enrichment based on normalized volumes of analyzed proteins

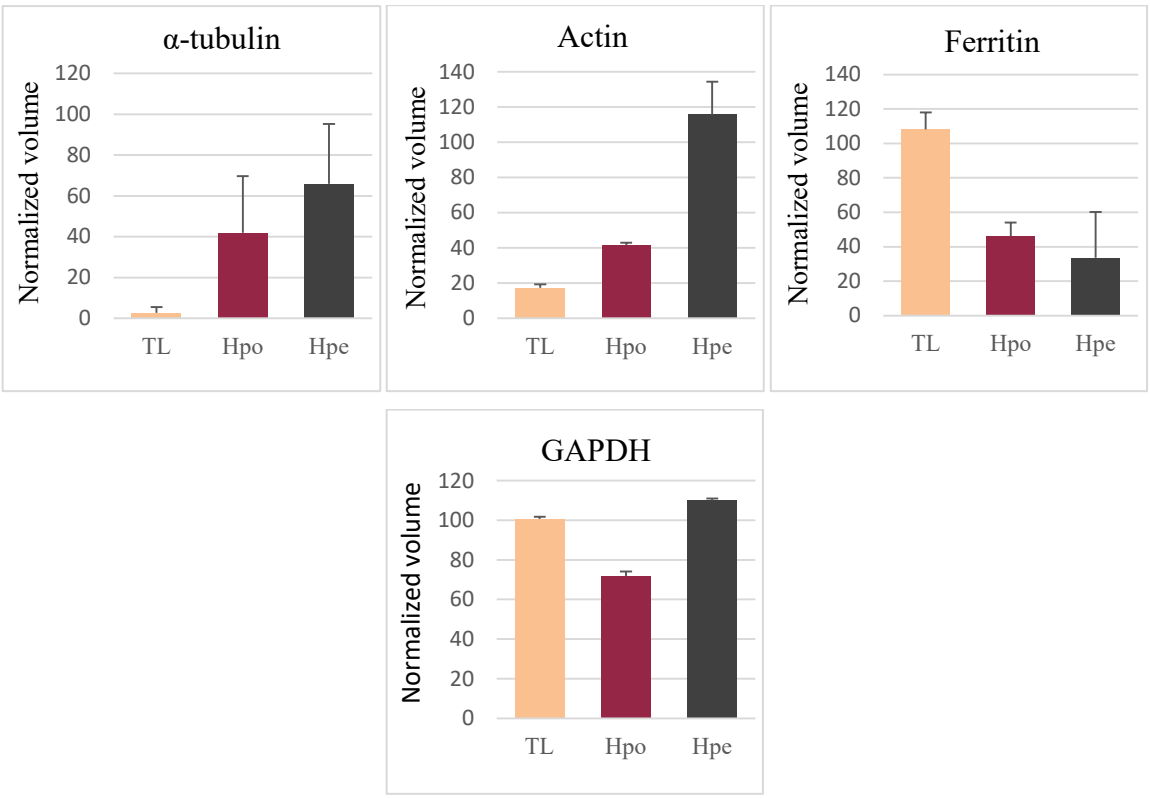

Figure 5.7 continued: The chart representation of the enrichment based on normalized volumes of analyzed proteins (TL = tissue lysate; Hpo = hypotonic membrane fraction, Hpe = hypertonic membrane fraction)

6. DISCUSSION

IMPs play an essential role in many different cellular processes and as such are considered to be attractive pharmaceutical targets. In fact, up to 60% of currently approved drugs target IMPs (Yildirim *et al*., 2007). However, analysis of IMPs is hindered by challenges associated with their physico-chemical properties and low abundance. To overcome these challenges, specific methods designed to cope with the unique nature of IMPs need to be employed.

One of the larger projects at our laboratory is the development of new approaches aiming towards the analysis of membrane proteome. The second project our laboratory focuses on is a study of molecular mechanisms of chronic heart failure with emphasis on identification of novel diagnostic and prognostic biomarkers and, more importantly, on the discovery of new drug targets. The work presented here is the first attempt to merge these two projects and to use the methods presented in this thesis for analysis of myocardial membrane proteome.

To explore the rat myocardial proteome, we used four different approaches – N-glyco-FASP and solid phase extraction of N-glycopeptides (SPEG) to target hydrophilic N-glycosylated extra-membrane segments of IMPs, hpTC to target hydrophobic transmembrane helices of IMPs and SDC-trypsin to non-selectively capture IMPs that might avoid being captured by the N-glyco-FASP, SPEG and hpTC. We also performed N-glyco-FASP and SPEG on two different starting materials – total tissue lysate (TL) and membrane-enriched fraction (MF) – to evaluate the impact of the starting material on the number of identified IMPs. Both MF and TL samples as well as samples for hpTC and SDC-trypsin were prepared from the right heart ventricle of male Wistar rats (*Rattus norvegicus).* The LC-MS analysis of peptide samples prepared with these methods resulted in six datasets that were computationally analyzed.

6.1 NUMBERS OF IDENTIFIED IMPS AND COMPLEMENTARITY OF USED METHODS

In the context of this thesis, IMPs were characterized as proteins with at least one TMH (trans-membrane helix) predicted by TMHMM prediction algorithm (Krogh *et al*., 2001). By using TMHMM on the lists of identified proteins we assessed the number of identified IMPs in each dataset (**Table 5.1**). The highest number of identified IMPs was associated with SPEG MF dataset (483) followed by SDC-trypsin (374), SPEG TL (349), N-glyco-FASP MF (314), hpTC (308) and N-glyco-FASP TL (273).

The combination of the identified IMPs of each individual method resulted in identification of 1006 unique IMPs in rat myocardial cells.

One of the goals of this work was to determine the complementarity of chosen methods on analysis of rat cardiomyocyte membrane proteome. We hypothesized that a combination of different methods, each targeting IMPs or peptides of different physico-chemical properties, may be efficient in providing higher membrane proteome coverage then employing only one of the strategies alone. Our results showed in **Figure 5.1A** and **B** supports the hypothesis by showing that each individual protocol provides a significant portion of unique IMP identifications and that the combination of all protocols results in only a limited overlap. This hypothesis was recently tested, proved and published as "Pitchfork" strategy by our laboratory (Vít *et al*., 2019).

In the original "Pitchfork" study, the membrane proteome of human lymphoma cells was analyzed by the combination of SPEG, hpTC and SDC-trypsin and led to the identification of 1214 unique IMPs in the membrane fraction of the lymphoma cells. In the study, SPEG was performed only with membrane fraction and not with total tissue lysate. The N-glyco-FASP was done but it was not included in the final results because of extensive overlap of its results with SPEG and low number of unique IMP identifications.

However, results presented in this thesis showed that N-glyco-FASP significantly contributed to the 1006 unique identification (128 unique IMPs, 12.7%) compared to N-glyco-FASP in the "Pitchfork" (46 unique IMPs, 3.7%). This difference may be attributed to the fact that in the "Pitchfork" study significantly more IMPs were identified with each individual method compared to numbers of identified IMPs in this thesis (880 in hpTC compared to 308 in this thesis, 675 in SDC-trypsin compared to 374, 642 in SPEG of MF compared to 483). What possibly contributed to this difference was the fact that in the "Pitchfork" study, each method was performed in three replicates compared to two replicates per method in this thesis. This difference in the technical replicates might have led to the smaller overlap between the N-glycopeptide enrichment methods (SPEG and N-glyco-FASP) in this thesis.

The work presented in this thesis is to our knowledge one of the most complete analysis of myocardial membrane proteome up to date (1006 identified IMPs). However, it should be kept in mind that these results were obtained by analysis of only two technical replicates of used methods. More technical replicates would likely result in higher number of identifications. Some of the notable works in this area include in-depth profiling of left

ventricular tissue proteome from end-stage dilated cardiomyopathy patients that resulted in identification of 760 IMPs (Liu *et al*., 2017), analysis of membrane proteome of cardiac fibroblasts which led to identification of 774 IMPs (Sebastião *et al*., 2018) or proteomic analysis of left ventricular tissue of mice that resulted in identification of 534 IMPs (Tian *et al*., 2014). Although not a study of cardiac membrane proteome, the work of the team around Sophia Doll at Max Planck Institute of Biochemistry (Doll *et al*., 2017) resulted in identification of over 10,700 proteins in human heart tissue and it is up to date the most complete analysis of the heart proteome. We acquired the proteomic data from http://maxqb.biochem.mpg.de and used TMHMM prediction algorithm (Krogh *et al*., 2001) on the FASTA sequences of all identified proteins in this study. The TMHMM predicted that 2151 proteins were IMPs. However, it must be noted that obtaining this number of identifications required 400 MS runs compared to 12 MS runs performed in our work.

6.2 PROTEIN/PEPTIDE HYDROPHOBICITY AND TRANSMEMBRANE ALPHA-HELICES

The GRAVY analysis confirmed that methods of the multi-pronged approach target IMPs of different physico-chemical properties. The results showed that hpTC was oriented towards more hydrophobic proteins and peptides. This is because of the hpTC method design in which a membrane fraction is isolated, "shaved" from hydrophilic segments by trypsin and solubilized. Released hydrophobic tryptic peptides are then re-digested with CNBr (Vít *et al*., 2016). The GRAVY values of both N-glyco-capture strategies as well as SDC-trypsin predominantly targeted less hydrophobic IMPs. This preference was well reflected on the peptide level as well. The hpTC, on the opposite, showed the preference for more hydrophobic values. The data for hpTC, SPEG and SDC-trypsin roughly corresponded to the GRAVY data published in the "Pitchfork" study (Vít *et al*., 2019). However, there was a significant shift in N-glyco-FASP TL dataset towards hydrophilic values (**Figure 5.5**) that correlated with a higher share of deamidation of peptides (**Table 5.4**) and a higher proportion of identified IMPs associated with plasma membrane (**Table 5.6**) in this dataset. This correlation may provide an explanation to this shift in peptide GRAVY distribution and it will be discussed in **chapter 6.5**.

The complementarity of the individual methods was also reflected in the number of predicted alpha-TMHs (**Figure 5.3**). The results showed that hpTC was oriented

towards proteins with more than one TMHs compared to both N-glyco-capture approaches and SDC-trypsin that were more efficient in the identification of single-span IMPs. This may suggest that single-span IMPs are more hydrophilic compared to multi-span IMPs. These results also correspond to the data published by Vít *et al*. (2019).

The comparison between our datasets with predicted rat membrane proteome showed that all four methods used in this work were underperforming in the identification of 7-span IMPs. This is likely caused by the fact that majority of 7-span IMPs are G-protein coupled receptors (GPCR), similarly as in human proteome where 623 out of 737 7-span IMPs are GPCRs (Vít *et al*., 2019). In fact, an overall map of the GPCR subset of rat genome created by Gloriam *et al*. (2007) contains more than 1800 genes coding for GPCRs. The analysis of expression of more than 350 GPCRs in 41 different tissues (Regard *et al*., 2008) also showed that the GPCR expression is tissue-specific suggesting that it is impossible for all GPCRs to be present in a single sample of a single tissue. However, this comparison has its limits because rat proteome is incompletely annotated. Nevertheless, the results roughly correspond to the TMHs distribution in human proteome (Fagerberg *et al*., 2010) and also to the data published by Vít *et al*. (2019)

6.3 DEAMIDATION IN IDENTIFIED IMPS

N-glycosylation is one of the most common post-translational modification and it is prominent among membrane proteins (Chandler and Costello, 2016). The determination of N-glycosylation among identified IMPs was based on deamidation of asparagine (Asn) to aspartate (Asp) caused by PNGase F in both SPEG and N-glyco-FASP. The deamidation is characterized by a mass change of 0.98 Da (Yang and Zubarev, 2010) that is detected during LC-MS and serves as a marker of N-glycosylation (Tian *et al*., 2007).

The results showed that the deamidation was present in 38-50% (depending on the dataset) of identified IMPs in all four datasets (**Table 5.4**). However, we cannot conclude that IMPs in this share are N-glycoproteins. We can only conclude that up to 38-50% IMPs were identified thanks to the deamidation that marks the former presence of N-glycan moiety. The true share of N-glycoproteins can be, in fact, different. First, according to Palmisano *et al*. (2012), proteins spontaneously deamidate during sample processing. The spontaneous deamidation results in the same modification

and is detected during LC-MS resulting into false-positive characterization of some IMPs as N-glycosylated IMPs. The true share of N-glycosylated IMPs would therefore be lower. Secondly, not all (formerly) N-glycosylated peptides of identified IMPs are identified during MS. This means that the true share of N-glycoproteins can be also higher.

The results also showed that the relative enrichment of N-glycopeptides related to identified IMPs (presence of deamidation) was 64% in N-glyco-FASP tissue lysate (TL) dataset, but was significantly lower in N-glyco-FASP membrane fraction (MF) dataset (41% of identified IMP peptides) and both SPEG datasets (31% peptides in SPEG TL and 26% peptides in SPEG MF dataset). This showed that N-glyco-FASP protocol is more specific towards N-glycopeptides compared to SPEG. This data also means that majority of peptides in N-glyco-FASP MF, SPEG MF and TL datasets non-specifically interacted with either hydrazide-coated agarose beads in case of SPEG, with lectins or used ultrafiltration columns in N-glyco-FASP or non-specifically aggregated to bound N-glycosylated peptides, even though extensive washing with high salt buffers was employed in both methods. The non-specifically bound peptides were then eluted together with N-glycopeptide after PNGase F treatment.

The data also showed that the deamidation of peptides from IMPs was always more prominent in TL samples (31% compared to 26% in SPEG MF, 64% compared to 41% in N-glyco-FASP MF). The reason why TL samples contained higher proportion of deamidated IMP peptides is unknown. Theoretically, it can reflect the different physico-chemical properties of the samples and therefore different conditions of peptide competition and binding to lectins in N-glyco-FASP samples and reaction of peptides with hydrazide in SPEG samples.

6.4 THE IMPACT OF USED STARTING MATERIAL – MEMBRANE FRACTION OVER TISSUE LYSATE

The results of the comparison of membrane fraction and tissue lysate (**Figure 5.2**) showed that in both N-glyco-capture methods there was an increase in the number of identified IMPs in MF compared to TL. In SPEG dataset, the total number of IMP identifications was 38% higher in MF (483 IMPs) compared to TL (349 IMPs). However, in N-glyco-FASP the increment was significantly lower - only 15% higher in MF dataset (314 IMPs) compared to TL (273 IMPs).

The lower increment between MF and TL in N-glyco-FASP datasets compared to SPEG datasets may be caused by different specificity of SPEG and N-glyco-FASP towards N-glycosylated peptides from IMPs. As shown in **Table 5.4**, N-glyco-FASP is generally more specific for N-glycosylated peptides related to identified IMPs (41-64% of identified IMP peptides carried the deamidation of Asn to Asp) compared to SPEG. In general, the more specific the N-glyco-capture protocol is, the less sensitive this protocols is to the presence of multiple non-glycosylated (non-IMP) proteins present in the tissue lysate.

Although preparation of membrane fraction is more time consuming compared to preparation of standard tissue lysate, the LC-MS analysis of MF samples results in significantly higher numbers of identified IMPs (38% and 15% higher in SPEG and N-glyco-FASP dataset, respectively) and for this reason MF preparation should be employed as a standard step in membrane proteome analysis workflow. However, it must be noted, that the recommended amount of peptide digest for both N-glyco-FASP and SPEG is approximately 100 µg of peptide material. Such amount of peptides roughly corresponds to about 100 mg of tissue that is required for preparation of membrane fraction. Gaining 100 µg of peptide material from unfractionated tissue lysate requires only units of mg. This demand for a large amount of sample material may be an issue when analysis of tissues of limited availability is intended.

6.5 CELLULAR LOCALIZATION OF IDENTIFIED IMPS

To test whether our multi-pronged strategy is able to access IMPs in membranes of all major cellular compartments we manually filtered the list of 1006 identified IMPs according to their GO annotations (see **Table 5.5** for the list of GO annotation used in the filtering). The combined approach identified IMPs associated with plasma membrane (16.6% of identified IMPs), endoplasmic reticulum and Golgi apparatus (ER and GA, 12.1%) and mitochondria (11.2%). IMPs associated with membranes of other cellular compartments such as vesicles, nucleus, peroxisomes, lysosomes and endosomes were also found.

We then looked at the capabilities of each individual method to capture plasma membrane, ER/GA and mitochondrial IMPs. The results in **Table 5.6** confirmed that each individual method is effective in isolation of IMPs of membranes of major cellular compartments. It also showed that there is a preference in both N-glyco-capture

methods for IMPs associated with plasma membrane. This preference can be attributed to the fact that both SPEG and N-glyco-FASP are methods selectively targeting N-glycosylated proteins. This suggests that N-glycosylation is more prominent in plasma membrane IMPs than in IMPs of other compartments.

The results also show that there is a significant increase in the relative share of plasma membrane IMPs in N-glyco-FASP TL dataset. This correlates with the higher number of deamidated peptides in same dataset (**Table 5.4**) suggesting that TL is richer in plasma membrane N-glycoproteins compared to MF and it also correlates with lower hydrophobicity of identified peptides in this dataset (**Figure 5.5**). One of the possible explanations is that peptides in the MF samples are more hydrophobic than peptides in TL samples. More hydrophobic peptides can then non-specifically interact with lectins and block binding spots for the actual IMPs resulting in a lower count of IMPs in MF samples. The absence of the increased enrichment of plasma membrane IMPs in SPEG TL dataset could be attributed to the different chemistry and consumables (hydrazide and agarose beads) utilized in SPEG, but we are not able to conclude the exact cause of this difference.

The second explanation could be that some of the identified plasma membrane IMPs in the TL dataset were, in fact, false-positively identified as IMPs. Among the 73 plasma membrane proteins we found 9 proteins with GO annotation for extracellular space (GO: 0005615), that is defined as "part of a multicellular organism outside the cells proper, usually taken to be outside the plasma membranes, and occupied by fluid". The TMHMM false-positively recognizes such extracellular (excreted) proteins as IMPs because they also possess the signal peptide sequence, a sequence of usually 20-30 amino acid residues for protein translocation (von Heijne, 1990) resulting in wrong TMHs prediction, i.e. false-positive prediction of IMPs. The extracellular proteins can also reside for some time on the plasma membrane and as such may have received the "plasma membrane" GO annotation. Without these 9 extracellular proteins the plasma membrane IMPs in N-glyco-FASP TL dataset account for 23%. The updated proportion of identified plasma membrane IMPs without proteins with the "extracellular space" GO annotation is displayed in **Table 6.1**. Nevertheless, even with the updated shares of plasma membrane IMPs, the N-glyco-FASP TL still shows the highest enrichment of plasma membrane IMPs.

Dataset	Initial share of plasma membrane IMPs	Number of identified extracellular proteins	Updated proportion of plasma membrane IMPs
N-glyco-FASP MF	17%	5	16%
N-glyco-FASP TL	27%	9	23%
SPEG MF	17%	5	16%
SPEG TL	15%	4	14%
hpTC	10%	1	9%
SDC-trypsin	10%	1	10%

Table 6.1: Updated proportion of identified plasma membrane IMPs after the exclusion of potential extracellular proteins

The hpTC, on the other hand, results in identification of higher proportion of mitochondrial IMPs and lower share of plasma membrane-bound IMPs, compared to both N-glyco-capture strategies.

This could be attributed to the fact that the surface of mitochondrial membranes in myocardial tissue is most likely significantly higher compared to the surface of the plasma membrane due to the fact that at least 30% of the cardiomyocyte cell volume is occupied by mitochondria (Piquereau *et al*., 2013). The hpTC then selectively targets the hydrophobic portion of IMPs resulting in higher numbers of identified mitochondrial IMPs compared to other methods. This is also supported by the fact that in the non-selective SDC-trypsin dataset, the mitochondrial IMPs accounted for 18% (also a higher proportion than plasma membrane and ER/GA IMPs). The SDC-trypsin dataset could, in this case, serve as a "control". Both N-glyco-capture methods showed lower numbers of mitochondrial IMPs compared to SDC-trypsin, likely because of the preference for enriched N-glycosylated plasma membrane IMPs. In hpTC and SDC-trypsin no such N-glycoprotein enrichment was used. In hpTC and SDC-trypsin the difference in copy numbers of mitochondrial and plasma membrane IMPs in the tissue likely led to the higher share of identified mitochondrial IMPs. Hence, non-targeted SDC-trypsin, as well as hpTC, identified more mitochondrial IMPs compared to plasma membrane IMPs.

6.5.1 FUNCTIONAL ANALYSIS OF IDENTIFIED IMPS

Among the identified 1006 IMPs, we found proteins associated with plasma membrane, membranes of mitochondria, endoplasmic reticulum, Golgi apparatus, nucleus and membranes of vesicles, according to their GO annotations.

We subjected the identified IMPs to functional annotation using the Kyoto Encyclopedia of Genes and Genomes (KEGG) via Database for Annotation, Visualization and Integrated Discovery (DAVID). The functional analysis revealed that among identified IMPs were receptors, transporters, adhesive molecules, proteins with enzymatic activity, proteins playing a role in metabolism, or proteins associated with the immune system. The plasma membrane was represented by 19 CD (cluster of differentiation) antigens, 44 G-protein coupled receptors and various transporter and channel proteins. Mitochondrial IMPs included 39 proteins associated with oxidative phosphorylation including subunits of all four electron transport chain complexes and ATP synthase, TIM translocase subunits, regulators of apoptosis BAK and Bcl-2 protein. Among the IMPs associated with endoplasmic reticulum we found proteins playing a role in calcium metabolism and calcium-mediated pathways, sarcoplasmic/endoplasmic reticulum ATPases and ER to Golgi transport vesicular proteins. The Golgi apparatus proteins included, among others, SNAP receptor complex members and Syntaxins.

We also looked at IMPs functionally associated with heart physiology or pathophysiology. This way we identified 21 IMPs playing a role in arrhythmogenic right ventricular cardiomyopathy (ARVC) including cadherin 2 (Mayosi *et al*., 2017), desmoglein 2 (Awad *et al*., 2006) and desmocollin 2 (Syrris *et al*., 2006) which play a direct role in the pathogenesis of ARVC. 17 proteins related to dilated cardiomyopathy were represented by dystroglycan 1, emerin, sarcoglycans α-δ (McNally and Mestroni, 2017) which all have a role in the pathogenesis of this disease. 16 proteins were associated with hypertrophic cardiomyopathy. These proteins included calcium voltage-gated channel subunit alpha1 C (Boczek *et al*., 2015) and ryanodine receptor 2 (Landstrom and Ackerman, 2012), both having a role in the development of hypertrophic cardiomyopathy. Proteins with a role in physiology of the heart muscle included 20 proteins involved in cardiac muscle contraction such as cytochrome b, 6 subunits of cytochrome c oxidase and 3 subunits of ubiquinol-cytochrome c reductase, 27 proteins playing role in calcium-mediated signaling pathways such as plasma membrane calcium-transporting ATPase 1-4, sarcoplasmic/endoplasmic reticulum Ca^{2+} transporting

ATPase 1-3, inositol 1,4,5-trisphosphate receptor 1 or platelet-derived growth factor receptor beta (PDGFRβ). We also identified 37 cell adhesion molecules including 4 integrin subunits and cadherin 2 and 5.

6.6 COMPARISON OF TWO METHODS OF MEMBRANE FRACTION PREPARATION BY WESTERN BLOT ANALYSIS OF SELECTED PROTEINS

To evaluate the relative efficiency of two membrane enrichment strategies used as initial steps in our proteomic workflows, we verified the relative enrichment of 7 different proteins corresponding to different cellular compartments - 3 IMPs (caveolin 1, connexin 43 and transferrin receptor protein 1 (TfR1)), 2 cytoskeletal proteins (α-tubulin and actin) and 2 major soluble cytosolic proteins (ferritin and GAPDH) - by western blot analysis (**Figure 5.7B** and **C**).

The western blot analysis confirmed the enrichment of selected membrane proteins in analyzed membrane fractions compared to total tissue lysate. It also suggested that there are quantitative differences in the enrichment of membrane material between the different enrichment techniques or buffers used for homogenization. These differences can be best seen on blots of Caveolin-1 and Connexin 43, where the membrane fractions prepared in a hypertonic buffer (according to Nielsen *et al*., 2005) showed 36.5% and 25.3% higher enrichment compared to hypotonic membrane preparation (according to Vít *et al*., 2016), respectively. The relative enrichment of TfR1 was comparable in both membrane fractions.

However, both membrane enrichment techniques were only partially successful in reducing the amount of cytosolic proteins. The western blot analysis showed that the amount of ferritin was reduced by 57.8% in hypotonic fraction and by 68.5% in hypertonic fraction compared to the total tissue lysate.

The western blot verification of GAPDH led to more ambivalent results. As expected, GAPDH was present in total tissue lysate samples. It was partially removed from hypotonic fraction. However, it was not depleted from the hypertonic membrane fraction. This may be caused by used high-salt treatment during the preparation of membrane fraction. The high concentration of salt (2 M NaCl in our case) may lead to the salting out of some proteins resulting in precipitation of GAPDH during membrane

fraction preparation and subsequent co-enrichment of GAPDH alongside with IMPs. The same outcome was observed by Sawa *et al*. (1997) in the study of nuclear translocation where GAPDH failed to elute from nuclear pellet fractions when employing up to 5 M concentrations of NaCl.

The enrichment of membrane fraction by both methods also resulted in enrichment of cytoskeletal proteins as seen on blots for both actin and α-tubulin. One of the possible explanations is that many cytosolic proteins, cytoskeletal proteins included, are known to closely interact with plasma membrane lipids and with membrane proteins (Branton *et al*., 1981). Newer studies of membrane lipid rafts also show that cytoskeletal interaction plays an important role in the regulation of membrane lipid rafts (Haller *et al*., 2001; Yanagida *et al*., 2007; Head *et al*., 2014). This close interaction results in conjoined enrichment of both IMPs as well as cytoskeletal proteins which remain bound to IMPs. The second reason behind the cytoskeletal proteins co-enrichment is that cytoskeletal proteins are known to form supramolecular structures (actin forms actin filaments, α-tubulin forms microtubules). These structures of high molecular weight then sediment together with the membrane material resulting in conjoined enrichment.

The western blot analysis showed that both membrane enrichment strategies are effective in the enrichment of membrane material but are only partially efficient in depletion of cytosolic proteins. The western blot data also showed that the membrane fractions are still heavily contaminated by cytoskeletal proteins as well as some of the most abundant soluble proteins. The data also suggested that an employment of different methods of membrane fraction preparation can also be beneficial in the combined approach as it can further increase the complementarity of the individual methods of the combined approach.

6.7 SECRETS HIDDEN WITHIN MEMBRANE PROTEOME

Membrane proteins are generally under-represented in proteomic studies (Vít *et al*., 2019) and for this reason many IMPs have not yet been detected on protein level. Proteins that have not been detected on protein level are called “missing” and have recently become one of the interests of the Human Proteome Project. Missing proteins include proteins that were previously evidenced on transcript level (PE2 – protein evidence level 2), proteins inferred from homology (PE3) and proteins which existence was only theoretically predicted (PE4; Baker *et al*., 2017). The list of missing proteins, created

and maintained by neXtProt database, now (July 2019) contains 2705 missing proteins (Gaudet *et al*., 2017). However, the neXtProt database lists only human missing proteins. Other organisms, rat included, were not catalogued in such manner. Fortunately, the UniProt protein database provides such information, although the quality of the annotation of rat proteome is not nearly as complete as the annotation of the human proteome is.

We searched the list of identified 1006 IMPs against the rat subset of Uniprot database and found 250 proteins that were previously observed only on transcript level (PE2), the existence of 104 proteins was inferred from homology (PE3) and the existence of 99 proteins was only predicted (PE4). We cannot call these 453 proteins "missing" as the term officially only implies to human proteins but it provides us with a hint of how much information may be hidden in the phospholipid bilayer.

The data presented in this work, as well as the data presented by Vít *et al*. (2019), show many benefits of the multi-pronged strategy. It seems that a combination of strategies that are specifically targeting different parts of IMPs together with the non-targeted "classical" approach of SDC-trypsin may lead into uncovering of some of the most hidden parts of membrane proteome.

6.8 PERSPECTIVES OF THE MULTI-PRONGED APPROACH

The work presented in this thesis is, to our knowledge, one of the most complete analyses of the myocardial membrane proteome up to date. In general, a deep knowledge of a membrane proteome is desired in a research for novel biomarkers as well as new therapeutic targets. We believe that the multi-pronged approach presented in this thesis is a valuable tool that could accelerate the search for both biomarkers and novel drug targets, and could, eventually, help to decrease the alarming mortality numbers of cardiovascular diseases.

7. CONCLUSIONS

The main goal of this thesis was to uncover as large as possible portion of the rat myocardial membrane proteome using four different methods targeting IMPs of different physico-chemical properties (SPEG, N-glyco-FASP, hpTC and SDC-trypsin) and to determine if the used methods are complementary to each other.

The combined approach led to the identification of total 1006 unique rat myocardial IMPs. We performed an extensive bioinformatic analysis of the data which confirmed the complementarity of used methods on several levels: Each of the individual methods of the multi-pronged approach provided a significant portion of unique IMP identifications. The hpTC showed a preference for hydrophobic IMPs on both protein and peptide level compared to the remaining methods. The hpTC was also more efficient in the identification of IMPs with more than one transmembrane domain compared to SPEG, N-glyco-FASP and SDC-trypsin that were more effective for single-span IMPs. The complementarity was also reflected on the cellular localization of identified IMPs. We showed that hpTC and SDC-trypsin identified higher proportion of mitochondrial IMPs while both N-glyco-capture methods were more effective for the identification of plasma membrane IMPs.

Nevertheless, it must be noted, that each method was performed only in duplicates. More technical replicates of the methods would probably lead to higher numbers of uniquely identified IMPs by each individual method which could affect the conclusions to some extent. However, the complementarity of used methods was already confirmed by the "Pitchfork" study (Vít *et al*., 2019).

The second goal of this work was to determine the impact of used starting material (membrane fraction and tissue lysate) on the number of identified IMPs in used N-glyco-capture methods (SPEG and N-glyco-FASP).

We concluded that the preparation of membrane fraction results in significantly higher numbers of identified IMPs in both SPEG (38% higher compared to the tissue lysate) and N-glyco-FASP (15% higher compared to the tissue lysate). This suggests that preparation of membrane fraction should be considered as a standard step in the analysis of membrane proteome when using SPEG or N-glyco-FASP in membrane proteomic workflow.

The third goal of this thesis was to determine the efficiency of two membrane fraction preparation techniques, the first isolated after lysis in hypotonic buffer with carbonate washes according to Vít *et al.* (2016), and the second isolated after lysis in hypertonic buffer with carbonate washes according to Nielsen at al. (2005).

The results showed that both protocols for preparation of membrane fraction were efficient in the enrichment of IMPs, more so in the case of the high salt buffer technique. However, the membrane-enriched fractions were also heavily contaminated by major cytoskeletal proteins, especially in the case of the high salt buffer technique. The results also showed that the removal of major soluble cytosolic proteins is mostly efficient, but as showed on the example of GAPDH with the high salt buffer technique, can also lead to their undesired enrichment.

We believe that this work provided deeper insight into the rat myocardial proteome and that the data presented in this thesis could help to speed up the process of finding both novel biomarkers as well as novel drug targets of cardiovascular diseases.

8. REFERENCES

Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, 422(6928), 198-207.

Antharavally, B., Mallia, K., Rosenblatt, M., Salunkhe, A., Rogers, J., Haney, P., & Haghdoost, N. (2011). Efficient removal of detergents from proteins and peptides in a spin column format. *Analytical Biochemistry*, 416(1), 39-44.

Apweiler, R., Hermjakob, H., & Sharon, N. (1999): On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et biophysica acta*, 1473, 4-8.

Awad, M. M., Dalal, D., Cho, E., Amat-Alarcon, N., James, C., Tichnell, C., Tucker A, Russell, S. D., Bluemke, D. A., Dietz, H. C., Calkins, H., & Judge, D. P. (2006). DSG2 mutations contribute to arrhythmogenic right ventricular dysplasia/cardiomyopathy. *American journal of human genetics*, *79*(1), 136–142.

Baker, M. S., Ahn, S. B., Mohamedali, A., Islam, M. T., Cantor, D., Verhaert, P. D., Fanayan, S., Sharma, S., Nice, E. C, Connor, M., & Ranganathan, S. (2017). Accelerating the search for the missing proteins in the human proteome. *Nature Communications*, 8(14271), 1-13.

Bakry, N., Kamata, Y., & Simpson, L. (1991). Lectins from Triticum vulgaris and Limax flavus are universal antagonists of botulinum neurotoxin and tetanus toxin. *Journal of Pharmacology and Experimental Therapeutics*, 258(3), 830-836.

Becker, J., Reeke, G., Wang, J., Cunningham, B., & Edelman, G. (1974). The covalent and three-dimensional structure of concanavalin A. III. Structure of the monomer and its interactions with metals and saccharides. *Journal of Biological Chemistry*, 250, 1513-1524.

Bers, D. M., Eisner, D. A, & Valdivia, H. H. (2003). Sarcoplasmic Reticulum Ca^{2+} and heart failure: roles of diastolic leak and Ca^{2+} transport. *Circulation Research*, 93(6), 487–490.

Betancourt, L. H., Sanchez, A., Pla, I., Kuras, M., Zhou, Q., Andersson, R., & Marko-Varga, G. (2018). Quantitative assessment of urea in-solution Lys-C/trypsin digestions reveals superior performance at room temperature over traditional proteolysis at 37 °C. *Journal of Proteome Research*, 17(7), 2556–2561.

Bian, Y., Song, C., Cheng, K., Dong, M., Wang, F., Huang, J., Sun, D., Wang, L., Ye, M., & Zou, H. (2014). An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *Journal of Proteomics*, 96, 253-262.

Blackler, A. R., Speers, A. E., Ladinsky, M. S., & Wu, C. C. (2008). A shotgun proteomic method for the identification of membrane-embedded proteins and peptides. *Journal of Proteome Research*, 7(7), 3028–3034.

Blonder, J., Conrads, T. P., Yu, L.-R., Terunuma, A., Janini, G. M., Issaq, H. J., Vogel, J. C., & Veenstra, T. D. (2004). A detergent- and cyanogen bromide-free method for integral membrane proteomics: Application to Halobacterium purple membranes and the human epidermal membrane proteome. *Proteomics*, 4(1), 31–45.

Boczek, N. J., Ye, D., Jin, F., Tester, D. J., Huseby, A., Bos, J. M., Johnson, A. J., Kanter, R., & Ackerman, M. J. (2015). Identification and functional characterization of a novel CACNA1C-mediated cardiac disorder characterized by prolonged QT intervals with hypertrophic cardiomyopathy, congenital heart defects, and sudden cardiac death. *Circulation. Arrhythmia and electrophysiology*, 8(5), 1122–1132.

Bosserhoff, A., Wallach, J., & Frank, R. (1989). Micropreparative separation of peptides derived from sodium dodecyl sulphate-solubilized proteins. *Journal of Chromatography A*, 473, 71-77.

Bowie, J. U. (1997). Helix packing in membrane proteins. *Journal of Molecular Biology*, 272(5), 780–789.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1-2), 248–254.

Branton, D., Cohen, M. C., & Tyler, J. (1981). Interaction of cytoskeletal proteins on the human erythrocyte membrane. *Cell*, 24(1), 24–32.

Bryk, A., & Wiśniewski, J. (2017). Quantitative analysis of human red blood cell proteome. *Journal of Proteome Research*, 16(8), 2752-2761.

Bunkenborg, J., Pilch, B. J., Podtelejnikov, A. V., & Wiśniewski, J. R. (2004). Screening forN-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics*, 4(2), 454–465.

Burns, J., Butler, J., Moran, J., & Whitesides, G. (1991). Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *The Journal of Organic Chemistry*, 56(8), 2648-2650.

Cañas, B., Piñeiro, C., Calvo, E., López-Ferrer, D., & Gallardo, J. M. (2007). Trends in sample preparation for classical and second generation proteomics. *Journal of Chromatography A*, 1153(1-2), 235–258.

Catherman, A. D., Skinner, O. S., & Kelleher, N. L. (2014). Top Down proteomics: Facts and perspectives. *Biochemical and Biophysical Research Communications*, 445(4), 683–693.

Chandler, K. B., & Costello, C. E. (2016). Glycomics and glycoproteomics of membrane proteins and cell-surface receptors: Present trends and future opportunities. *Electrophoresis*, 37(11), 1407–1419.

Chen, E., Cociorva, D., Norris, J., & Yates, J. (2007). Optimization of Mass Spectrometry-Compatible Surfactants for Shotgun Proteomics. *Journal of Proteome Research*, 6(7), 2529-2538.

Cleland, W. (1964). Dithiothreitol, a new protective reagent for SH groups. *Biochemistry*, 3(4), 480-482.

Cox, H. D., Chao, C.-K., Patel, S. A., & Thompson, C. M. (2008). Efficient digestion and mass spectral analysis of vesicular glutamate transporter 1: a recombinant membrane protein expressed in yeast. *Journal of Proteome Research*, 7(2), 570–578.

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26, 1367-1372.

Crimmins, D., Mische, S., & Denslow, N. (2005). Chemical cleavage of proteins in solution. *Current Protocols in Protein Science*, 11.4.1-11.4.11.

Da Cruz, S., Xenarios, I., Langridge, J., Vilbois, F., Parone, P. A., & Martinou, J.-C. (2003). Proteomic analysis of the mouse liver mitochondrial inner membrane. *Journal of Biological Chemistry*, 278(42), 41566–41571.

Deeb, S. J., Cox, J., Schmidt-Supprian, M., & Mann, M. (2013). N-linked glycosylation enrichment for in-depth cell surface proteomics of diffuse large B-cell lymphoma subtypes. *Molecular & Cellular Proteomics*, 13(1), 240–251.

Doll, S., Dreßen, M., Geyer, P., Itzhak, D., Braun, C., Doppler, S., Meier, F., Deutsch, M., Lahm, H., Lange, R., Krane, M., & Mann, M. (2017). Region and cell-type resolved quantitative proteomic map of the human heart. *Nature Communications*, 8(1).

Doll, S., Dreßen, M., Geyer, P., Itzhak, D., Braun, C., Doppler, S., Meier, F., Deutsch, M., Lahm, H., Lange, R., Krane, M., & Mann, M. (2017). Region and cell-type resolved quantitative proteomic map of the human heart. *Nature Communications*, 8(1).

Dormeyer, W., van Hoof, D., Mummery, C. L., Krijgsveld, J., & Heck, A. J. R. (2008). A practical guide for the identification of membrane and plasma membrane proteins in human embryonic stem cells and human embryonal carcinoma cells. *Proteomics*, 8(19), 4036–4053.

Doucette, A. A., Vieira, D. B., Orton, D. J., & Wall, M. J. (2014). Resolubilization of precipitated intact membrane proteins with cold formic acid for analysis by mass spectrometry. *Journal of Proteome Research*, 13(12), 6001–6012.

Drake, P., Schilling, B., Niles, R., Braten, M., Johansen, E., Liu, H., Lerch, M., Sorensen, D., Li, B., Allen, S., Hall, S., Witkowska, H., Regnier, F., Gibson, B., & Fisher, S. (2011). A lectin affinity workflow targeting glycosite-specific, cancer-related carbohydrate structures in trypsin-digested human plasma. *Analytical Biochemistry*, 408(1), 71-85.

Drapeau, G., Boily, Y., & Houmard, J. (1971). Purification and properties of an extracellular protease of Staphylococcus aureus. *Journal of Biological Chemistry*, 227(20), 6720-6726.

Dupuy, A. D., & Engelman, D. M. (2008). Protein area occupancy at the center of the red blood cell membrane. *Proceedings of the National Academy of Sciences*, 105(8), 2848–2852.

Eng, J., McCormack, A., & Yates, J. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 5(11), 976-989.

Estigoy, C. B., Pontén, F., Odeberg, J., Herbert, B., Guilhaus, M., Charleston, M., Ho, J. W. K., Cameron, D., & dos Remedios, C. G. (2009). Intercalated discs: multiple proteins perform multiple functions in non-failing and failing human hearts. *Biophysical Reviews*, 1(1), 43–49.

Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M., & Berglund, L. (2010). Prediction of the human membrane proteome. *Proteomics*, 10(6), 1141–1149.

Fenn, J., Mann, M., Meng, C., Wong, S., & Whitehouse, C. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926), 64–71.

Fernandes, H., Cohen, S., & Bishayee, S. (2000). Glycosylation-induced conformational modification positively regulates receptor-receptor association. *Journal of Biological Chemistry,* 276(7), 5375–5383.

Fischer, F., & Poetsch, A. (2006). Protein cleavage strategies for an improved analysis of the membrane proteome. *Proteome Science*, 4(1), 2.

Fujiki, Y. (1982). Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. *The Journal of Cell Biology*, 93(1), 97-102.

Gardini, S., Cheli, S., Baroni, S., Di Lascio, G., Mangiavacchi, G., Micheletti, N., Monaco, C., Savini, L., Alocci, D., Mangani, S., & Niccolai, N. (2016). On nature's strategy for assigning genetic code multiplicity. *PLOS ONE*, 11(2), e0148174.

Gauci, S., Helbig, A. O., Slijper, M., Krijgsveld, J., Heck, A. J. R., & Mohammed, S. (2009). Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Analytical Chemistry*, 81(11), 4493–4501.

Gaudet, P., Michel, P. A., Zahn-Zabal, M., Britan, A., Cusin, I., Domagalski, M., Duek, P. D., Gateau, A., Gleizes, A., Hinard, V., Rech de Laval, V., Lin, J., Nikitin, F., Schaeffer, M., Teixeira, D., Lane, L., & Bairoch, A. (2017). The neXtProt knowledgebase on human proteins: 2017 update. *Nucleic acids research*, 45(D1), D177–D182.

Giansanti, P., Tsiatsiani, L., Low, T. Y., & Heck, A. J. R. (2016). Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nature Protocols*, 11(5), 993–1006.

Gloriam, D. E., Fredriksson, R., & Schiöth, H. B. (2007). The G protein-coupled receptor subset of the rat genome. *BMC genomics*, 8, 338.

Granseth, E., von Heijne, G., & Elofsson, A. (2005). A study of the membrane–water interface region of membrane proteins. *Journal of Molecular Biology*, 346(1), 377-385.

Gregorich, Z. R., Chang, Y. H., & Ge, Y. (2014). Proteomics in heart failure: top-down or bottom-up? *Pflugers Archiv: European journal of physiology*, 466(6), 1199–1209.

Gross, E., & Witkop, B. (1961). Selective cleavage of the methionyl peptide bonds in ribonuclease with cyanogen bromide. *Journal of the American Chemical Society*, 83(6), 1510–1511.

Gstaiger, M., & Aebersold, R. (2009). Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nature Reviews Genetics*, 10(9), 617–627.

Haller, P. D. von, Donohoe, S., Goodlett, D. R., Aebersold, R., & Watts, J. D. (2001). Mass spectrometric characterization of proteins extracted from Jurkat T cell detergent-resistant membrane domains. *Proteomics*, 1(8), 1010–1021.

Head, B. P., Patel, H. H., & Insel, P. A. (2014). Interaction of membrane/lipid rafts with the cytoskeleton: impact on signaling and function: membrane/lipid rafts, mediators of cytoskeletal arrangement and cell signaling. *Biochimica et biophysica acta*, 1838(2), 532–545.

Hemström, P., & Irgum, K. (2006). Hydrophilic interaction chromatography. *Journal of Separation Science*, 29(12), 1784–1821.

Hildebrand, P., Preissner, R., & Frömmel, C. (2004). Structural features of transmembrane helices. *FEBS Letters*, 559(1-3), 145-151.

Hilgemann, D. W. (2004). New insights into the molecular and cellular workings of the cardiac Na^{+}/Ca^{2+} exchanger. *American Journal of Physiology*, 287, C1167–C11.72

Hofmann, A., Gerrits, B., Schmidt, A., Bock, T., Bausch-Fluck, D., Aebersold, R., & Wollscheid, B. (2010). Proteomic cell surface phenotyping of differentiating acute myeloid leukemia cells. *Blood*, 116(13), e26–e34.

Horigome, T., & Sugano, H. (1984). Rapid method for removal of detergents salts from protein solutions using toyopearl HW-40F. *Journal of Chromatography A*, 283, 315–322.

Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., & Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. *Journal of Mass Spectrometry*, 40(4), 430–443.

Hustoft, H. K., Reubsaet, L., Greibrokk, T., Lundanes, E., & Malerod, H. (2011). Critical assessment of accelerating trypsination methods. *Journal of Pharmaceutical and Biomedical Analysis*, 56(5), 1069–1078.

Illergård, K., Kauko, A., & Elofsson, A. (2010). Why are polar residues within the membrane core evolutionary conserved? *Proteins: Structure, Function, and Bioinformatics*, 79(1), 79–91.

Kaiser, R. and Metzka, L. (1999). Enhancement of cyanogen bromide cleavage yields for methionyl-serine and methionyl-threonine peptide bonds. *Analytical Biochemistry*, 266(1), 1-8.

Karas, M., Bachmann, D., & Hillenkamp, F. (1985). Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry*, 57(14), 2935-2939.

Kaszuba, K., Grzybek, M., Orłowski, A., Danne, R., Róg, T., Simons, K., Coskun, Ü., & Vattulainen, I. (2015). N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes. *Proceedings of the National Academy of Sciences*, 112(14), 4334-4339.

Kellie, J., Tran, J., Lee, J., Ahlf, D., Thomas, H., Ntai, I., Catherman, A., Durbin, K., Zamdborg, L., Vellaichamy, A., Thomas, P., & Kelleher, N. (2010). The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. *Molecular BioSystems*, 6(9), 1532.

Khan, H. M., He, T., Fuglebakk, E., Grauffel, C., Yang, B., Roberts, M. F., Gershenson, A., & Reuter, N. (2016). A role for weak electrostatic interactions in peripheral membrane protein binding. *Biophysical Journal*, 110(6), 1367–1378.

Kollipara, L., & Zahedi, R. P. (2013). Protein carbamylation: In vivo modification or in vitro artefact? *Proteomics*, 13(6), 941–944.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3), 567–580.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1), 105–132.

Lakshmanan, R., & Loo, J. (2019). Top-down protein identification using a time-of-flight mass spectrometer and data independent acquisition. *International Journal of Mass Spectrometry*, 435, 136-144.

Landstrom, A. P., & Ackerman, M. J. (2012). Beyond the cardiac myofilament: hypertrophic cardiomyopathy- associated mutations in genes that encode calcium-handling proteins. *Current molecular medicine*, 12(5), 507–518.

Li, A., Sowder, R., Henderson, L., Moore, S., Garfinkel, D., & Fisher, R. (2001). Chemical cleavage at aspartyl residues for protein identification. *Analytical Chemistry*, 73(22), 5395-5402.

Lin, Y., Zhou, J., Bi, D., Chen, P., Wang, X., & Liang, S. (2008). Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Analytical Biochemistry*, 377(2), 259-266.

Liu, S., Xia, Y., Liu, X., Wang, Y., Chen, Z., Xie, J., Qian, K., Shen, H., & Yang, P. (2017). In-depth proteomic profiling of left ventricular tissues in human end-stage dilated cardiomyopathy. *Oncotarget*, *8*(29), 48321–48332.

Loo, R. R. O., & Loo, J. A. (2007). Matrix-assisted laser desorption/ionization-mass spectrometry of hydrophobic proteins in mixtures using formic acid, perfluorooctanoic acid, and sorbitol. *Analytical Chemistry*, 79(3), 1115–1125.

Loo, R. R. O., Dales, N., & Andrews, P. C. (1994). The effect of detergents on proteins analyzed by electrospray ionization. protein and peptide analysis by mass spectrometry, *Methods in Molecular Biology* 61, 141–160.

Lowry, O., Rosebrough, N., Farr, A., & Randall, R. (1951). Protein measurement with the folin phenol reagent. *Journal of Biological Chemistry*, 193, 265.

Marks A. R. (2003). Calcium and the heart: a question of life and death. *The Journal of clinical investigation*, 111(5), 597–600.

Marshall, A. G., Hendrickson, C. L., & Jackson, G. S. (1998). Fourier transform ion cyclotron resonance mass spectrometry: A primer. *Mass Spectrometry Reviews*, 17(1), 1–35

Masuda, T., Tomita, M., & Ishihama, Y. (2008). Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *Journal of Proteome Research*, 7(2), 731–740.

Mayosi, B. M., Fish, M., Shaboodien, G., Mastantuono, E., Kraus, S., Wieland, T., Kotta, M. C., Chin, A., Laing, N., Ntusi, N. B., Chong, M.m Horsfall, C., Pimstone, S. N., Gentilini, D., Parati, G., Strom, T. M., Meitinger, T., Pare, G., Schwartz, P. J., & Crotti, L. (2017). Identification of Cadherin 2 (CDH2) mutations in arrhythmogenic right ventricular cardiomyopathy. *Circulation: Cardiovascular Genetics*, 10(2), e001605.

McNally, E. M., & Mestroni, L. (2017). Dilated Cardiomyopathy. *Circulation Research*, 121(7), 731–748.

Meyer, J. G., Kim, S., Maltby, D. A., Ghassemian, M., Bandeira, N., & Komives, E. A. (2014). Expanding proteome coverage with orthogonal-specificity α-lytic proteases. *Molecular & cellular proteomics: MCP*, *13*(3), 823–835.

Moest, H., Frei, A. P., Bhattacharya, I., Geiger, M., Wollscheid, B., & Wolfrum, C. (2013). Malfunctioning of adipocytes in obesity is linked to quantitative surfaceome changes. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1831(7), 1208–1216.

Monje-Galvan, V., & Klauda, J. B. (2016). Peripheral membrane proteins: Tying the knot between experiment and computation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858(7), 1584–1593.

Mosterd, A., & Hoes, A. W. (2007). Clinical epidemiology of heart failure. *Heart (British Cardiac Society)*, 93(9), 1137–1146.

Müller, T., & Winter, D. (2017). Systematic evaluation of protein reduction and alkylation reveals massive unspecific side effects by iodine-containing reagents. *Molecular & Cellular Proteomics*, 16(7), 1173-1187.

Nag, A. C., & Zak, R. (1979). Dissociation of adult mammalian heart into single cell suspension: an ultrastructural study. *Journal of anatomy*, 129(3), 541–559.

Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., & Mann, M. (2011). Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular systems biology*, 7, 548.

Neurath, H. (1994). Proteolytic enzymes past and present: The second golden era. *Protein Science*, 3(10), 1734-1739.

Nielsen, P. A., Olsen, J. V., Podtelejnikov, A. V., Andersen, J. R., Mann, M., & Wiśniewski, J. R. (2005). Proteomic mapping of brain plasma membrane proteins. *Molecular & Cellular Proteomics*, 4(4), 402–408.

Noble, J., Knight, A., Reason, A., Di Matola, A., & Bailey, M. (2007). A comparison of protein quantitation assays for biopharmaceutical applications. *Molecular Biotechnology*, 37(2), 99-111.

Olsen, J., Ong, S., & Mann, M. (2004). Trypsin cleaves exclusively c-terminal to arginine and lysine residues. *Molecular & Cellular Proteomics*, 3(6), 608-614.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics*, 1(5), 376–386.

Ongay, S., Boichenko, A., Govorukhina, N., & Bischoff, R. (2012). Glycopeptide enrichment and separation for protein glycosylation analysis. *Journal of Separation Science*, 35(18), 2341–2372.

Palmisano, G., Melo-Braga, M. N., Engholm-Keller, K., Parker, B. L., & Larsen, M. R. (2012). Chemical deamidation: a common pitfall in large-scale n-linked glycoproteomic mass spectrometry-based analyses. *Journal of Proteome Research*, 11(3), 1949–1957.

Paul, W. (1990). Electromagnetic traps for charged and neutral particles. *Reviews of Modern Physics*, 62(3), 531–540.

Perkins, D., Pappin, D., Creasy, D., & Cottrell, J. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18), 3551-3567.

Piquereau, J., Caffin, F., Novotova, M., Lemaire, C., Veksler, V., Garnier, A., Ventura-Clapier, R., & Joubert, F. (2013). Mitochondrial dynamics in the adult cardiomyocytes: which roles for a highly specialized cell? *Frontiers in physiology*, *4*, 102.

Pocock, S., Ariti, C., McMurray, J., Maggioni, A., Køber, L., Squire, I., Swedberg, K., Dobson, J., Poppe, K., Whalley, G., & Doughty, R. (2012). Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *European Heart Journal*, 34(19), 1404-1413.

Rahali, V., & Gueguen, J. (1999). Chemical cleavage of bovine beta-lactoglobulin by BNPS-skatole for preparative purposes: comparative study of hydrolytic procedures and peptide characterization. *Journal of Protein Chemistry*, 18(1), 1–12.

Raimondo, F., Corbetta, S., Savoia, A., Chinello, C., Cazzaniga, M., Rocco, F., Bosari, S., Grasso, M., Bovo, G., Magni, F., & Pitto, M. (2015). Comparative membrane proteomics: a technical advancement in the search of renal cell carcinoma biomarkers. *Molecular BioSystems*, 11(6), 1708-1716.

Regard, J. B., Sato, I. T., & Coughlin, S. R. (2008). Anatomical profiling of G protein-coupled receptor expression. *Cell*, 135(3), 561–571.

Russell, W. K., Park, Z.-Y., & Russell, D. H. (2001). Proteolysis in mixed organic−aqueous solvent systems: applications for peptide mass mapping using mass spectrometry. *Analytical Chemistry*, 73(11), 2682–2685.

Sawa, A., Khan, A. A., Hester, L. D., & Snyder, S. H. (1997). Glyceraldehyde-3-phosphate dehydrogenase: nuclear translocation participates in neuronal and nonneuronal cell death. *Proceedings of the National Academy of Sciences of the United States of America*, 94(21), 11669–11674.

Schroeder, M., Shabanowitz, J., Schwartz, J., Hunt, D., & Coon, J. (2004). A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Analytical Chemistry*, 76(13), 3590-3598.

Sebastião, M. J., Pereira, R., Serra, M., Gomes-Alves, P., & Alves, P. M. (2018). Unveiling human cardiac fibroblast membrane proteome. *Proteomics*, 18(13), 1700446.

Seddon, A. M., Curnow, P., & Booth, P. J. (2004). Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1666(1-2), 105–117.

Shi, Y., Xiang, R., Horváth, C., & Wilkins, J. A. (2004). The role of liquid chromatography in proteomics. *Journal of Chromatography A*, 1053(1-2), 27–36.

Simon, L. M., Kotormán, M., Garab, G., & Laczkó, I. (2001). Structure and Activity of α-chymotrypsin and trypsin in aqueous organic media. *Biochemical and Biophysical Research Communications*, 280(5), 1367–1371.

Simon, L. M., László, K., Vértesi, A., Bagi, K., & Szajáni, B. (1998). Stability of hydrolytic enzymes in water-organic solvent systems. *Journal of Molecular Catalysis B: Enzymatic*, 4(1-2), 41–45.

Smith, P., Krohn, R., Hermanson, G., Mallia, A., Gartner, F., Provenzano, M., Fujimoto, E., Goeke, N., Olson, B., & Klenk, D. (1985). Measurement of protein using bicinchoninic acid. *Analytical Biochemistry*, 150(1), 76-85.

Speers, A. E., & Wu, C. C. (2007). Proteomics of integral membrane proteins theory and application. *Chemical Reviews*, 107(8), 3687–3714.

Sun, D., Wang, N., & Li, L. (2012). Integrated SDS removal and peptide separation by strong-cation exchange liquid chromatography for SDS-assisted shotgun proteome analysis. *Journal of Proteome Research*, 11(2), 818–828.

Swatkoski, S., Gutierrez, P., Ginter, J., Petrov, A., Dinman, J. D., Edwards, N., & Fenselau, C. (2007). Integration of residue-specific acid cleavage into proteomic workflows. *Journal of Proteome Research*, 6(11), 4525–4527.

Syrris, P., Ward, D., Evans, A., Asimaki, A., Gandjbakhch, E., Sen-Chowdhry, S., & McKenna, W. J. (2006). Arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2. *American journal of human genetics*, 79(5), 978–984.

Tian, Y., Kelly-Spratt, K. S., Kemp, C. J., & Zhang, H. (2008). Identification of glycoproteins from mouse skin tumors and plasma. *Clinical Proteomics*, 4(3-4), 117–136.

Tian, Y., Koganti, T., Yao, Z., Cannon, P., Shah, P., Pietrovito, L., Modesti, A., Aiyetan, P., DeLeon-Pennell, K., Ma, Y., Halade, G., Hicks, C., Zhang, H., & Lindsey, M. (2014). Cardiac extracellular proteome profiling and membrane topology analysis using glycoproteomics. *Proteomics - Clinical Applications*, 8(7-8), 595-602.

Tian, Y., Zhou, Y., Elliott, S., Aebersold, R., & Zhang, H. (2007). Solid-phase extraction of N-linked glycopeptides. *Nature Protocols*, 2(2), 334–339.

Tsiatsiani, L., & Heck, A. J. R. (2015). Proteomics beyond trypsin. *FEBS Journal*, 282(14), 2612–2626.

Ulmschneider, M. B., & Sansom, M. S. P. (2001). Amino acid distributions in integral membrane protein structures. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1512(1), 1–14.

Vandermarliere, E., Mueller, M., & Martens, L. (2013). Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrometry Reviews*, 32(6), 453-465.

Vit, O., & Petrak, J. (2017). Integral membrane proteins in proteomics. How to break open the black box? *Journal of Proteomics*, 153, 8–20.

Vit, O., Harant, K., Klener, P., Man, P., & Petrak, J. (2019). A three-pronged “Pitchfork” strategy enables an extensive description of the human membrane proteome and the identification of missing proteins. *Journal of Proteomics*, 204, 103411.

Vit, O., Man, P., Kadek, A., Hausner, J., Sklenar, J., Harant, K., Novak, P., Scigelova, M., Woffendin, G., & Petrak, J. (2016). Large-scale identification of membrane proteins based on analysis of trypsin-protected transmembrane segments. *Journal of Proteomics*, 149, 15-22.

Von Heijne, G. (1989). Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature*, 341(6241). 456-458.

Von Heijne, G. (1990). The signal peptide. *The Journal of Membrane Biology*, 115(3), 195–201.

Vuckovic, D., Dagley, L. F., Purcell, A. W., & Emili, A. (2013). Membrane proteomics by high performance liquid chromatography-tandem mass spectrometry: Analytical approaches and challenges. *Proteomics*, 13(3-4), 404–423.

Waas, M., Bhattacharya, S., Chuppa, S., Wu, X., Jensen, D., Omasits, U., Wollscheid, B., Volkman, B., Noon, K., & Gundry, R. (2014). Combine and conquer: surfactants, solvents, and chaotropes for robust mass spectrometry based analyses of membrane proteins. *Analytical Chemistry*, 86(3), 1551-1559.

Warburg, O., & Christian, W. (1941). Isolierung und kristallisation des Garungs-ferments Enolase. *Biochemische Zeitschrift*, 310, 384-421.

Washburn, M., Wolters, D., & Yates, J. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, 19(3), 242-247.

Wasinger, V., Cordwell, S., Cerpa-Poljak, A., Yan, J., Gooley, A., Wilkins, M., Duncan, M., Harris, R., Williams, K., & Humphery-Smith, I. (1995). Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis*, 16(1), 1090-1094.

Weickhardt, C., Moritz, F., & Grotemeyer, J. (1996). Time-of-flight mass spectrometry: State-of the-art in chemical analysis and molecular science. *Mass Spectrometry Reviews*, 15(3), 139–162.

Wiechelman, K., Braun, R., & Fitzpatrick, J. (1988). Investigation of the bicinchoninic acid protein assay: Identification of the groups responsible for color formation. *Analytical Biochemistry*, 175(1), 231-237.

Wilkins, E., Wilson, L., Wickramasinghe, K., Bhatnagar, P., Leal, J., Luengo-Fernandez, R., Burns, R., Rayner, M., & Townsend, N. (2017). European Cardiovascular Disease Statistics 2017. European Heart Network, Brussels.

Williams, D. A., Delbridge, L. M., Cody, S. H., Harris, P. J., & Morgan, T. O. (1992). Spontaneous and propagated calcium release in isolated cardiac myocytes viewed by confocal microscopy. *American Journal of Physiology-Cell Physiology*, 262(3), C731–C742.

Wimley W. C. (2002). Toward genomic identification of beta-barrel membrane proteins: composition and architecture of known structures. *Protein science : a publication of the Protein Society*, 11(2), 301–312.

Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, 6(5), 359–362.

Wiśniewski, J., Zougman, A., & Mann, M. (2009). Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *Journal of Proteome Research*, 8(12), 5674-5678.

Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Schiess, R., Aebersold, R., & Watts, J. D. (2009). Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature Biotechnology*, 27(4), 378–386.

Wu, F., Sun, D., Wang, N., Gong, Y., & Li, L. (2011). Comparison of surfactant-assisted shotgun methods using acid-labile surfactants and sodium dodecyl sulfate for membrane proteome analysis. *Analytica Chimica Acta*, 698(1-2), 36-43.

Yanagida, M., Nakayama, H., Yoshizaki, F., Fujimura, T., Takamori, K., Ogawa, H., & Iwabuchi, K. (2007). Proteomic analysis of plasma membrane lipid rafts of HL-60 cells. *Proteomics*, 7(14), 2398–2409.

Yang, H., & Zubarev, R. A. (2010). Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides. *Electrophoresis*, 31(11), 1764–1772.

Yau, W., Wimley, W., Gawrisch, K., & White, S. (1998). The preference of tryptophan for membrane interfaces. *Biochemistry*, 37(42), 14713-14718.

Yildirim, M. A., Goh, K.-I., Cusick, M. E., Barabási, A.-L., & Vidal, M. (2007). Drug-target network. *Nature Biotechnology*, 25(10), 1119–1126.

Yu, Y.-Q., Gilar, M., Lee, P. J., Bouvier, E. S. P., & Gebler, J. C. (2003). Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Analytical Chemistry*, 75(21), 6023–6028.

Zhang, H., Li, X., Martin, D. B., & Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature Biotechnology*, 21(6), 660–666.

Zhang, H., Yi, E., Li, X., Mallick, P., Kelly-Spratt, K., Masselon, C., Camp, D., Smith, R., Kemp, C., & Aebersold, R. (2004). High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Molecular & Cellular Proteomics*, 4(2), 144-155.

Zhou, J., Dann, G., Shi, T., Wang, L., Gao, X., Su, D., Nicora, C., Shukla, A., Moore, R., Liu, T., Camp II, D., Smith, R., & Qian, W. (2012). Simple Sodium dodecyl sulfate-assisted sample preparation method for LC-MS-based proteomics applications. *Analytical Chemistry*, 84(6), 2862-2867.

Zielinska, D., Gnad, F., Wiśniewski, J., & Mann, M. (2010). Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell*, 141(5), 897-907.