FACULTY OF MEDICINE AND HEALTH SCIENCES

New epitopes on the surface of cancer cells as targets for immunotherapy

Bram Van Rie

Student number: 00705826

Promotor: Prof. dr. Kris Gevaert,
Supervisor: dr. Esperanza Fernández
Co-supervisor: An Staes

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Master of Science in the Biomedical Sciences

Academic year: 2018 - 2019

GHENT UNIVERSITY

FACULTY OF MEDICINE AND HEALTH SCIENCES

Preface

This dissertation was written to obtain the degree Master of Science in Biomedical Sciences at Ghent University. A provisional stop in a long but interesting journey.

I do not have any well-developed talent, but I have a restless mind that cannot stop asking questions. The entire world is interesting and focussing on a single thing is definitely not my strong suit. Why travel from A to B in a straight line if there is an entire world of opportunities and scenic routes around it. You might occasionally end up in a dead end street, but that is always a nice opportunity to ask for directions and reflect on how you got there. I do realise this has to be exhausting for the people around me, especially if they are in a hurry to get somewhere, but therefore I would like to take advantage of the opportunity to thank everyone who helped in creating an environment in which I could quench my thirst for knowledge. In particular my beloved parents and my multi-talented girlfriend for the love and support they give me on a daily basis, but also everyone else from friends and family to professors and staff who kept believing in me and helped me to get the correct answer by asking the wright questions. And while writing this I cannot help but wonder: "What will be the next question in this series of adventures?"

Table of Contents

List of abbreviations

aTIS:	alternative translation initiation site
CD:	cluster of differentiation
CSB:	cell-surface biotinylation
CSC:	cell-surface capture
Her2:	Human Epidermal growth factor Receptor 2 (ERBB2)
IAA:	Iodoacetamide
LFQ :	label-free quantification
mAbs:	monoclonal antibodies
MS:	mass spectrometry
MS/MS:	tandem mass spectrometry
NHS:	*N*-hydroxysuccinimide
ER:	oestrogen receptor
PgR:	progesterone receptor
PNGaseF:	peptide N-glycosidase F
TFA:	trifluoroacetic acid
TNBC:	triple negative breast cancer
uORFs:	upstream open reading frames

FACULTY OF MEDICINE AND HEALTH SCIENCES

Summary

Despite medical advances, breast cancer remains the world's leading cause of cancer-related deaths in women. Recent breakthroughs in immune therapies have fuelled a shift in the cancer treatment paradigm towards the use of targeted approaches. An important obstacle in both classical antitumoural approaches as well as in the application of immunotherapy is the ability to distinguish cancer cells from healthy cells.

All cells interact with each other and their surrounding environment mainly via their cell surface membrane proteins. Since cancer cells are genetically unstable and have different environmental needs compared to normal cells, the hypothesis of this master's thesis was that there should be detectable differences in cell membrane proteins between cancer cells and normal cells. These differences can ultimately lead to new targets for immunotherapy.

Mass spectrometry (MS) remains the best approach to study protein composition, but despite technological improvements, MS keeps struggling with the high dynamic range and challenges posed by the chemical properties of membrane proteins.

To cover these difficulties in the detection of membrane proteins multiple enrichment strategies have been developed. The aim of this master thesis was to compare the efficiency of two enrichment strategies, namely biotinylation of primary amines and the hydrazide capturing of oxidized glycoproteins.

The hypothesis was that the use of both orthogonal strategies could work complementary in unravelling the protein composition of the cell membrane. However, the results suggest that the hydrazide capturing of glycoproteins was much more efficient in comparison to the biotinylation approach, thereby rendering this approach practically obsolete.

1. Introduction

Currently, breast cancer is the most common cancer among females. With an estimated worldwide incidence of 1.7 million and ~500,000 annual deaths it is the leading cause of cancer-related death in women worldwide [1]. Next to surgery, the classic cytotoxic treatments, like chemo- and radiotherapy, remain the most common treatment options. These classical antitumoural therapies target rapidly dividing cells by interfering with cell proliferation. However, this strategy has the important drawback that therapy can fail to eradicate dormant cancer cells while interfering with rapidly dividing healthy cells [2, 3].

1.1 Targeted therapies

Unsatisfactory results using classical antitumoural techniques such as chemo- and radiotherapy combined with recent breakthroughs in immune therapies have fuelled a shift in the treatment paradigm for cancer towards the use of personalized or targeted approaches such as monoclonal antibodies (mAbs) [4], the use of Activity-on-Target cytokines (Actakines[1]) [5, 6] and others. Targeted approaches against cancer have the advantage that they spare healthy cells because they specifically focus on malignant cells. A second advantage is that they are also better suited for the treatment of distinct cancer subtypes as it was noticed that a lot of the variability in therapy response and prognosis can be explained by distinct expression profiles of certain subpopulations of malignant cells. Exemplary to this are metastases that developed resistance against the therapy that was used to treat the primary tumour [7, 8]. This recent shift towards these targeted approaches has already led to the approval, by the FDA, of more than sixty antitumoural mAbs and 17 targeted therapies specific for breast cancer [9, 10].

Genomic instability is known to be one of the major hallmarks of cancer cells [11]. The accumulation of DNA aberrations such as mutations, translocations, amplifications and deletions can change the composition of their proteins [12], but can also cause expression of different variations of proteins [13, 14]. Increasing evidence, provided by the Proteomics lab of professor Gevaert and others, showed that alternative translation initiation sites (aTIS), in addition to alternative splicing and post translational modifications, contribute to the complexity of proteomes and play an important role in cancer by influencing the translation efficiency of certain cancer-related mRNAs [15, 16]. Recently it was also noticed that during tumour initiation, the translation of proteins was redirected towards upstream open reading frames (uORFs) [17].

[1] *These are mutant cytokines that have a reduced receptor-binding affinity in order to avoid systemic toxicity. They are coupled to a Nanobody directed against a surface antigen of a specific cell population. This does not only restore the activity of the cytokine, but also makes it possible to direct the cytokine to specific cells that express a biomarker. This specificity greatly reduces systemic toxicity and side effects [3].*

1.2 Tumour heterogeneity

Targeted therapies have the potential to revolutionise cancer treatment with the promise of substantially increased survival and less side effects. However, despite often strong initial responses, almost all tumours develop resistance to targeted therapies [18]. This loss of treatment response can be attributed to tumour heterogeneity [19]. Historically cancers were considered to be clonal expansions of a single malignant cell, but there is now evidence of important spatial and temporal heterogeneity in most tumours [20]. Spatial heterogeneity describes the uneven distribution of genetically distinct tumour subpopulations across a single tumour. Temporal heterogeneity is used to denote the changes in composition of these subpopulations over time. The origin of this heterogeneity is currently explained by a Darwinian view on tumorigenesis that is called clonal evolution [20]. This is a reiterative process of clonal expansion and natural selection. Genetic instability of malignant cells results in diversification during the clonal expansion and the emerging subpopulations are evolutionary tested by their surroundings [20] (Figure 1).

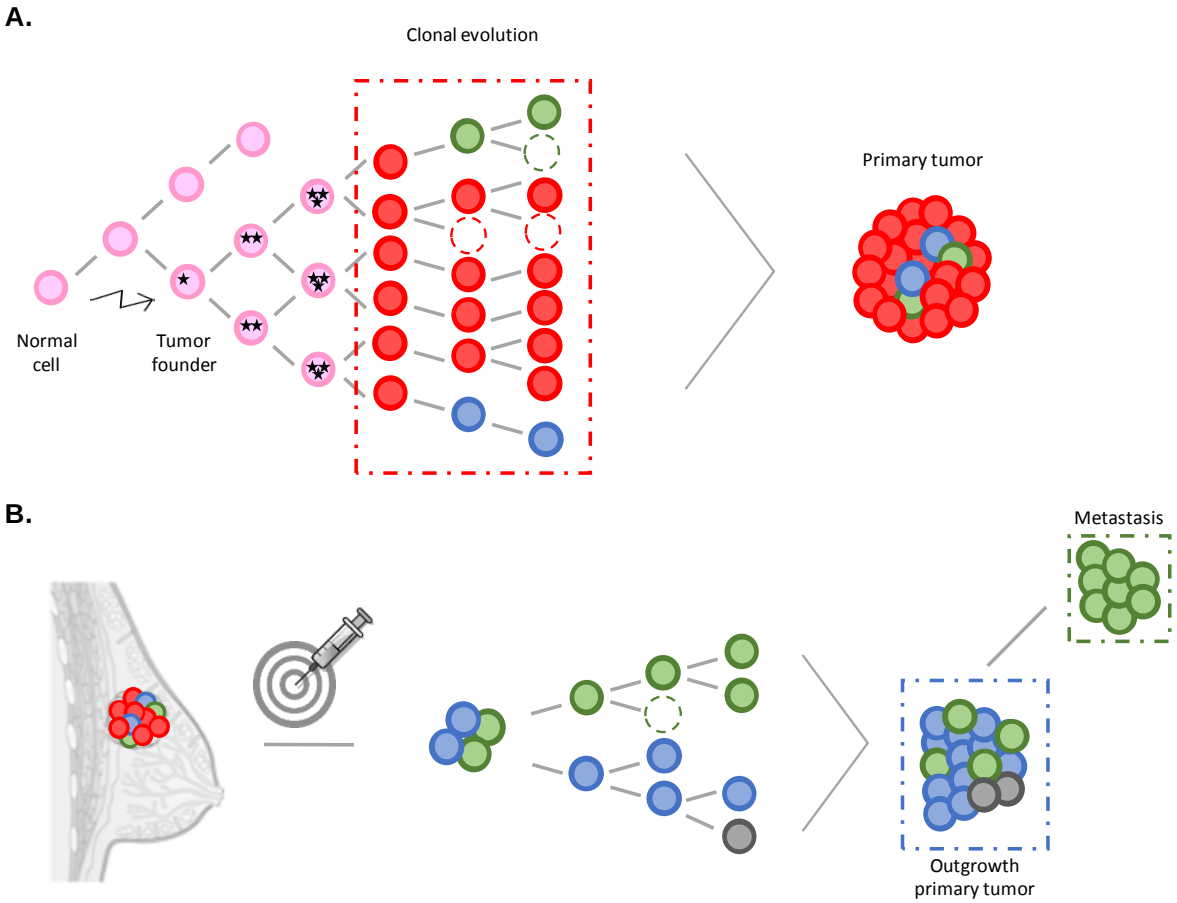

*Figure 1: **A.** Tumour heterogeneity caused by clonal evolution. Tumorigenesis and the accumulation of DNA-damage (*) results in genetically unstable tumour cells. During clonal evolution, selective pressure allows expansion of some subpopulations while others remain dormant or die out. Because of the genetic instability a majority of primary tumours have a very heterogeneous composition. **B.** Therapeutic failure of targeted therapy. Changes in selective pressure can result in an altered composition of the tumour. Targeted therapy can therefore result in selection of therapy resistant cancer subtypes.*

Therapeutic intervention, which results in massive cell death, can provide strong selective pressure for therapy resistant subpopulations. This increase in selective pressure can result in the expansion of resistant variants and the more aggressive re-outgrowth of the tumour due to a lack of competition like is seen in virulent metastasis. Tumour heterogeneity is the main reason for therapeutic failure and this is especially the case for targeted therapies as several studies demonstrate that higher tumour heterogeneity predisposes patients to inferior outcome [21, 22]. This also means that the use of a single targeted therapy can entail major risks in the case of very heterogenic tumours [18]. Therefore, an accurate assessment of the heterogeneity and composition of the tumour could assist in achieving better results with targeted therapies.

Similar to antibiotics, a combination of different targeted therapies is expected to be the best suited strategy to overcome resistance. Mathematical models and preclinical studies in mice predict that combination therapy of two or three targeted therapies should result more often in long-term remission compared to single or sequential therapies [23, 24]. By targeting multiple distinct pathways, this should limit the outgrowth of resistant subpopulations resulting in a sustained therapy response. This research also suggests that dual therapy should be sufficient to result in long-term disease control for most patients while for patients with advanced disease stages and very heterogeneous tumours, triple therapy is more suited.

Targeted treatments have been developed for breast cancer types that are hormone dependent because they overexpress either oestrogen receptors (ER), progesterone receptors (PgR) or the human epidermal growth factor receptor 2 (Her2). However, these therapies often result in therapy resistant tumours [25-27]. Furthermore, a significant proportion of primary breast cancers do not express any of these receptor proteins and are called triple negative breast cancer (TNBC) category. The current lack of suitable targeted therapies and the aggressive nature of these tumours result in a poor prognosis for women diagnosed with TNBC [28]. Key in effectively using targeted therapies or immunotherapy, is the very precise recognition of cancer cells by the therapeutic. Knowing the exact composition of healthy and cancerous cells is essential for the design of such selective therapeutics. The higher cost of targeted therapies in combination with the need to select the correct treatment for a given tumour increases the need for 'companion diagnostics'. These are biomarker assays that stratify patients into responders and non-responders and assist in choosing an appropriate therapy based on the molecular background of the patient [29]. An interesting type of biomolecules in this regard are the membrane proteins that are present on the cell-surface.

1.3 Importance of cell membrane proteins

Cell-surface proteins constitute the main interface through which cells interact with their environment. They are involved in cell-cell interaction, signal transduction and a multitude of other essential cellular functions that also play an important role in tumour development [30-32]. The transmembrane or plasma membrane-anchored proteins that are responsible for these interactions are mainly called cluster of differentiation (CD) proteins and these can be used to identify differences in cellular phenotypes. Because the composition of cell-surface proteins has a strong influence on the interactions and cellular function of a cell, it is highly distinctive for a cell type. This cellular fingerprint is called the surfaceome (cell-surface proteome) [33, 34]. Nowadays global proteome diversity is thought to be much larger than originally expected based on the number of genes. This higher complexity can only be explained by the fact that a single gene can give rise to more than one molecular form of a protein, due to genetic variation and multilevel gene regulation. To denote these different variants of a protein, Smith and Kelleher suggested the use of the term “proteoforms” [14]. This term is now used to group all of the different molecular protein forms coming from a single gene that are caused by biological sources of variation such as allelic variation, post- and co-translational modifications, alternatively splicing and other forms of mRNA editing. Cell-surface proteins are, in addition to being cell type specific, also easily accessible. Making the surfaceome an extremely attractive subpopulation of proteins for drug development [31, 32].

1.4 Mass spectrometry

Mass spectrometry has become one of the key techniques to study proteins, but regardless the importance of membrane proteins in various cellular processes and cancer, as well as their potential to serve as drug targets, the surfaceome has historically been understudied using mass spectrometry (MS). This is partly due to the challenges posed by the hydrophobic nature of membrane proteins, which results in low water solubility. Additionally they have a heterogeneous and highly dynamic composition, contain lower numbers of tryptic cleavage sites and they are less abundant than the intracellular proteins [35]. There are two main approaches for MS-based proteomics, namely top-down and bottom up proteomics. Top-down proteomics starts from intact proteins, while bottom-up proteomics injects enzymatic generated peptides into the mass-spectrometer. The advantages of top-down MS is the nearly complete sequence coverage and the ability to detect posttranslational modifications. Both are necessary for the characterisation of unique proteoforms. Despite the progress being made, top-down proteomics remains computational and experimentally challenging and bottom-up MS remains the most used approach [36]. Therefore, LC-MS/MS was used in this Master thesis project which is explained in the following paragraphs and depicted in Figure 2.

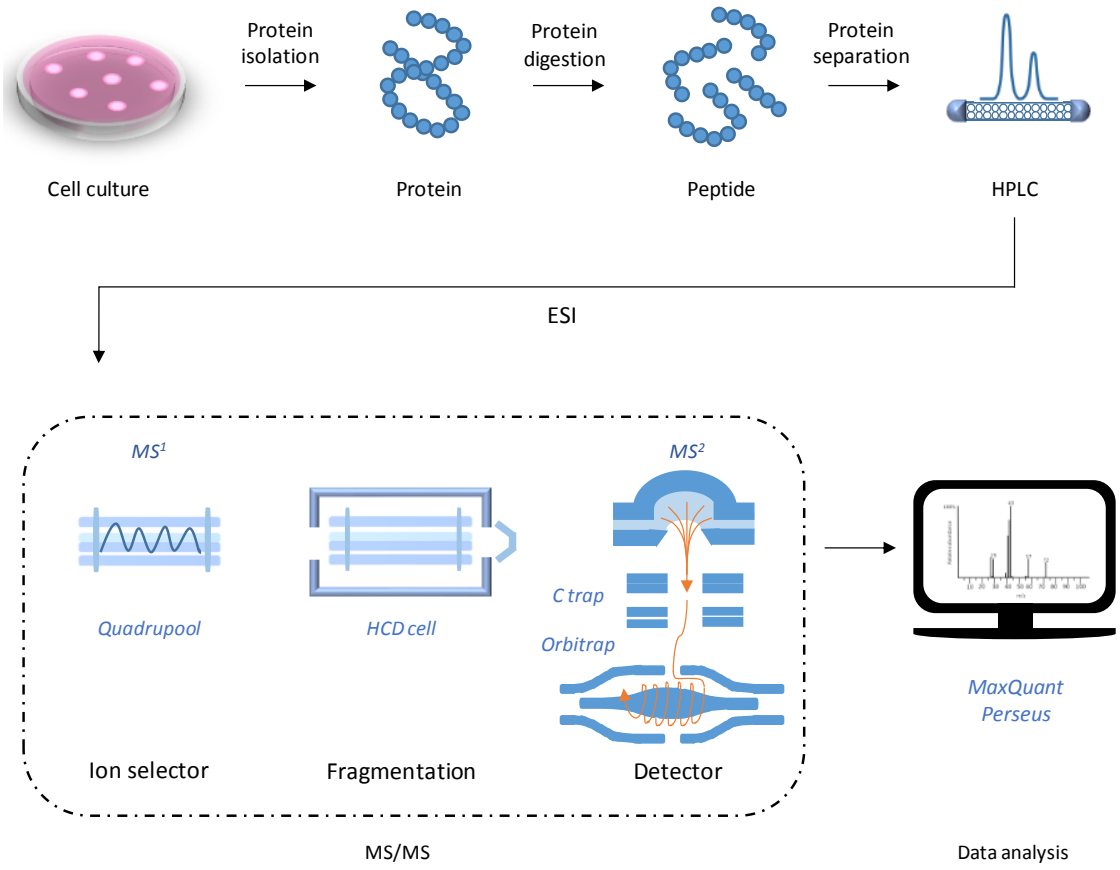

Figure 2: Overview bottom-up proteomics approach in LC-MS/MS. After isolation of the proteins, bottom-up proteomics starts with a proteolytic digestion of proteins into peptides. The resulting peptide mixture is commonly separated by reversed-phase liquid chromatography (RP-LC), followed by ionisation and transfer of the peptides into the gaseous phase. The peptide ions are then separated based on their mass over charge ratio (m/z ratio) and the (n) most abundant (top (n)) peptides are selected and further fragmented into their fragment ions. These are measured and give the MS² spectra that are computationally analysed and processed.

1.4.1 LC-MS/MS

MS-based identification of proteins was originally done by peptide mass fingerprinting (PMF), which measured the mass-over-charge (m/z) values of tryptic peptides and compared this with the m/z values in a database. Tandem mass spectrometry (MS/MS) has greatly improved the accuracy while gel-free approaches resulted in higher throughput. Both techniques revolutionised the proteomics field that is now dominated by liquid chromatography coupled tandem mass spectrometry (LC-MS/MS) [37].

1.4.1.1 Liquid chromatography

The liquid chromatography (LC) step in LC-MS/MS is used to separate the peptides based on their hydrophobicity by forcing the sample mixture (mobile phase) through a LC-column (stationary phase). This fractionation of the sample makes the analysis of the sample less complex, because not all of the peptides enter the mass spectrometer at the same time [38, 39].

A reversed-phase high-performance LC setup (RP-HPLC), as was used for this thesis, uses a hydrophobic (nonpolar) stationary phase in combination with an aqueous (polar) mobile phase. The most frequently used stationary phase in RP-HPLC are silica beads coated with C 18 carbon chains. After loading the sample onto the column, elution of the peptides is done by gradually increasing the percentage of organic solvent in the mobile phase over time. This in known as gradient elution. The higher the affinity of the peptide for the stationary phase the longer it will take to elute from the column and as a result the less hydrophobic peptides elute first [40].

1.4.1.2 Electrospray Ionisation (ESI)

Electrospray ionisation (ESI) (Figure 3) is the most common method to generate charged ions and is considered to be a soft ionisation technique because it generates peptide ions with low internal energies without inducing structural changes in the ionised products. Ionization is achieved by forcing the liquid sample through a cone-shaped capillary (injection needle) at a high electric potential (e.g., 3-5 kV) resulting in a spray of charged droplets [41]. The use of an acidic mobile phase (e.g., aqueous solution containing 2% acetonitrile (ACN) and 0.1% trifluoroacetic acid (TFA)) ensures that positively charged droplets are formed when leaving the injection needle. This is called ESI in positive mode. The positively charged droplets get smaller due to evaporation until they reach the point that the electrostatic repulsion between the ionised particles becomes high enough to overcome the surface tension (Rayleigh limit) and the droplet splits in multiple smaller droplets. This happens several times until the droplets have disappeared and are converted into gas phase ions that can enter into the mass spectrometer [42].

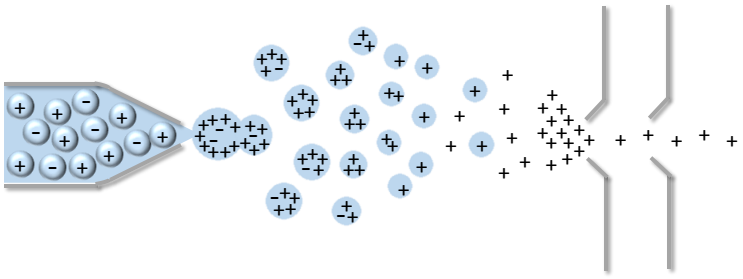

Figure 3: The principle of electrospray ionisation (ESI). The tip of the capillary ejects a spray of charged droplets. Due to evaporation the repulsion between the charges overcomes the surface tension and the droplets repeatedly split into smaller and smaller droplets until they become gaseous peptide ions that can enter the mass spectrometer. In case of positive mode ESI, the use of an acidic mobile phase ensures that positively charged droplets are formed when leaving the injection needle.

1.4.1.3 Collision-induced dissociation (CID)

Collision-induced dissociation (CID) is a fragmentation technique that uses an inert gas, such as N_2 or Ar, to fragment peptide ions into fragment ions. The peptide bonds in the peptide backbone break when the peptide ions collide with the gas particles. Higher-energy collisional dissociation (HCD) is a CID technique specific to the orbitrap technology in which the fragmentation takes place in an external collision cell [43].

1.4.1.4 MS/MS

MS/MS is the sequential use of two mass spectrometers. The reason why there is a second mass spectrometry step is that these m/z values are not always enough to be able to determine the exact composition of the detected peptide. This is especially the case in complex samples. Traditional MS, as explained above, generates MS1 spectra. This is only the first step in MS/MS. After this first step, the (n) most abundant peptide ions are selected and further fragmented into fragment ions that are then measured a second time via MS, generating MS2 spectra. The combination of these two MS spectra with the information of the initial peptide ions (precursor ions) allows to obtain detailed information about the sequence and even posttranslational modifications [37].

1.4.2 Label free quantification

Quantitative proteomics provides information about the abundance of proteins present in a sample instead of providing a list of identified proteins. This quantification provides an additional layer of information because it can be used to look at physiological differences between multiple samples. Quantification can be divided in relative and in absolute quantification. Relative quantification compares the difference in abundance of proteins between different samples and is represented with fold changes. Absolute quantification gives a specific amount for each detected proteins. Relative quantification can be achieved via two distinct approaches: differential labelling and label-free quantification (LFQ). Differential labelling peptides with stable isotopes introduces a difference in mass between the labelled and unlabelled peptides. This difference between samples can be used for relative quantification, but by adding a synthetic peptide at a known concentration this approach can also be used to get an idea about the absolute amount [44].

LFQ is a method to measure and compare the relative abundance of proteins across multiple samples without using isotopic tags to discriminate between samples [45]. A major constraint of the stable isotope-labelling approach is that it requires that all samples are differentially labelled before combining them into a single sample. This is not the case in LFQ because every sample is measured separately. Therefore in a LFQ experiment there are no limitations on the number of samples that can be compared. Another advantage of LFQ is that it can be used for all types of samples. This is very interesting for the clinical samples that otherwise cannot be metabolically labelled. Finally, LFQ is easier and cheaper to preform, because it requires less sample preparation and no expensive stable isotope labels [46].

There are two commonly used methods for label-free quantification: spectral counting and area under the curve (AUC). Spectral counting uses the amount of peptide spectra of a peptide that is specific for one protein and compares these amounts between samples. High numbers of peptide identifications are essential for the accuracy of the extrapolation [44].The AUC is calculated by integrating the peak intensity versus retention time. AUC is more robust compared to the spectral counting approach, but technically more demanding because m/z and retention time need to strictly calibrated between LC-MS/MS runs [47].

Despite technological improvements, MS keeps struggling with the large dynamic range of proteins present in total cell lysates. In an effort to mitigate these difficulties and to allow the identification of some of the more rare proteins various, strategies have been developed to isolate membrane proteins.

1.5 Strategies to enrich for membrane proteins

The extraction and analysis of surface membrane proteins is complicated by the fact that they reside in the insoluble lipid bilayer and are a lot less abundant than the intracellular proteome. Therefore, plasma membrane proteins require an enrichment and solubilisation step prior to mass spectrometric analysis [35]. Over the years, a huge number of strategies have been developed to enrich for membrane proteins. For the purpose of this Master dissertation the focus will be on chemical enrichment strategies, instead of enzymatic methods or isolation based on the physical properties of plasma membranes. This choice was mainly prompted by the fact that chemical enrichment strategies were favoured in recent years. For a comprehensive overview on other approaches I would like to refer to Kuhlmann *et al*. [30].

1.5.1 Affinity purification of biotinylated cell-surface proteins

Cell-surface biotinylation is a chemical labelling method in which the extracellular part of plasma membrane proteins are covalently labelled by a reactive biotin ester. After cell lysis and affinity purification, stringent wash steps ensure the removal of the unspecific bound proteins.

A biotin labelling reagent is typically composed out of three parts (Figure 4), namely a reactive ester group, the biotin label and a linker moiety [48].The reactive ester determines the location of the covalent bond between the label and the target biomolecule. The linker functions as a spacer between the biotin moiety and the reactive ester that limits the steric hindrance. The biotin residue itself is essential for the affinity purification [30].

Figure 4: Three constituents of a biotin labelling reagent for membrane protein enrichment. Reactive ester (orange), linker (green) and biotin moiety (blue).

Reactive ester group

The reactive ester group can be designed to covalently bind to amino groups, thiol- or carboxyl groups [31]. *N*-hydroxysuccinimide (NHS) esters react with primary amino groups. Since primary amino groups are present both at free (non-acetylated) N-termini and in the amino acid lysine, these groups can be found in majority of proteins [49]. NHS esters are very hydrophobic, but by adding a sulfo group the labelling reagent becomes charged and hydrophilic. The combination between the increased polarity and electrostatic repulsion between phospholipids and the sulfonic acid of the label make it less prone to cross the cell membrane and label internal proteins [31, 49]. Thiol groups on cysteine and methionine residues are a lot less prevalent compared to primary amines. Furthermore, because they are exposed in the oxidative environment of the extracellular medium they are likely to form disulphide bridges making them unavailable [50]. Carboxyl groups can also be targeted, but these have the drawback that some of the membrane proteins carry a lipid anchor on their C-terminal amino acid [51].

Biotin

The biotin moiety is the functional group that is essential for the affinity purification. The interaction between biotin and avidin (or the deglycosylated version Neutravidin) has a dissociation constant (K_d) around an order of magnitude of 10^{-15} M, which makes it one of the strongest non-covalent biological interaction. This allows stringent washes to remove unspecific interactors without losing the biotin labelled proteins [31].

Linker

The linker that acts as a spacer between the reactive ester moiety and the biotin group can serve several functions. A first general function of the linker is that it helps to reduce the steric hindrance and allow the membrane protein being captured on the avidin resin. Secondly, by increasing the size of the label it also prevents cellular internalisation via vesicular uptake [49]. Thirdly, because of the strong interaction between biotin and avidin the protein cannot be retrieved from the beads under normal conditions. This can be overcome by using a cleavable linker such as a disulphide bridge which opens up under reducing conditions [52].

1.5.2 Enrichment of glycoproteins

Glycoproteins are proteins that are covalently linked with an oligosaccharide (glycan) chain to the side-chain of an amino acid. With an estimated prevalence of 90% is glycosylation the most common modification of the surfaceome in multicellular organisms. It is essential in important cellular functions such as cell adhesion, but is also necessary for correct protein functioning because of its role in folding, intracellular migration and ligand interaction [53].

There are many types of glycosylation, but the two best known classes of glycoproteins are N-glycoproteins and O-glycoproteins [54]. The distinction is based on the type of link between the sugar molecules and the protein. In the case of N-linked glycosylation the carbohydrates are covalently bound to the amide nitrogen of an asparagine side chain, while O-glycosylation uses the oxygen of the hydroxyl-group of a serine or threonine side chain (Figure 5).

All mammalian glycans are composed out of only ten monosaccharides: N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc), fucose (Fuc), galactose (Gal), glucose (Glc), glucuronic acid (GlcA), iduronic acid (IdoA), mannose (Man), xylose (Xyl) and sialic acid (SA). However, because glycan chains can branch out, they can reach a much higher structural diversity compared to the linear character of amino acids in proteins [55]. Additional complexity comes from the fact that glycosylation is not template-driven but synthesized in intricate biosynthetic pathways that are influenced by variety of external factors such as enzyme accessibility or availability of sugar-nucleotide precursors [56]. While glycosylation is considered to be a posttranslational modification, N-glycosylation is mostly a co-translational modification that takes place during translation in the endoplasmic reticulum. Most plasma membrane and secreted proteins are translated and processed in the ER, therefore it is not surprising that a majority of the glycosylated membrane proteins are N-glycosylated [55].

Owing to the importance of this protein subclass, multiple approaches have been developed to select for glycoproteins. Affinity purification using lectins used to be a popular approach to capture glycoproteins, but because of the variation in affinity for specific carbohydrates chemical enrichment strategies are now being used more often [31]. The main approaches are cell-surface capture and aminooxy biotinylation.

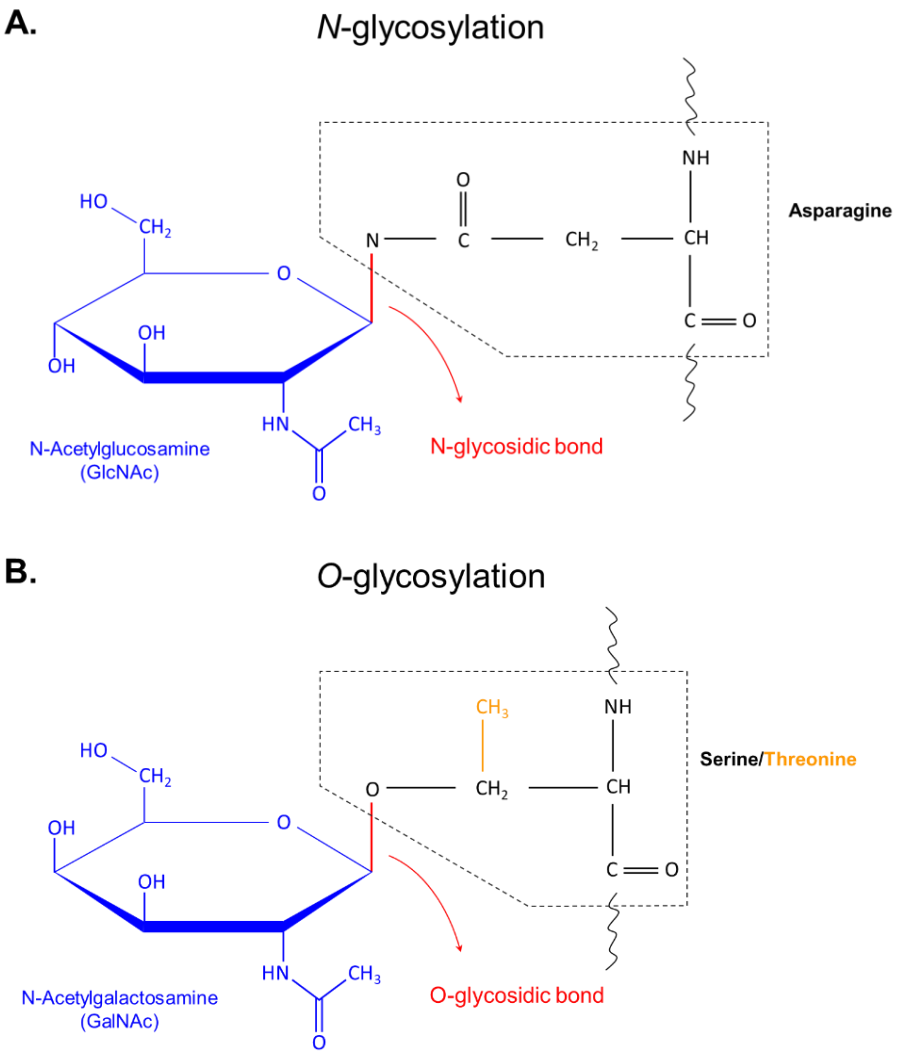

Figure 5: Two most important classes of Glycoproteins. ***A.*** *N-glycosylation covalently links sugar moiety to a protein by forming amide bond (N-glycosidic bond) between asparagine and N Acetylglycosamine (GlcNAc).* ***B.*** *O-glycosylation uses an ether bond (O-glycosidic bond) to couple N-Acetylgalactosamine (GalNAc) onto serine or threonine side chain.*

1.5.2.1 Cell-surface capture (CSC)

Cell-surface capture (CSC) is a chemical capture method that was developed by Wollscheid *et al*. [57] and uses hydrazide chemistry to enrich for N-glycosylated peptides. Sodium metaperiodate ($NaIO_4$) is used to oxidize vicinal diols, in extracellular exposed carbohydrates, to adjacent aldehydes. For example sialic acid is a monosaccharide that is present in both N- and O-linked glycan chains and is already oxidized at sodium metaperiodate concentrations of 1 mM [58]. The formed aldehydes are then covalently labelled with biocytin hydrazide. The hydrazide binds to the aldehydes to form hydrazone. After cell lysis and digestion, affinity chromatography with streptavidin is used to capture the biotin-tagged glycosylated proteins. Washing removes the unspecific binders and subsequently the captured peptides are released with peptide N-glycosidase F (PNGaseF) which is an amidase that specifically cleaves the N-glycosidic bond [59, 60]. CSC is only suited to study cultured cells because of the requirement of membrane integrity and the substantial amount of starting material ($> 10^7$ cells) which makes it difficult to use for tissue biopsies [30].

1.5.2.2 Aminooxy biotinylation

Aminooxy biotinylation uses mild oxidation $NaIO_4$ [1mM] of sialic acid residues followed by oxime ligation with a biotin tag for subsequent affinity purification. The oxime ligation can be accelerated using aniline as nucleophilic catalyst. This allows the use of the labelling reagent at a neutral pH and in a low concentration thus maintaining cell viability [58].

Hormann *et al*. compared the isolation of cell membrane proteins using sulfo-NHS-SS-biotinylation with aminooxy-biotinylation [52]. Aminooxy-biotinylation resulted in the detection of 468 proteins while using sulfo-NHS-SS-biotinylation was able to detect 1306 proteins. However, 340 (74%) of these proteins were annotated as plasma membrane proteins in the aminooxy-biotinylation protocol compared to 650 (49%) in sulfo-NHS-SS-biotinylation protocol, proving the efficiency of the approach at the cost of a lower yield.

The majority of the other strategies that have been developed to capture glycoproteins suffer from varying selectivity and poor reproducibility. These drawbacks do not apply for the cell-surface capture (CSC) approach [30], but the selectivity for N-glycoproteins is a disadvantage in this thesis, because our interest is on the protein level rather than the glycan level. McDonald *et al*. proposed an alternative, but very similar approach that allows to detect both N- and O-glycosylated proteins. Following the oxidation with sodium periodate, cell lysates are mixed with hydrazide beads to capture glycosylated proteins instead of peptides (Figure 6).

After washing, the captured proteins are digested with trypsin while they are still attached to the beads (on-bead digest). Treatment with PNGaseF is an optional step that allows the retrieval of the N-glycosylated peptide that was bound to the hydrazide. An amidase that cleaves the O-glycosidic bond is currently not available [61].

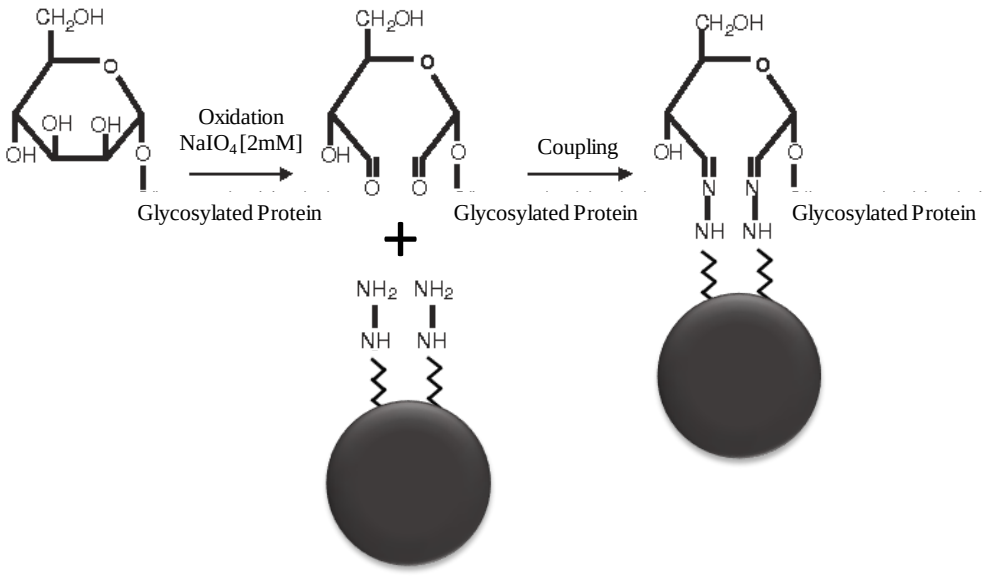

Figure 6: Oxidation of a cis-diol in a carbohydrate to an aldehyde followed by covalent coupling to hydrazide beads. Adapted from: Zhang et al. (Nat Biotechnol., 2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labelling and mass spectrometry.

1.6 Aim and objectives

The aim of this Master thesis was to study the composition of the surfaceome of breast cancer cells. To achieve this aim, the objective was to optimise the enrichment of membrane proteins. Since all of the above discussed approaches have theoretical and practical limitations that only give a partial image of the surfaceome and it was seen that the use of different approaches targeting glycoproteins greatly improved the coverage of the glyco-proteome [61]. The hypothesis is that the parallel use of several of these approaches could mitigate the drawbacks of an individual approach and result in a more complete overview of the surfaceome.

Herefore, we selected two methods: biotinylation using a sulfo-NHS-SS-biotin labelling reagent and affinity purification of oxidized glycoproteins using hydrazide beads (for an overview see Figure 8).The sulfo-NHS-SS-biotin labelling reagent (Figure 7) has the advantage that it selectively targets primary amines, without passing through the cell membrane into the cytosol. Furthermore, it has the possibility of retrieving the proteins after reduction of the disulphide bridge in the linker.

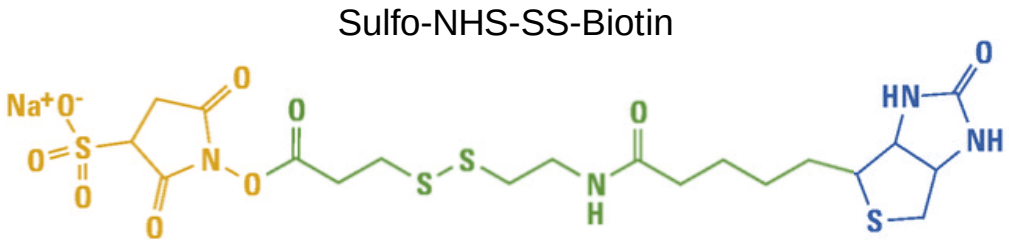

Figure 7: Selected labelling reagent: Sulfo-NHS reactive ester (orange), cleavable linker with a disulphide bridge (green) and biotin moiety (blue). Adapted from: ThermoFisher product website

The approach of McDonald *et al.*, that allows to detect both N- and O-glycosylated proteins, was selected because it offers a more complete picture of the (glycosylated) surfaceome.

An additional motivation to select these methods is their complementary strategy. The hydrazide approach only targets glycosylated proteins and would therefore miss the non-glycosylated proteins. This would be unfortunate as non-glycosylated membrane proteins, despite their minority, could still contain interesting information. Because the biotinylation approach targets primary amines, the assumption is that especially those membrane proteins that are not heavily glycosylated are more accessible. The drawback of this approach however is the possibility that a relevant membrane protein does not have an accessible primary amine. This can be the case if there is an N-terminal modification or if there are no lysines in the extracellular part of the protein. This is mostly mitigated with the enrichment strategy of the glycoproteins.

The fulfilment of this objective will lead to an optimised enrichment strategy that will enable us to capture as much membrane proteins as possible. This can then be used to compare different cell lines with each other and ultimately to determine the surfaceome of breast tumours which eventually can aid to select the best suited targeted therapy.

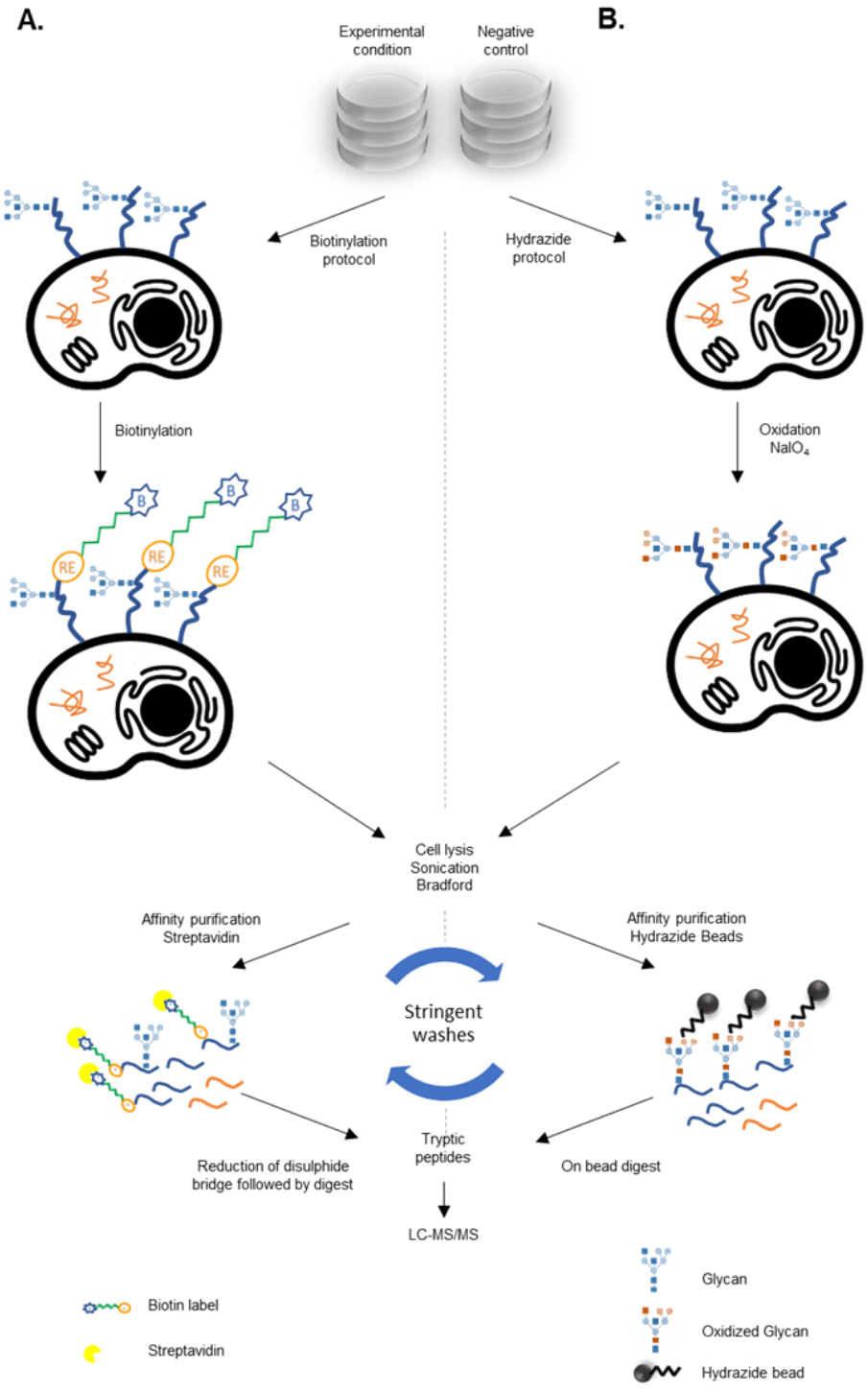

Figure 8: Comparison of the two used chemical enrichment strategies. ***A.*** *Cell surface biotinylation. Intact cells are incubated with reactive biotin esters that covalently bind to primary amines. After cell lysis and sonication, protein concentration is determined and equal amount of proteins are added to streptavidin beads to enrich for biotinylated proteins. Following affinity purification stringent washes are used to remove unspecific binders. Captured biotinylated proteins are released from the beads using 15 mM DTT to reduce the disulphide bridge and digested with endoproteinase-LysC and trypsin overnight.* ***B.*** *Hydrazide capture of oxidized membrane proteins. Intact cells are treated with sodium metaperiodate (NaSO4), which leads to formation of di-aldehydes on carbohydrates. Following cell lysis, sonication and determination of protein concentration, the glycoproteins that have a di-aldehyde are enriched from the complex protein mixture by hydrazide beads. Non-glycosylated proteins and intracellular (non-oxidized) glycoproteins are washed away before on bead digest. The tryptic peptides generated by both protocols are then analysed via LC-MS/MS.*

2. Materials and methods

2.1 Cell culture

The setup and the optimisation of both protocols was done in MCF-7 cells (an epithelial breast cancer cell line). Cell lines were preferred over native/patient derived material for convenience (i), in order to have sufficient and stable material for protocol optimization (ii), cultured cells form a homogeneous population and this ensured that the identified proteins were from the actual tumour cells and not from the tumour microenvironment (TME) (iii). MCF-7 is an epithelial breast cancer cell line from a 69-year old Caucasian female derived from the metastatic site [62]. The cells express the WNT7B oncogene and oestrogen receptor (ER). MCF-7 cells were obtained from the American Type Culture Collection (ATCC, Manassas, Virginia, US) and cultured in Dulbecco's Minimum Essential Medium (DMEM) with 10 % foetal bovine serum (FBS) and 2% penicillin/streptomycin at 37° in a 5% CO_2 incubator.

Per experimental condition, of which three replicates were performed, 1-4 x10^7 cells were used, per replicate. Cells were plated on 150x20 mm cell culture dish (Nunclon Delta surface/145 cm² from Thermo Fisher Scientific, Waltham, Massachusetts, US) ~5 days prior to the cell surface protein enrichment experiment and grown till ~95% confluence.

2.2 Isolation of cell-surface membrane proteins

For the isolation of membrane proteins, two different approaches were investigated in parallel (Figure 8). All steps in both protocols were performed on ice (4° C) to prevent protein processing. Milli-Q water (18.2 MΩ·cm at 25 °C) was used for all preparations.

2.2.1 Surface biotinylation protocol

This approach uses the conjugation of a biotin label to the primary amine groups of proteins. This protocol is adapted from Scheurer *et al.* [63], Roesli *et al.* [8], Karhemo *et al.* [64] and the entire protocol can be found in the supplementary section.

Six dishes (3x biotin labelling condition (B) and 3x negative control (C^-)) with ~95% confluent MCF-7 cells were washed three times with ice-cold, cation free Dulbecco phosphate buffered saline (DPBS) (Thermo Fisher Scientific). This extensive washing of the cells before the labelling reaction ensures the removal of excreted proteins and already proteins present in the medium. To prevent biotin reduction during cell lysis, cultured plates were incubated for 5 minutes with 30 mM iodoacetamide (IAA) in 10 mL DPBS at 4 °C covered by aluminium foil. After a washing step to remove the IAA, the labelling of half the dishes was done by a 30-minute incubation with 500 µM Pierce™ Premium Grade Sulfo-NHS-SS-Biotin (Thermo Fisher Scientific) on ice. The other half was incubated with DPBS as a control condition. Since sulfo-NHS-SS-biotin is an impermeable membrane reagent, it should only covalently modify the extracellular parts of the membrane proteins. The cold temperatures minimise the vesicular uptake of the biotin label. Cells were washed twice with DPBS and the non-reacted biotin was blocked with 20 mM glycine for 15 minutes. Cell lysis was done with 1 mL of lysis buffer (pH 7.4) (1% IGEPAL, 0.5% DOC, 0.1% SDS, 1 mM EDTA, 50 mM phosphate buffer, 150 mM NaCl, 30 mM IAA, 1X protease inhibitor (Roche, Basel, Switzerland) in Milli-Q) was added to the cells. Lysed cell extracts were scraped off the plates and transferred to a 2 mL Eppendorf tube. These lysates were sonicated followed by 30 min incubation on a rotor at 4 °C. Next, the lysates were centrifuged for 10 min (16000×g, @ 4 °C) to pellet the insoluble material. Bradford protein assay was used to determine protein concentrations. Equal amounts of protein (~4.5 mg) were added to 40 µL Neutravidin beads (Thermo Fischer Scientific), and incubated on rotor for 3 hours at 4 °C. Next, beads were washed four times with the lysis buffer

containing 300 mM NaCl and twice with 100 mM TEAB (pH = 8.0) to achieve detergent-free buffer. Proteins were eluted twice with 50 µL of 15 mM DTT dissolved in 100 mM TEAB (pH 8.0) for 15 min at 55 °C to reduce the disulphide bridge in the biotin label. Both elutions were pooled and digested overnight at 37 °C with 10 µL of sequence grade trypsin [0.2 µg/µL] (Promega, Fitchburg, Wisconsin, US). The digested peptide samples were acidified to 1% trifluoroacetic acid (TFA) and centrifuged for 15 min at 1780 x g to precipitate insoluble material. To remove the buffer and salts from the peptide mixture, the cleared samples underwent C18 pipette-based solid phase extraction (SPE) (Ziptip Bond Elut OMIX from Agilent, Santa Clara, California, US). The tips were activated by adding five times 150 µL of pre-wash buffer (80% ACN / 20% Milli-Q / 0.1% TFA) that was each time discarded. Next, the tips were washed five times with 150 µL Solvent A (= wash buffer) (100% Milli-Q / 0.1% TFA). The acidified and cleared samples were loaded on the resin by pipetting the samples 20-30 times up and down. Hereafter they were washed 3 times with solvent A and eluted by sequential use of two times 75 µL solvent B (= elution buffer) (60% ACN / 40% Milli-Q / 0.1% TFA) that passed ten times through the resin. The resulting peptide samples were dried and re-suspended in 20 µL loading buffer (0.1% TFA, 2% ACN and 98% Milli-Q (v/v)) and subsequently analysed by mass spectrometry. The difference in amount of membrane proteins between both conditions allowed to demonstrate the enrichment.

2.2.2 Hydrazide protocol

This protocol is adapted and optimised from Zhang *et al*. [65] and McDonald *et al*. [61] and can be found in full in the supplementary section. This approach is based on the conjugation of oxidised glycans to hydrazide beads that allows to capture and selectively enrich for glycosylated membrane proteins. There are two conditions in which protein samples will be collected: the oxidised condition (Ox) and the negative control (C^-) each performed in three replicates.

The six dishes with ~95% confluent MCF-7 cells were washed three times with ice-cold cation free DPBS (pH = 6-6.5). Three dishes were treated for 10 min with 2 mM sodium metaperiodate ($NaIO_4$) in 10 mL DPBS (pH = 6-6.5) in the dark on ice. $NaIO_4$ oxidises the carbohydrates present on the extracellular membrane proteins and the slightly acidic conditions expedite the oxidation. After the oxidation step, the $NaIO_4$ reagent was aspirated and the cells were washed three times with ice-cold DPBS (pH = 7). Cell were lysed with 1 mL of lysis buffer (pH = 7) (1% IGEPAL, 0.1% SDS, 50 mM Sodium Acetate, 300 mM NaCl, 1X protease inhibitor (Roche, Basel, Switzerland) in Milli-Q) and scraped into an Eppendorf tube using a cell scraper. After sonication, cell lysates were rotated for 30 minutes at 4° C. The lysates were cleared by centrifugation at 16000 x g for 10 min. at 4 °C. The salt concentration in the cleared samples was increased from 0.3 M to 1.5 M NaCl to minimise non-specific interaction with the hydrazide beads. After determining the protein concentration with the Bradford assay, the samples were acidified to pH 5.5. The hydrazide beads that were used (Affi-Gel Hz # 153-6047) have a capacity between 1-5 mg/mL for oxidized carbohydrates. The beads were washed three times with Milli-Q and re-suspended in 50 mM sodium acetate buffer (pH = 5.5). Equal amounts of proteins (~4.5 mg) were added to 100 µL of hydrazide beads slurry. This was followed by overnight affinity purification @ 4 °C on a rotor. Unbound proteins were removed by centrifugation for 2 min at 2000 x g at 4° C. Beads were re-suspended in 1 mL of Urea buffer (pH = 8.0) (8 M Urea / 0.1 M TEAB / 0.1% SDS) to denature the captured proteins. Disulphide bridges were reduced by incubation with 15 mM DTT for 1 h at room temperature and alkylation of free thiols was performed using 30 mM IAA for 30 min in the dark. Released interactors were removed by washing three times with urea buffer and the urea buffer was subsequently replaced with 100 mM TEAB (pH = 8.0). Proteins were digested on-beads with 3 µL endoproteinase Lys-C [1µg/µL] in 200 µL 100mM TEAB (3 h) followed by digestion with 10 µL of sequence grade trypsin [0.2 µg/µL] (Promega) overnight in a heat block at 37 °C. Samples were then acidified to 1% TFA and cleared (15 min at 1780 x g) prior to

SPE. To remove the buffer and salts from the peptide mixture, the cleared samples underwent C18 Pipette-based solid phase extraction (SPE) (Agilent) The resulting peptide samples were dried and re-suspended in 20 µL loading buffer (0.1% TFA, 2% ACN and 98% Milli-Q (v/v)) and subsequently analysed by mass spectrometry. The difference in amount of membrane proteins between both conditions allowed to demonstrate the enrichment efficiency.

2.3 LC-MS/MS analyses

In the first two experiments (CMB 467 / 471), the eluted peptides were analysed with a Q Exactive HF Biopharma mass spectrometer (Thermo Fisher Scientific), but because the reverse-phase column was changed for other experiments on this mass spectrometer, the third and fourth experiments (CMB 479 / 488) were analysed using a Q Exactive mass spectrometer (Thermo Fischer Scientific).

Purified peptides were re-dissolved in 20 µL of loading buffer (0.1% TFA in water/acetonitrile (98:2, v/v)) and 15 µL (max volume that can be injected which approximates 3 µg of peptides) of each sample was injected in an Ultimate 3000 RSLC nano-system, equipped with a 20 µL loop, in-line connected to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific) or a Q Exactive mass spectrometer (Thermo Fisher Scientific). Trapping was performed at 10 µl/min for 4 min in loading solvent A (0.1% TFA in water/acetonitrile (98:2, v/v)) on a 20 mm trapping column (made in-house, 100 µm internal diameter (I.D.), 5 µm beads, C18 Reprosil-HD, Dr. Maisch, Germany) and the sample was loaded on a 200 mm analytical column (made in-house, 75 µm I.D., 1.9 µm beads C18 Reprosil-HD, Dr. Maisch). Prior to packing the column, the fused silica capillary had been equipped with a laser pulled electrospray tip using a P-2000 Laser Based Micropipette Peller (Sutter Instruments, Navato, CA, US). Peptides were eluted by a non-linear gradient from 2 to 56% solvent B (0.1% formic acid (FA) in water/acetonitrile (2:8, v/v)) over 145 min at a constant flow rate of 250 nl/min, followed by a 5 min ramp to 97% solvent B with an additional 10 min wash. The column was then re-equilibrated with 98% solvent A (0.1% FA in water) for 20 min. The column temperature was kept constant at 50 °C in a Butterfly column heater (Phoenix S&T, Chadds Ford, PA, US). The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS acquisition for the (n) most abundant ion peaks per MS spectrum (n = 12 for the analyses ran on the HF and 5 for the analysis on the QE). Full-scan MS spectra (375-1500 m/z) were acquired at a resolution of 60,000 in the Orbitrap analyzer after accumulation to a target value of 3E6. The (n) most intense ions above a threshold value of 1.3E4 were isolated for fragmentation at a normalized collision energy of 28% after filling the trap at a target value of 1E5 for maximum 80 ms. MS/MS spectra (200-2000 m/z) were acquired at a resolution of 15,000 in the Orbitrap analyser.

2.4 Data analysis

The protein identification and quantification was done in an automated manner by using the so called protein database approach. This approach uses a search algorithm that tries to match the generated peptide MS/MS spectra with in *silico* calculated MS/MS spectra starting from a protein sequence database. The search parameters, based on the experimental conditions, are essential to improve the accuracy of the matches and subsequently the protein identification [66].

2.4.1 MaxQuant

Database search was done using the open source MaxQuant software package [67] that contains the integrated peptide search engine Andromeda [68] and the MaxLFQ algorithm that uses delayed normalization to make LFQ fully compatible with up-front separation [46]. The peptide identifications of all experiments in this thesis were done using the latest MaxQuant (version 1.6.5.0). with the default search settings including a FDR set at 1% on both the peptide and protein level. Spectra were searched against the human SwissProt database (January 2019, www.uniprot.org).

2.4.1.1 Search parameters biotinylation protocol (CMB 471)

- Group specific parameters (**group 0 = biotin condition**)
  - Variable modifications:
    - Oxidation (M)
    - Acetyl (Protein N-term)
    - Thioacyl (L)
    - Carbamidomethyl (C)
  - Digestion
    - Trypsin/P with a maximum of 4 missed cleavages

- Group specific parameters (**group 1 = control**)
  - Variable modifications:
    - Oxidation (M)
    - Acetyl (Protein N-term)
    - Carbamidomethyl (C)
  - Digestion
    - Trypsin/P with a maximum of 2 missed cleavages

- Global parameters
  - Protein quantification
    - Unique + razor peptides
    - Modifications used in protein quantification:
      Oxidation (M), Acetyl (Protein N-term), Carbamidomethyl (C), Thioacyl (DSP)

The rationale behind these search parameters was the following. N-terminal acetylation is a prevalent protein modification which affects around 80% of all human proteins [69]. Protein Nt-acetylation refers to the covalent attachment of an acetyl group (CH_3CO) to the free α-amino group (NH_3^+) at the N-terminal end of a polypeptide [70]. Despite the fact that many membrane proteins have lost this N-terminal acetyl group, Nt-acetylation was used as a variable modification to help identify some of the background peptides.

The thioacyl group is a remnant of the labelling reagent and was therefore only used in the biotin condition. Since this remnant on lysines could interfere with the tryptic cleavage a higher number of missed cleavages was allowed. Carbamidomethylation is an expected modification of cysteines caused by the incubation with IAA prior to labelling.

2.4.1.2 Search parameters hydrazide protocol (CMB 467, 479, 488)

- Group specific parameters (No grouping)
  - Variable modifications:
    - Oxidation (M)
    - Acetyl (Protein N-term)
  - Fixed modifications
    - Carbamidomethyl (C)
  - Digestion
    - Trypsin/P with a maximum of 2 missed cleavages

- Global parameters
  - Protein quantification
    - Unique + razor peptides
    - Modifications used in protein quantification:
      Oxidation (M), Acetyl (Protein N-term), Carbamidomethyl (C)

The search parameters for the hydrazide protocols were similar as those of the biotin protocol as mentioned above (section 2.4.1.1). However, carbamidomethylation is here also caused by alkylation with IAA, but this time the step was performed later in the protocol (just before trypsinisation).

2.4.2 Perseus

Data analysis was performed with the Perseus software [67, 71] (version.1.6.5.0) using the ProteinGroup.txt generated by MaxQuant. Protein groups only identified by site, reverse hits and potential contaminants were filtered out. The biological replicates were grouped and LFQ intensity values were log2 transformed. Next, only the protein groups that showed 3 LFQ valid values in at least one group were retained. These remaining and normalised protein groups were checked whether they followed a normal distribution prior to imputation of the missing values. This imputation was done with values from the lower part of the normal distribution, close to the detection limit. Following this, a two-sided t-test was performed with a permutation-based FDR (FDR = 0.01, s = 0.1 and 1000 permutations) that was visualised in a Volcano plot (Figure 9, Figure 11, Figure 13 and Figure 15) The significant protein groups were then isolated, z-scored and plotted into a heat map with hierarchical clustering (Figure 10, Figure 12, Figure 14 and Figure 16).

2.4.3 Functional enrichment analysis

Functional enrichment analysis was performed to have an idea about the origin and properties of the identified proteins that were found to significantly differ between the enrichment and negative control group. Enrichment scores were obtained by using the Database for Annotation, Visualization and Integrated Discovery (DAVID) 6.8 (https://david.ncifcrf.gov/tools.jsp). This is a bioinformatics tool that pools information from major protein and gene databases, with the aim of analysing large datasets in a high-throughput manner [72]. The analysis was performed by searching the majority proteins, of the protein groups that were significantly more present in the enrichment condition, against both the background of the experiment (all retained protein groups) as well as the *Homo sapiens* proteome as background.

3. Results

This Master thesis project focussed on the detection of plasma membrane proteins by two different approaches: affinity purification of biotin labelled surface proteins and affinity purification of oxidised surface proteins with hydrazide beads. The results of both approaches were then compared to check whether these merely confirmed each other or worked complementary in order to provide a completer picture of the composition of the surfaceome.

3.1 Results of the biotinylation experiment

Three replicates of two conditions, membrane protein enrichment using biotin labelling (B) and the negative control (C), were analysed via LC-MS/MS. An overview of the analysis features are shown in table 1.

Samples	*MS/MS*	*MS/MS Identified*	*Identification rate*	*Protein identifications*
B1	37905	8373	22.09%	618
B2	39942	7224	18.09%	636
B3	41228	8577	20.80%	909
C1	40296	5307	13.17%	517
C2	38349	4736	12.35%	522
C3	39607	6916	17.46%	822

Table 1: Analysis features overview of the MaxQuant search for biotinylation protocol (CMB 471). The number of MS/MS spectra (column 2), the number of these spectra that were identified (column 3), the proportion of MS/MS spectra that were identified was derived by the deviation of column 3 by column 2) (column 4) and the total amount of proteins that were identified in each sample (column 5).

Since no major differences were seen in the number of protein identifications between the biotin and control condition, as can appreciated form column 5 in table 1, the biotin labelling was checked by looking at the number of identified thioacyl modification events since this is specific for the biotin labelling condition. The identification of 91 thioacyl events confirmed that biotinylation took place.

From the total of 1154 proteins groups that were identified in the MaxQuant search, 1027 remained after filtration and 584 protein groups (634 majority protein IDs) had 3 valid LFQ values in at least one group (biotin or control) which are shown in figure 9. Of these 584 retained protein groups, 266 protein groups showed a significant difference between both groups with 260 protein groups that were significant enriched in the biotin condition compared to 6 protein group that were enriched in the control condition. This higher enrichment in the biotin condition was seen as a right shift (shift towards the biotin side). This shift was also observed below the significance threshold.

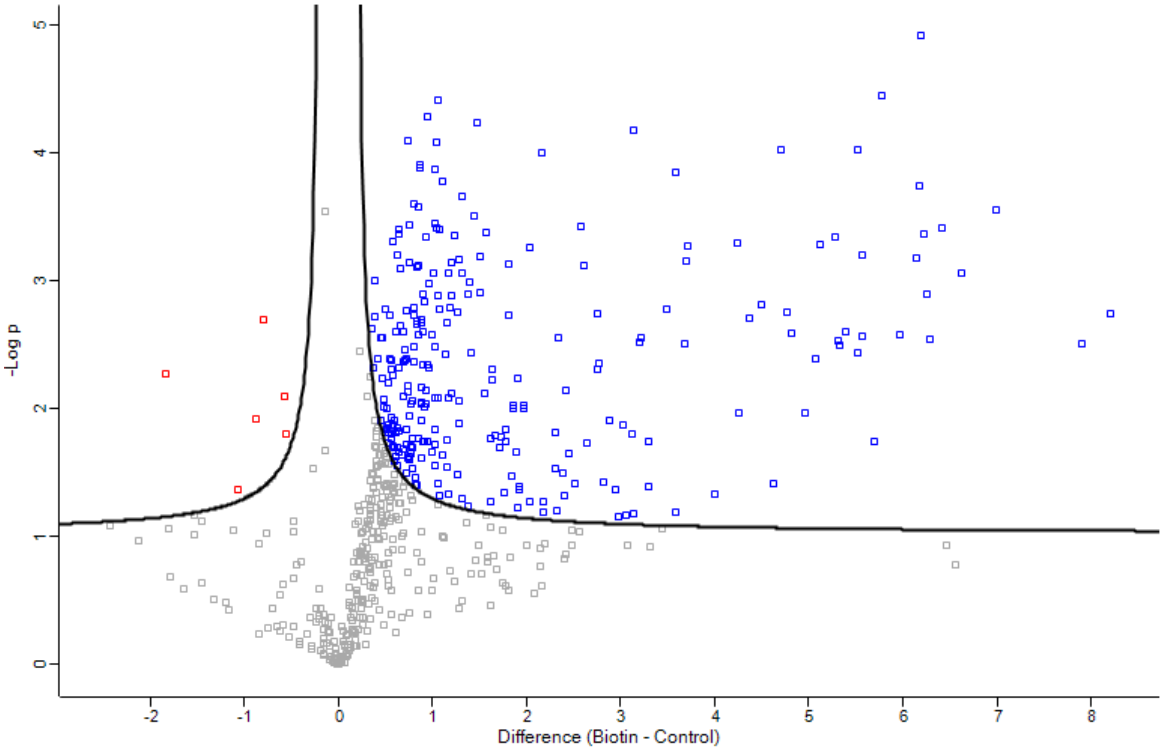

Figure 9: Volcano plot CMB 471 biotinylation protocol. The significance was plotted on the y-axis versus fold-change on the x-axis. The black line indicates threshold of significance (FDR = 0.01). The 260 proteins that were found significantly enriched in the biotin condition compared to the control condition were marked in blue. The 6 proteins that were significantly enriched in the control condition compared to the biotin condition were marked in red.

To check the degree of inter-sample variability heat maps were generated (Figure 10). Protein abundance, using all retained proteins, clustered together into the two conditions (biotin and control condition) (panel A). To remove the background, only the 266 proteins that were significantly increased were clustered according to their abundance (panel B). Although some variability remained, clusters were more pronounced. Two main clusters were formed, with 6 protein groups that clustered together in the control condition and 260 protein groups that clustered together in the oxidation condition.

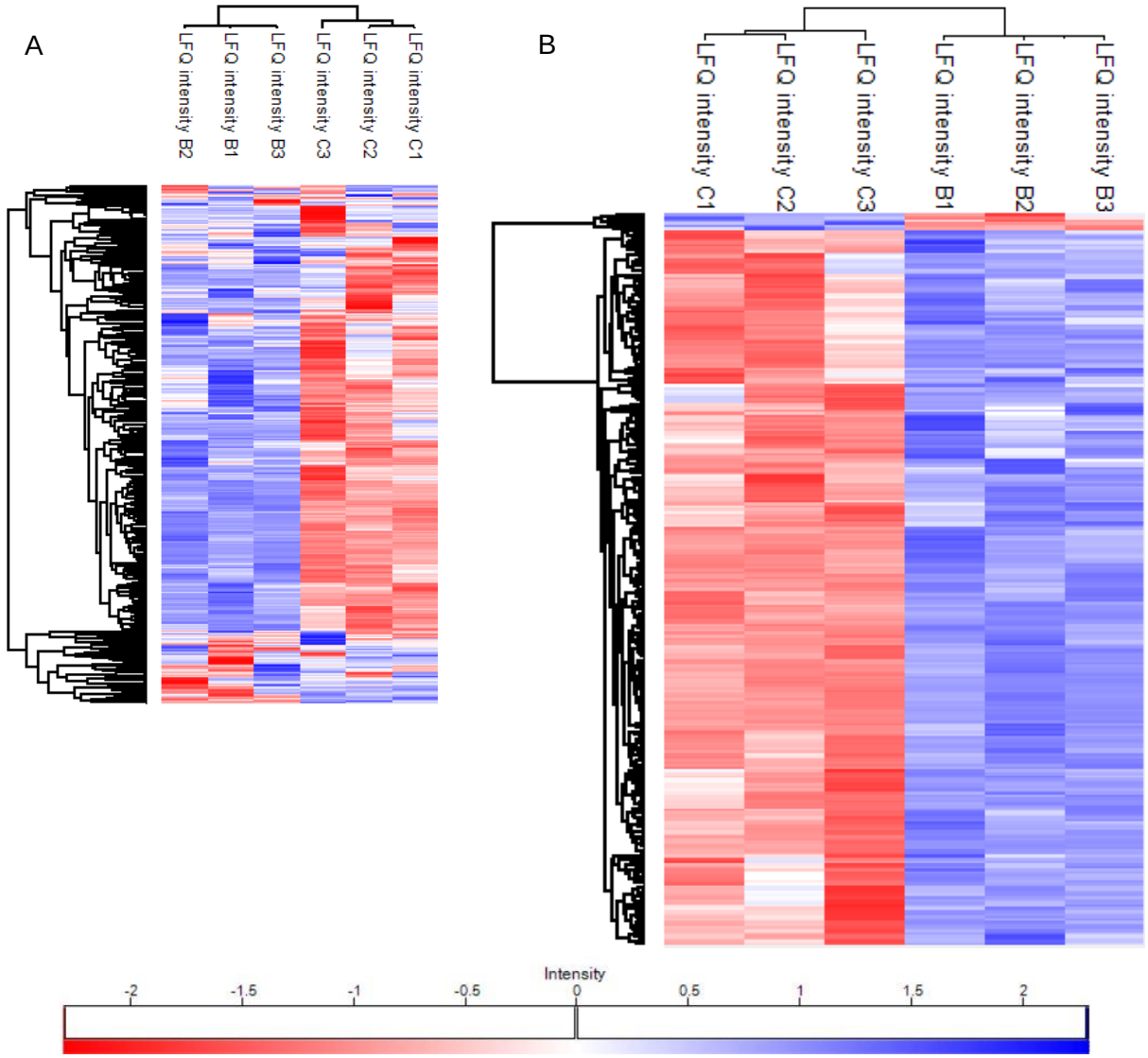

Figure 10: Heat maps of CMB 471 ***A****. Clustering of all the 584 protein groups that were present in each replicate (n=3) of at least one condition group.* ***B****. Clustering of the 266 protein groups that were detected significantly different between condition groups. Clustering was based on protein abundance. Less abundant protein groups are represented in shades of red whereas protein groups that are more abundant are shown in shades of blue.*

Functional enrichment analysis (Table 2) was performed using DAVID 6.8 on the protein groups that were found significantly more in the biotin group compared to the control condition. The amount of surface membrane proteins that were identified confirmed the enrichment of membrane proteins.

A

Category	Term	# (%)	FE	EASE Score	FDR
GOTERM_CC_DIRECT	plasma membrane	87 (31.9)	1.4	**2.6 E^{-5}**	**1.1 E^{-2}**
UP_KEYWORDS	Cell membrane	62 (22.7)	1.5	**1.6 E^{-6}**	**1.3 E^{-4}**
GOTERM_CC_DIRECT	cell surface	25 (9.2)	1.6	**2.8 E^{-3}**	3.4 E^{-1}
KEGG_PATHWAY	Proteoglycans in cancer	16 (5.9)	1.7	**6.6 E^{-3}**	4.8 E^{-1}
B					
GOTERM_CC_DIRECT	plasma membrane	87 (31.9)	1.4	**3.6 E^{-4}**	**5.7 E^{-3}**
UP_KEYWORDS	Cell membrane	62 (22.7)	1.5	**1.4 E^{-3}**	**1.1 E^{-2}**
GOTERM_CC_DIRECT	cell surface	25 (9.2)	3.1	**2.2 E^{-6}**	**6.1 E^{-5}**
KEGG_PATHWAY	Proteoglycans in cancer	16 (5.9)	3.0	**2.3 E^{-4}**	**1.1 E^{-2}**

*Table 2: Functional enrichment analysis of CMB 471 (performed with DAVID (6.8)). The 260 protein groups (273 majority protein IDs) that were found significantly more in the treatment condition (biotinylation) were searched against: **A.** Experimental background (634 majority protein IDs). **B.** Homo sapiens background. FE: Fold enrichment. EASE Score: Modified Fisher Exact p-value (value ranges from 0 to 1, with 0 representing perfect enrichment). FDR: p-value corrected against multiple hypothesis testing using Benjamini-Hochberg method. P-Value ≤ 0.05 are considered strongly enriched (highlighted in bold).*

3.2 Results of the hydrazide experiments

Three different variations of a protocol were used to evaluate the hydrazide approach. The baseline protocol (CMB 467) used hydrazide beads to capture glycosylated surface proteins of which the glycans were oxidised with 2 mM sodium metaperiodate ($NaIO_4$) for 10 min at 4 °C. The other variations used high NaCl concentrations 1.5 M in the wash steps (CMB 479) and in both the wash steps as well as the lysis buffer (CMB 488).

3.2.1 CMB 467: (Hydrazide baseline protocol)

Three replicates of two conditions, membrane protein enrichment using hydrazide capture of oxidised glycoproteins (Ox) and the negative control (C), were analysed using LC-MS/MS. An overview of the analysis features are shown in table 3.

Samples	*MS/MS*	*MS/MS Identified*	*Identification rate*	*Protein identifications*
Ox1	51524	28707	55,72%	2793
Ox2	48223	25129	52,11%	2481
Ox3	50977	28538	55,98%	2951
C1	49219	26424	53,69%	2487
C2	52258	26561	50,83%	2600
C3	52420	27919	53,26%	2741

Table 3: Analysis features overview of the MaxQuant search for CMB 467 (Hydrazide protocol). The number of MS/MS spectra (column 2), the number of these spectra that were identified (column 3), the proportion of MS/MS spectra that were identified was derived by the deviation of column 3 by column 2) (column 4) and the total amount of proteins that were identified in each sample (column 5).

No major differences were seen in the number of identified proteins between the oxidised and control condition as can appreciated from column 5 in table 3.

From the total of 3959 proteins groups that were identified in the MaxQuant search, 3801 remained after filtration and 2485 protein groups (2587 majority protein IDs) had 3 valid LFQ values in at least one group (oxidised and control). Of these 2485 retained protein groups, 114 protein groups showed a significant difference between both groups, with 112 that were significantly increased in the oxidised condition compared to 2 proteins that were significantly increased in the control condition (Figure 11). Here there was no observed shift in the non-significant proteins, the proteins below the significance threshold are equally distributed.

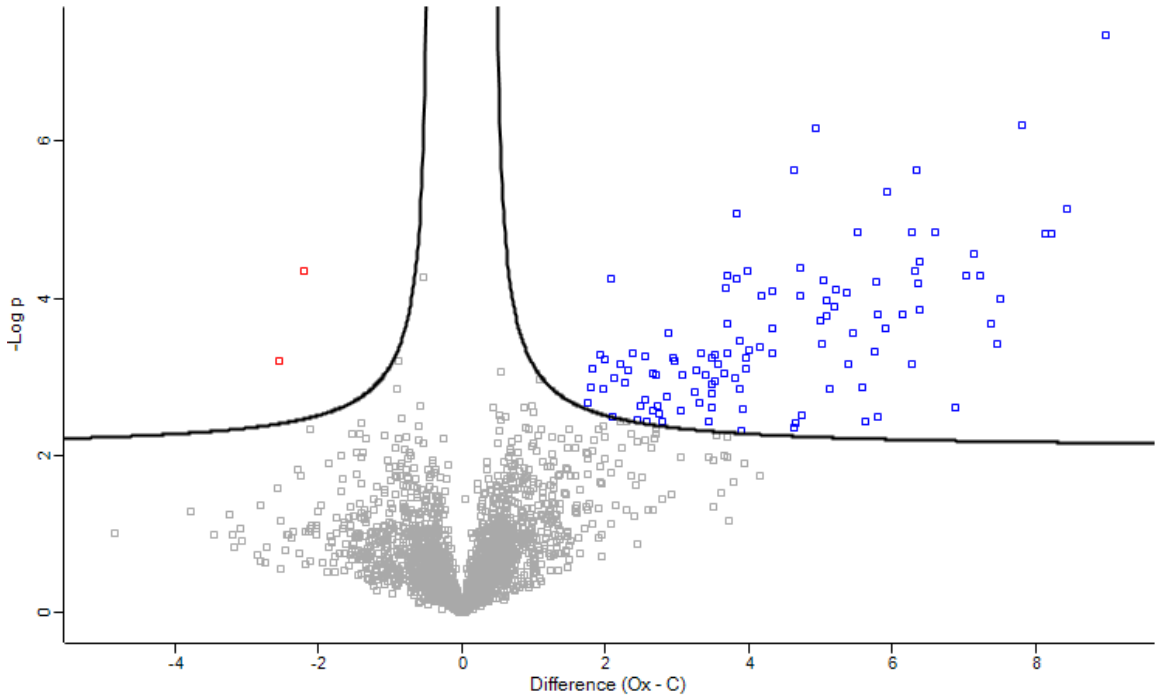

Figure 11 Volcano plot of baseline hydrazide protocol (CMB 467). The significance was plotted on the y-axis versus fold-change on the x-axis. The black line indicates threshold of significance (FDR = 0.01). The 112 protein groups that were significantly enriched in the biotin condition compared to the control condition were marked in blue. The 2 protein groups that were found significantly enriched in the control condition compared to the biotin condition were marked in red.

To check the degree of inter-sample variability heat maps were generated (Figure 12). The heat map in panel A was based on all identified proteins groups and shows considerable variability. The variability between the samples of the same condition is so high that the clustering of the oxidation and the control is not correct. The heat map in panel B was based on the 114 protein groups, where the difference between both conditions was classified as significant. This heat map showed a lot less variability between the samples of one condition and both conditions were grouped into two distinct clusters. Together with the high number of identifications, no difference in amount of identifications between the control and oxidation condition (column 5 table 3) was observed. This might be indicative a large background.

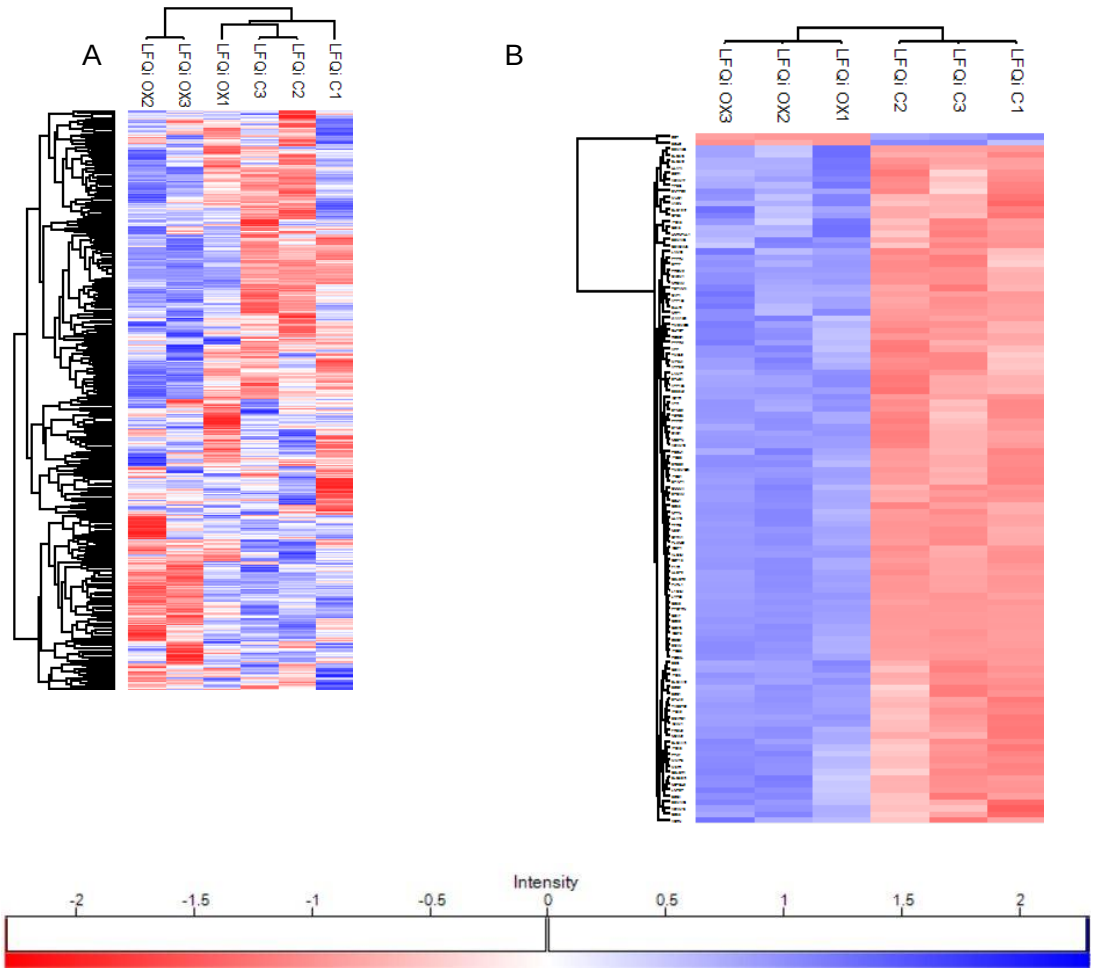

Figure 12: Heat maps of baseline hydrazide protocol (CMB 467) ***A.*** *Clustering of all the 2485 protein groups that were present in each replicate (n=3) of at least one condition group.* ***B****. Clustering of the 114 protein groups that were detected significantly different between condition groups. Clustering was based on protein abundance. Less abundant protein groups are represented in shades of red whereas protein groups that are more abundant are shown in shades of blue.*

Functional enrichment analysis (Table 4) was performed using DAVID 6.8 on the 114 protein groups that were found significantly more in the oxidised condition group compared to the control condition. The level of surface membrane proteins that were identified confirmed that the enrichment of membrane proteins.

A

Category	Term	# (%)	FE	EASE Score	FDR
UP_KEYWORDS	Glycoprotein	106 (93.0)	7.0	**$7.4\ E^{-88}$**	**$1.3\ E^{-85}$**
GOTERM_CC_DIRECT	Plasma membrane	85 (74.6)	3.8	**$3.5\ E^{-38}$**	**$3.7\ E^{-36}$**
UP_KEYWORDS	Cell membrane	69 (60.5)	5.2	**$3.6\ E^{-37}$**	**$1.1\ E^{-35}$**
GOTERM_CC_DIRECT	cell surface	44 (38.6)	9.4	**$5.5\ E^{-34}$**	**$3.9\ E^{-32}$**
KEGG_PATHWAY	Proteoglycans in cancer	11 (9.6)	4.3	**$1.2\ E^{-4}$**	**$1.7\ E^{-3}$**
B					
UP_KEYWORDS	Glycoprotein	106 (93.0)	4.2	**$4.3\ E^{-59}$**	**$7.6\ E^{-57}$**
GOTERM_CC_DIRECT	Plasma membrane	85 (74.6)	3.3	**$1.5\ E^{-31}$**	**$1.1\ E^{-29}$**
UP_KEYWORDS	Cell membrane	69 (60.5)	3.9	**$1.6\ E^{-27}$**	**$4.0\ E^{-26}$**
GOTERM_CC_DIRECT	cell surface	44 (38.6)	13.0	**$1.9\ E^{-36}$**	**$4.0\ E^{-34}$**
KEGG_PATHWAY	Proteoglycans in cancer	11 (9.6)	5.1	**$4.3\ E^{-5}$**	**$6.0\ E^{-4}$**

Table 4: Functional enrichment analysis of CMB 467 (performed with DAVID (6.8)). The 112 protein groups (114 majority protein IDs) that were found significantly more in the treatment condition (oxidised) were searched against: ***A.*** *experimental background (2587 majority protein IDs).* ***B.*** *homo sapiens background. FE: Fold enrichment. EASE Score: Modified Fisher Exact p-value (value ranges from 0 to 1, with 0 representing perfect enrichment). FDR: p-value corrected against multiple hypothesis testing using Benjamini-Hochberg method. P-Value ≤ 0.05 are considered strongly enriched (bold).*

3.2.2 CMB 479: (Hydrazide protocol with 1.5 M NaCl in the wash steps)

Three replicates of two conditions, membrane protein enrichment using hydrazide capture of oxidised glycoproteins (Ox) and the negative control (C), were analysed using LC-MS/MS. An overview of the analysis features are shown in table 5.

Samples	*MS/MS*	*MS/MS Identified*	*Identification rate*	*Protein identifications*
Ox1	37077	12840	34.63%	753
Ox2	36828	11701	31.77%	788
Ox3	37039	12622	34.08%	784
C1	35985	7993	22.21%	561
C2	35692	4916	13.77%	341
C3	36546	5471	14.97%	472

Table 5: Analysis features overview of the MaxQuant search for CMB 479 (Hydrazide protocol with 1.5 M NaCl in wash buffer). The number of MS/MS spectra (column 2), the number of these spectra that were identified (column 3), the proportion of MS/MS spectra that were identified was derived by the deviation of column 3 by column 2) (column 4) and the total amount of proteins that were identified in each sample (column 5).

The differences in the number of protein identifications between the oxidised and control condition as can appreciated form column 5 in table 5 was higher than in the previous experiment (~1.7 fold more proteins identified in the oxidised condition compared to similar numbers of identifications in the previous).

From the total of 1350 proteins groups that were identified in the MaxQuant search, 1213 remained after filtration and 601 protein groups (657 majority protein IDs) had 3 valid LFQ values in at least one group (oxidised and control). Of these 601 retained protein groups, 285 protein groups showed a significant difference between both groups, with 245 protein groups that were significantly more abundant in the oxidised condition compared to 40 protein groups that were found significantly more abundant in the control condition (Figure 13).

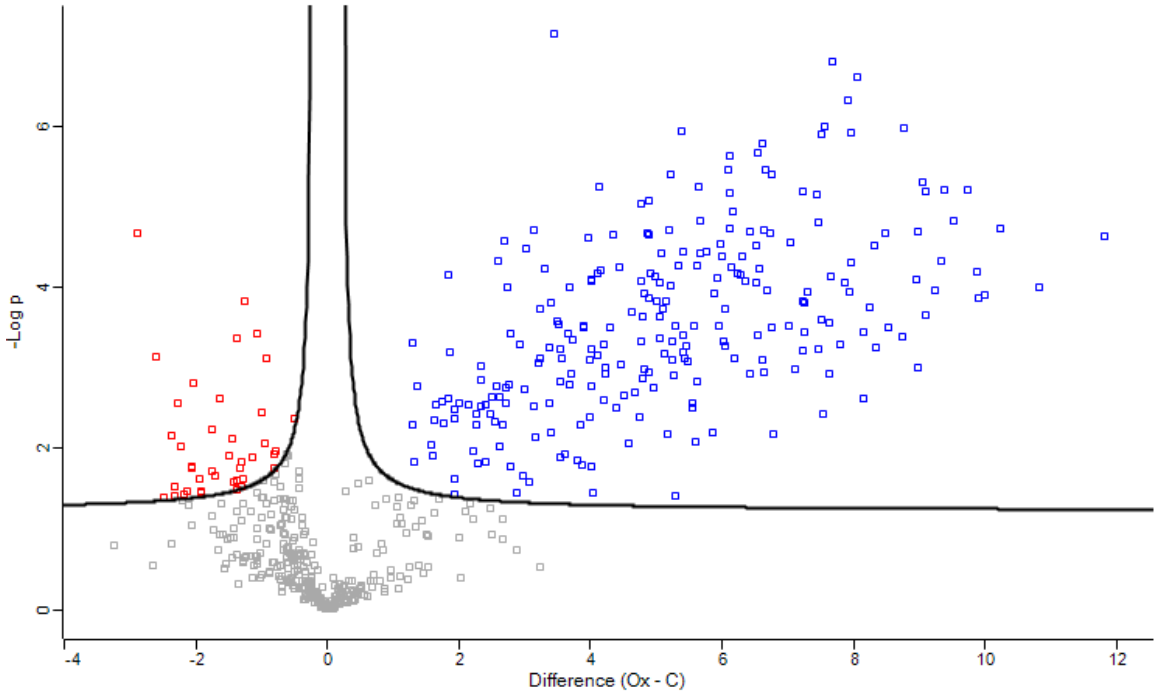

Figure 13: Volcano plot of hydrazide protocol using 1.5 M NaCl in wash steps (CMB 479). The significance was plotted on the y-axis versus fold-change on the x-axis. The black line indicates threshold of significance (FDR = 0.01). The 245 protein groups that were significantly enriched in the biotin condition compared to the control condition were marked in blue. The 40 protein groups that were significantly enriched in the control condition compared to the biotin condition were marked in red.

To check the degree of inter-sample variability heat maps were generated (Figure 14). Replicates form the same experimental condition clustered together with limited variability between the samples of a single condition.

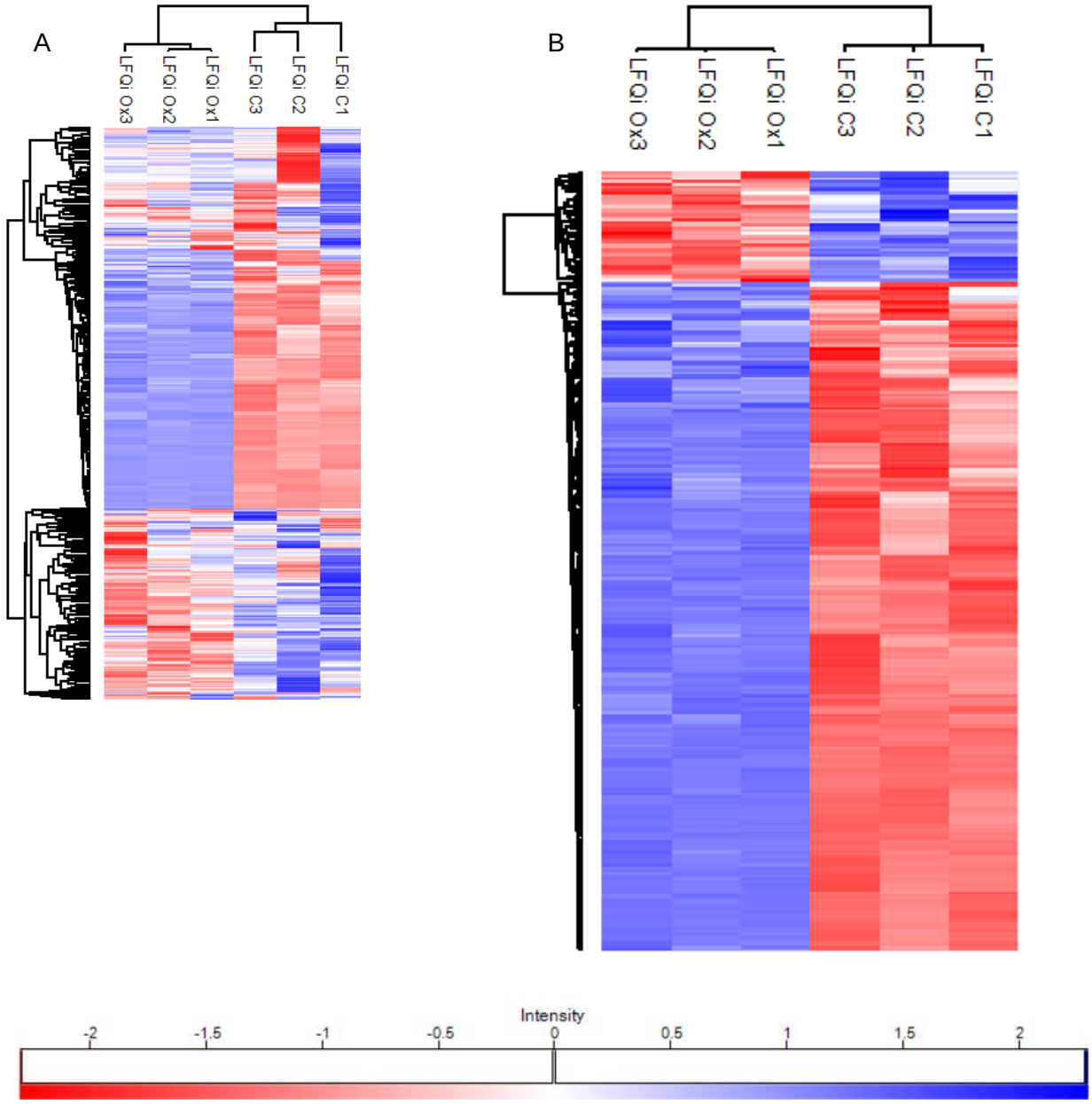

Figure 14:Heat maps of hydrazide protocol using 1.5 M NaCl in wash steps (CMB 479) ***A.*** *Clustering of all the 601 protein groups that were present in each replicate (n=3) of at least one condition group.* ***B.*** *Clustering of the 285 protein groups that were detected significantly different between condition groups. Clustering was based on protein abundance. Less abundant protein groups are represented in shades of red whereas protein groups that are more abundant are shown in shades of blue.*

Functional enrichment analysis (Table 6) was performed using DAVID 6.8 on the protein groups that were found significantly more in the oxidised condition group compared to the control condition. The level of surface membrane proteins that were identified confirmed that the enrichment of membrane proteins.

A

Category	Term	# (%)	FE	EASE Score	FDR
UP_KEYWORDS	Glycoprotein	229 (91.6)	2.3	**4.8 E^{-107}**	**1.1 E^{-104}**
GOTERM_CC_DIRECT	Plasma membrane	182 (72.8)	1.7	**2.2 E^{-37}**	**2.0 E^{-35}**
UP_KEYWORDS	Cell membrane	148 (59.2)	2.0	**5.5 E^{-36}**	**1.8 E^{-34}**
GOTERM_CC_DIRECT	cell surface	77 (30.8)	2.2	**1.6 E^{-22}**	**1.0 E^{-20}**
KEGG_PATHWAY	Proteoglycans in cancer	18 (7.2)	1.8	**5.9 E^{-3}**	1.7 E^{-1}
B					
UP_KEYWORDS	Glycoprotein	229 (91.6)	4.1	**4.7 E^{-124}**	**1.1 E^{-121}**
GOTERM_CC_DIRECT	Plasma membrane	182 (72.8)	3.2	**4.5 E^{-64}**	**6.0 E^{-62}**
UP_KEYWORDS	Cell membrane	148 (59.2)	3.8	**1.5 E^{-56}**	**5.1 E^{-55}**
GOTERM_CC_DIRECT	cell surface	77 (30.8)	10.4	**8.9 E^{-56}**	**7.9 E^{-54}**
KEGG_PATHWAY	Proteoglycans in cancer	18 (7.2)	4.5	**3.4 E^{-7}**	**1.3 E^{-5}**

*Table 6: Functional enrichment analysis of CMB 479 (performed with DAVID (6.8)). The 245 protein groups (250 majority protein IDs) that were found significantly more in the treatment condition (oxidised) were searched against: **A.** experimental background (657 majority protein IDs). **B.** homo sapiens background. FE: Fold enrichment. EASE Score: Modified Fisher Exact p-value (value ranges from 0 to 1, with 0 representing perfect enrichment). FDR: p-value corrected against multiple hypothesis testing using Benjamini-Hochberg method. P-Value ≤ 0.05 are considered strongly enriched (bold).*

3.2.3 CMB 488: (Hydrazide protocol with 1.5 M NaCl in the lysis buffer)

Three replicates of two conditions, membrane protein enrichment using hydrazide capture of oxidised glycoproteins (Ox) and the negative control (C), were analysed using LC-MS/MS. An overview of the analysis features are shown in table 7.

Samples	*MS/MS*	*MS/MS Identified*	*Identification rate*	*Protein identifications*
Ox1	22488	7724	35.48%	746
Ox2	20678	6251	31.05%	818
Ox3	21639	6126	28.87%	1179
C1	21532	5994	28.49%	781
C2	17319	4402	26.17%	665
C3	18191	4591	25.94%	619

Table 7: Analysis features overview of the MaxQuant search for CMB 488 (Hydrazide protocol with 1.5 M NaCl in lysis buffer). The number of MS/MS spectra (column 2), the number of these spectra that were identified (column 3), the proportion of MS/MS spectra that were identified was derived by the deviation of column 3 by column 2) (column 4) and the total amount of proteins that were identified in each sample (column 5).

No major differences were seen in protein identifications between the oxidised and control condition as can appreciated form column 5 in table 7.

From the total of 1681 proteins groups that were identified in the MaxQuant search, 1602 remained after filtration and 649 protein groups (720 protein IDs) had 3 valid LFQ values in at least one group (oxidised and control). Of these 649 retained protein groups, 45 protein groups showed a significant difference between both groups, with 43 that were found significantly more in the oxidised condition compared to 2 that were found significantly more in the control condition (Figure 15).

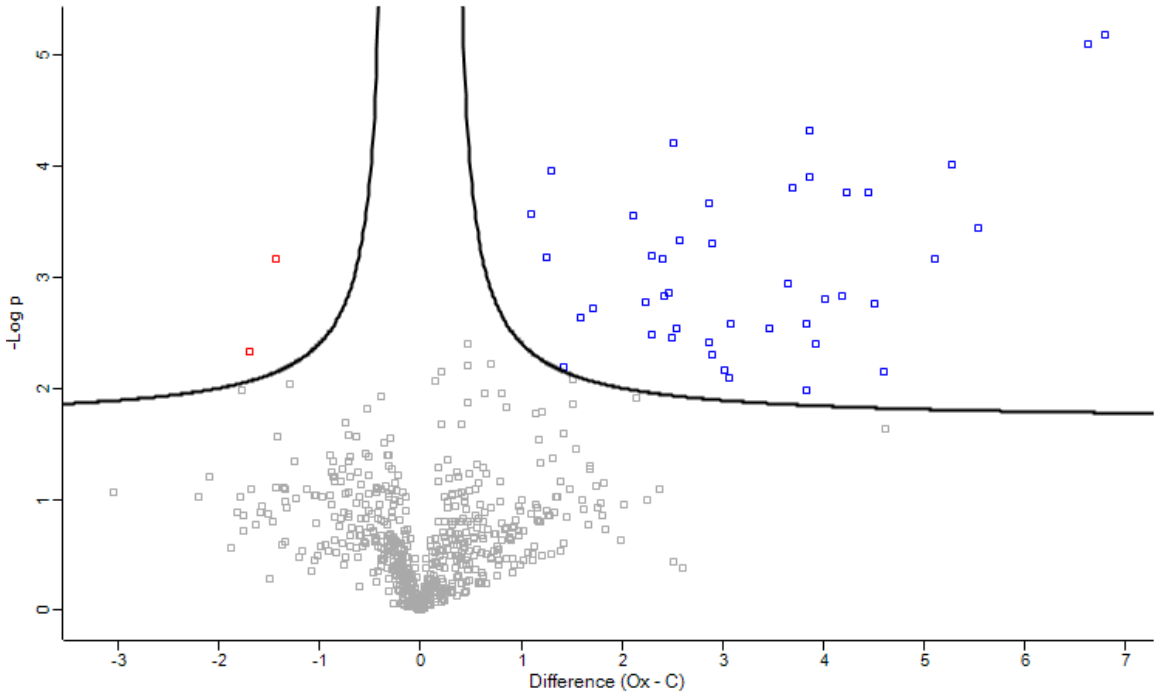

Figure 15: Volcano plot of hydrazide protocol using 1.5 M NaCl in wash steps and lysis buffer (CMB 488). The significance was plotted on the y-axis versus fold-change on the x-axis. The black line indicates threshold of significance (FDR = 0.01). The 43 protein groups that were significantly enriched in the biotin condition (blue) compared to the control condition were marked in blue. The 2 protein groups that were significantly enriched in the control condition compared to the biotin condition were marked in red.

To check the degree of inter-sample variability heat maps were generated (Figure 16). The heat map in panel A was based on all identified proteins groups and shows considerable variability between all samples and impaired clustering. The heat map in panel B was based on the 45 protein groups that were significantly different between both conditions (oxidised and control). These 45 proteins show limited inter-sample variability and form two distinct clusters.

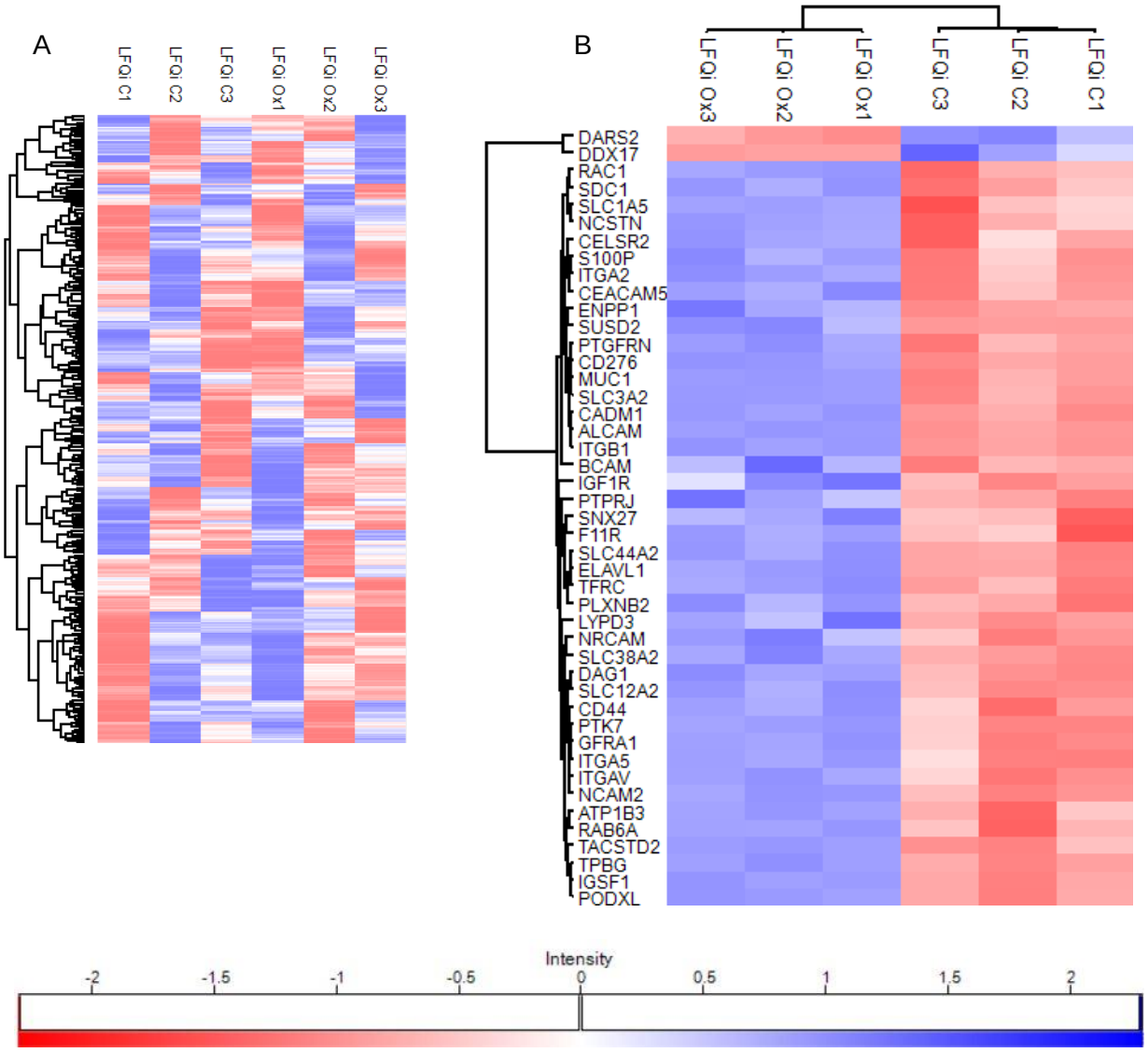

Figure 16: Heat maps of hydrazide protocol using 1.5 M NaCl in the wash steps and lysis buffer (CMB 488) ***A.*** *Clustering of all the 649 protein groups that were present in each replicate (n=3) of at least one condition group.* ***B.*** *Clustering of the 45 protein groups that were detected significantly different between condition groups. Clustering was based on protein abundance. Less abundant protein groups are represented in shades of red whereas protein groups that are more abundant are shown in shades of blue.*

Functional enrichment analysis (Table 8) was performed using DAVID 6.8 on the protein groups that were found significantly more in the oxidised condition group compared to the control condition which confirmed the enrichment of membrane proteins.

A

Category	Term	# (%)	FE	EASE Score	FDR
UP_KEYWORDS	Glycoprotein	40 (87.0)	6.1	**$2.3\ E^{-30}$**	**$2.4\ E^{-28}$**
GOTERM_CC_DIRECT	Plasma membrane	31 (67.4)	3.3	**$5.4\ E^{-12}$**	**$2.4\ E^{-10}$**
UP_KEYWORDS	Cell membrane	28 (60.9)	4.6	**$1.8\ E^{-14}$**	**$2.4\ E^{-13}$**
GOTERM_CC_DIRECT	cell surface	13 (28.3)	4.7	**$3.0\ E^{-6}$**	**$7.9\ E^{-5}$**
KEGG_PATHWAY	Proteoglycans in cancer	8 (17.4)	5.7	**$1.6\ E^{-4}$**	**$3.6\ E^{-3}$**
B					
UP_KEYWORDS	Glycoprotein	40 (87.0)	3.9	**$4.5\ E^{-20}$**	**$4.6\ E^{-18}$**
GOTERM_CC_DIRECT	Plasma membrane	31 (67.4)	3.0	**$3.4\ E^{-10}$**	**$1.5\ E^{-8}$**
UP_KEYWORDS	Cell membrane	28 (60.9)	3.9	**$1.1\ E^{-11}$**	**$3.7\ E^{-10}$**
GOTERM_CC_DIRECT	cell surface	13 (28.3)	9.5	**$5.0\ E^{-9}$**	**$1.1\ E^{-7}$**
KEGG_PATHWAY	Proteoglycans in cancer	8 (17.4)	9.5	**$1.1\ E^{-5}$**	**$3.5\ E^{-4}$**

*Table 8: Functional enrichment analysis of CMB 488 (performed with DAVID (6.8)). The 43 protein groups (46 majority protein IDs) that were found significantly more in the treatment condition (oxidised) were searched against: **A.** experimental background (720 majority protein IDs). **B.** Homo sapiens background. FE: Fold enrichment. EASE Score: Modified Fisher Exact p-value (value ranges from 0 to 1, with 0 representing perfect enrichment). FDR: p-value corrected against multiple hypothesis testing using Benjamini-Hochberg method. P-Value ≤ 0.05 are considered strongly enriched (bold).*

3.3 Comparison

In order to provide a completer picture of the composition of the surfaceome, the results of both approaches were compared to check whether these merely confirmed each other or worked complementary. The Venn diagrams below were generated using an online bioinformatics tool: http://bioinformatics.psb.ugent.be/webtools/Venn/

3.3.1 Comparison between hydrazide approaches

Comparison of the three different versions of the hydrazide approach: baseline hydrazide protocol (CMB 467), hydrazide protocol with 1.5 M NaCl in the wash steps (CMB 479) and hydrazide protocol with 1.5 M NaCl already in the lysis buffer (CMB 488) allowed to select the protocol with the highest enrichment. The differences in significant identifications are depicted in figure 17 where the high number of unique identifications by CMB 479 stands out. The use of a higher concentration of NaCl in the wash steps, resulted in less protein identifications, but with a higher enrichment (more identifications that were significant). Further adaptations to the protocol by using this high concentration of NaCl already in the lysis buffer resulted in equal amounts of identifications, but with lower numbers of significant modifications.

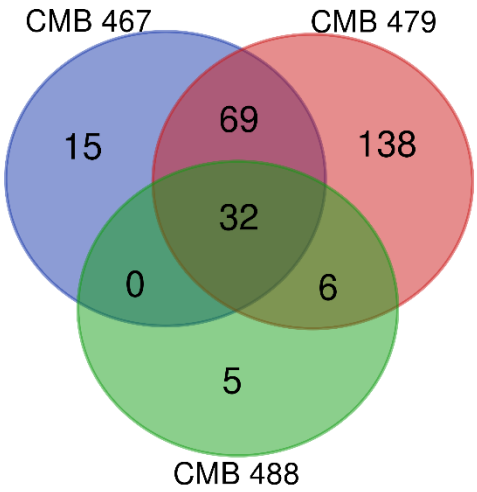

Figure 17: Venn diagram comparing protein identifications between the three different hydrazide protocols. The limited amount of unique identifications in the CMB 467 and CMB 488 demonstrates the higher performance of the protocol used in CMB 479 (= Hydrazide protocol with 1.5 M NaCl in the wash steps)

3.3.2 Comparison between hydrazide and biotin approach

Based on the comparison of the hydrazide approaches and the individual parameters we selected the second version of the hydrazide protocol (CMB 479) with 1.5 M in the wash steps. This was then compared to the biotin approach (CMB 471) to check whether the use of both orthogonal strategies could work complementary in unravelling the protein composition of the cell membrane. The observed limited overlap in figure 18 suggests that these two approaches can indeed complement each other.

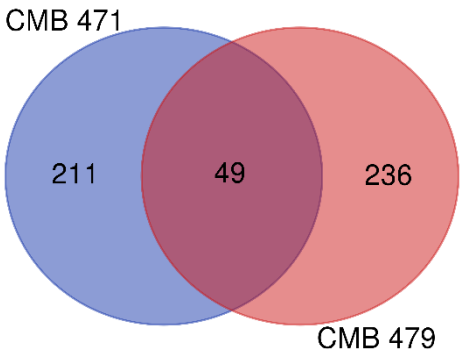

Figure 18: Venn diagram comparing majority protein identifications of biotinylation (CMB 471) and hydrazide (CMB 479) protocols. Limited overlap between both enrichment strategies.

An overview of the overlap between the two approaches was depicted in figure 20. In addition, a complete listing of the gene names of all identified majority proteins is provided in table 9. This allowed a more in depth analysis of the composition of both the intersection and disjoint sets. The sensitivity of these approaches was checked by the presence of known surface markers for breast cancer. This listing was screened for following tumour associated antigens: mucin-1 (MUC-1), carcinoembryonic antigen-related cell adhesion molecule 5 - 6 (CEACAM5 and CEACAM6), HER2 (ERBB2) and tyrosine kinase receptors (RTKs). The retrieved markers were indicated in bold.

Approach	#	Majority proteins
CMB 471 (Biotin)	(211)	ESYT1, MTDH, THOC2, PDCD11, PDIA3, HEATR1, HIP1R, IARS, VPS13A, CAND1, ARHGEF1, PKM, SPTBN1, COPA, HSD17B4, PRKDC, TARS, DDX21, CCT4, DHCR24, FASN, (ACTB;ACTG1), EFTUD2, KNTC1, UNC45A, XPO1, (RPS27A;UBA52;UBB;UBC), MYO5B, AP1G1, MTHFD1, FLII, ATRX, HSPA8, NNT, EZR, HK1, ECPAS, EVPL, DOP1B, SCRIB, LAMA5, IQGAP1, XRN2, SAMD9, PFKP, OGT, USP24, ALDH18A1, PARP1, RPL21, ILF3, (HSPA1B;HSPA1A), (RPL39P5;RPL39), MYH14, HNRNPK, (TUBA1B;TUBA4A), HIST1H1D, PDS5B, FANCI, TCOF1, SPTBN2, LONP1, ATP2A2, TRAP1, CHD4, GCN1, LGALS3, MACF1, DDX23, CYFIP1, RPL18, EPHX1, NRK, SMC3, DHX38, CANX, WASHC5, RPN1, UGDH, TLN1, LAMB2, CTNNB1, PRPF8, GAPVD1, CUL4A, ATAD2, PRPF6, EIF4A1, RANBP2, UBR5, TENM3, MTOR, GTF3C1, CTNNA2, MYO6, NSF, EDC4, UBR4, RPL15, RPL36, RPL3, RAD50, MYOF, TUBB, MADD, DNMT1, FAT1, NUP160, VAV2, AP3D1, CLTC, SMC1A, FAM129B, RPL6, HK2, SND1, LARS, POLR2B, EPPK1, ABCF1, PLEC, MCM7, RRP12, ATP1A1, ABAT, LMNA, DHTKD1, ARHGAP35, MYO1B, DENND4C, ATP5F1A, USP9X, SON, PLXNA3, TMEM263, ADAR, EIF3L, GLUD1, CNOT1, UPF2, RSL1D1, (DDX3Y;DDX3X), COPB2, NUP210, FERMT2, ESYT2, HIST1H1B, MMS19, TRIP12, PCCB, DDX54, BIRC6, QARS, DNAJC13, XAB2, HUWE1, P4HB, PLCG1, (HIST2H2BE;HIST1H2BB;HIST1H2BO;HIST1H2BJ;HIST3H2BB), CAD, VCP, CHD8, (EEF1A1;EEF1A1P5), ARFGEF2, CASK, NCBP1, SEC24C, MSH6, SF3B3, TOP2B, TNPO2, RRBP1, SNRNP200, HTT, VARS, PDCD6IP, AP2A1, TRRAP, BPTF, DYNC1H1, UPF1, PRKCD, NUP155, DNM2, HSPB1, MCM2, TOP1, JUP, RAB3GAP2, HECTD1, HSPA9, DSP, TUBB4B, IQGAP3, ACSL3, ACSL4, UBE3C, SMG1, URB1, ACACB, PREX1, NUP205, USP7, DHX9, RPL4, NCAPD2, LLGL2, EPRS, MYO5C, ATP5F1B, ITPR3, RECQL
CMB 471 + CMB 479	(49)	SLC39A10, ITGA2, PLXNB2, SLC39A6, ENPP1, ITGB5, CD59, BCAM, FREM2, ITGB4, PODXL, EPHB4, TFRC, ALCAM, DDR1, DSG2, PTPRK, ADAM10, CEMIP2, NRP1, **ERBB3**, NRCAM, SEMA3C, CELSR2, TRPM4, SLC3A2, PTGFRN, IGSF3, GFRA1, SDC1, CD44, HSPA5, CNNM4, PTPRF, CELSR1, DAG1, PLXNA1, **MUC1**, AGRN, ITGB1, CD109, CDH1, **PTK7**, **IGF1R**, PIEZO1, NAMPT, IGF2R, RPL13, **EPHA1**
CMB 479 (Hydrazide)	(236)	AIP, APMAP, CD276, SLC12A2, EPHB3, ADAM17, SDC4, RAB18, (ULBP2;RAET1L), (HLA-Cw12; HLA-Cw14; HLA-Cw16), DNAJA2, PLXNB1, TSPAN6, SLITRK6, VSIG10, MAL2, SORT1, SLC25A3, SMAGP, SCARB1, SDC3, GPRC5A, RET, SLC7A5, CA12, LNPEP, NPTN, PLPP1, CD99, FLRT3, BAG6, ATP6V0D1, YWHAZ, SLC44A1, PVR, CADM1, CD47, SPINT1, KIAA1549, FTSJ3, (RAC1;RAC3), **CEACAM5**, ATP5PD, PRSS23, MPZL1, ATRN, ETFA, RAB5C, RAB5B, NECTIN1, TMEM8A, SYT1, KLRG2, KIAA1324, AREG, AOC1, IFNAR1, S100A16, CXCL16, TRA2B, RPL7A, RPL23A, (RAB2A;RAB2B), FAM174B, PCDH1, RTN4R, TLR2, ST14, LAMP1, SEMA4B, EPCAM, TM9SF3, BST2, SLC4A7, IGSF1, SDC2, CAP1, RPL8, TXNDC12, EFNB2, PODXL2, SLC39A14, RAB7A, TSPAN13, L1CAM, RELL1, SLC38A2, EMB, PTPRG, ITFG1, NOTCH2, (HLA-B18;HLA-B38;HLA-B39;HLA-B41;HLA-B67), CD82, LGALS3BP, RHOC, EPHA2, HLA-A, M6PR, ITGAV, NCAM2, PRSS8, GLG1, CADM4, SLC5A6, SLC4A2, FAM234A, NECTIN2, F11R, SLC9A1, CD55, PSMC6, SUSD2, SEZ6L2, ATP1B1, RNF149, ITGA3, ICOSLG, GNS, NPY1R, DCBLD2, ITGA6, TACSTD2, SLC7A2, PRRT3, ANO6, FZD6, PTPRJ, MMP15, MEGF9, ECE1, NECTIN4, LINGO1, (RAB11A;RAB11B), MUC5B, NEO1, TMEM154, SRRM1, ROBO1, ENTPD2, CD151, CDH3, LRRC8A, LAP3, APP, ATP1B3, SLC38A1, ENTPD8, ADAM15, EFNB1, YBX1, FOLR1, RAB21, VASN, RAB1A, TM2D3, SDK1, TMEM30A, CD58, CD46, LDLR, RPL14, EPHA7, **ERBB2**, PROM2, SLC12A7, IGF2, ICAM1, CLU, (RAP1B;RP1BL), ERBB4, PSAP, LRRC8D, INSR, IL6ST, CD97, CD99L2, SLC6A14, ENG, MICB, SLC6A6, LYPD3, **CEACAM6**, CLIC3, SPINT2, PSMD12, PTPRA, RAB14, TSPAN15, SEMA4C, HLA-F, RAB5A, SEMA4D, SLC44A2, SRSF1, S100A11, RAB1B, SLC1A5, ADGRG6, ABCC1, ITGA5, ADGRL1, SLC29A1, BSG, CD9, TMEM9, TMEM132A, EIF4G1, SYPL1, IGFBP5, STC2, RPL18A, TMEM30B, PLXND1, MYORG, TMEM248, RAB3D, SULF2, NPNT, (ATP1A1;ATP1A3), CLDND1, CRB3, VSIG2, PTPRS, IL10RB, (CLDN3;CLDN4;CLDN6;CLDN9), NCSTN, SLC30A1, TPBG, ITGB6, NOTCH3, SLC2A1, SLITRK4, ADAM9, CPD, FZD2, CPM

Table 9: Listing of identified majority protein groups represented in figure 20. This table shows the gene names of the majority protein groups identified, using the biotin protocol (CMB 471) and the superior hydrazide approach (CMB 479). Majority proteins IDs of the same protein group are grouped using brackets. Important known cell surface markers are highlighted in bold: MUC-1, CEACAM5, CEACAM6, ERBB2 (=HER2) and multiple other tyrosine kinase receptors (RTKs) (ERBB3, PTK7, GF1R and EPHA1).

Figure 19 gives an overview of which of the eight surface tumour markers (MUC1, CEACAM5, CEACAM6, ERBB2, ERBB3, PTK7, GF1R, EPHA1) each protocol was able to pick up.

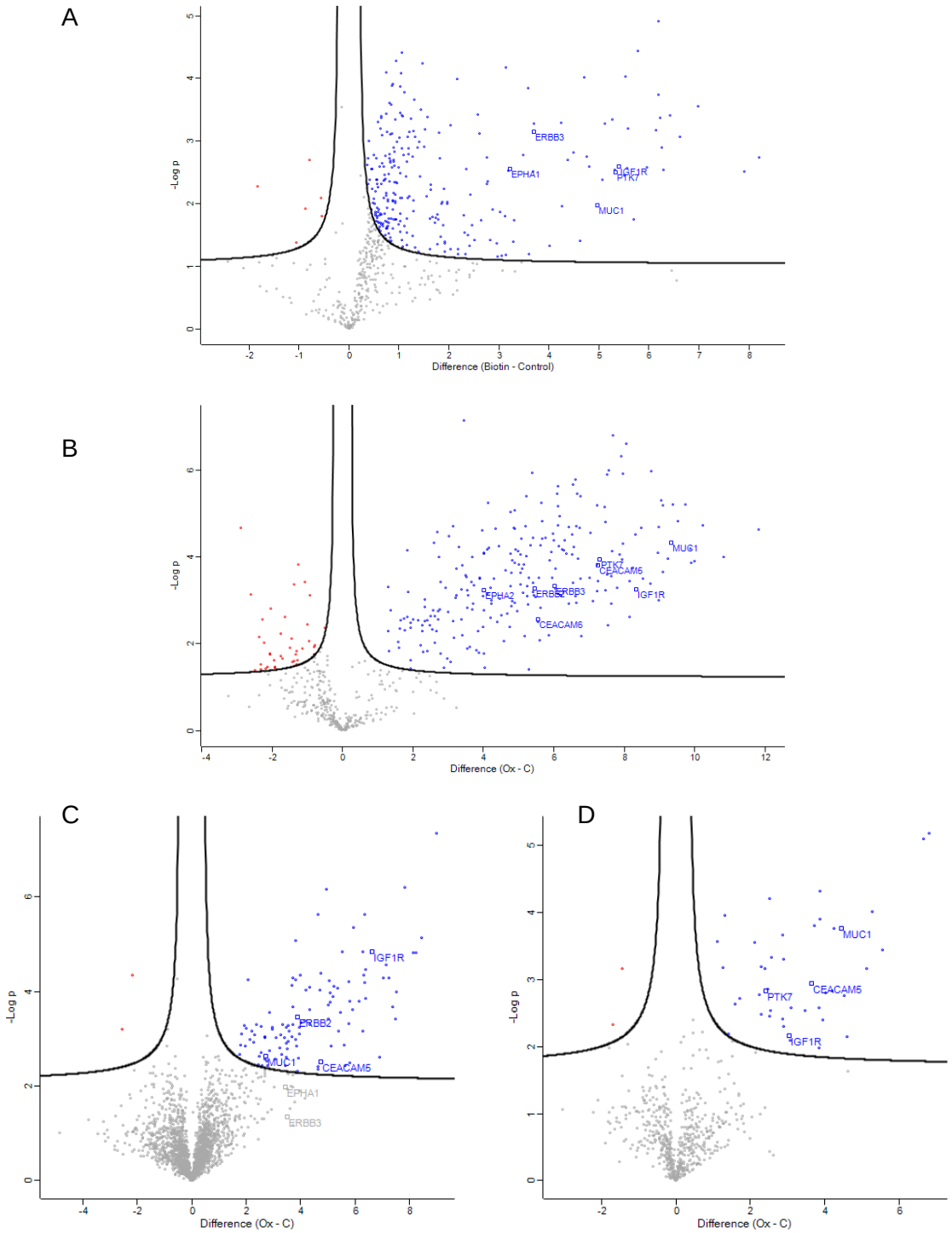

Figure 19: Overview of the scatterplots showing the tumour markers each protocol could detect. ***A.*** *The biotin protocol (CMB 471) picked up 5 tumour markers which were all significantly enriched.* ***B.*** *Hydrazide protocol with 1.5 M NaCl in the wash steps (CMB 479) picked up 8 tumour markers and which all significantly enriched.* ***C.*** *Baseline hydrazide protocol (CMB 471) picked up 6 tumour markers of which 4 were significant enriched.* ***D.*** *Hydrazide experiment that used 1.5 M NaCl in the lysis buffer (CMB 488) picked up 4 tumour markers which were all significantly enriched.*

Functional enrichment analysis (Table 10) using DAVID 6.8 was performed on the 475 significant enriched protein groups found by the combination of the biotin protocol (CMB 471) and the optimised hydrazide protocol (CMB 479).

Category	Term	# (%)	FE	EASE Score	FDR
UP_KEYWORDS	Glycoprotein	252 (53.1)	2.4	**$7.2\ E^{-50}$**	**$1.4\ E^{-47}$**
GOTERM_CC_DIRECT	Plasma membrane	230 (48.4)	2.1	**$9.4\ E^{-36}$**	**$8.1\ E^{-34}$**
UP_KEYWORDS	Cell membrane	175 (36.8)	2.4	**$9.9\ E^{-31}$**	**$6.3\ E^{-29}$**
GOTERM_CC_DIRECT	cell surface	81 (17.1)	5.7	**$8.9\ E^{-38}$**	**$1.2\ E^{-35}$**
KEGG_PATHWAY	Proteoglycans in cancer	27 (5.7)	3.2	**$2.1\ E^{-7}$**	**$1.6\ E^{-5}$**

Table 10: Functional enrichment analysis of combination between CMB 471 and CMB 479 (performed with DAVID (6.8)). The 457 protein groups (475 protein IDs) that were found significantly more in the treatment condition (biotinylation oxidised) of both experiments were searched against the Homo sapiens background. FE: Fold enrichment. EASE Score: Modified Fisher Exact p-value (value ranges from 0 to 1, with 0 representing perfect enrichment). FDR: p-value corrected against multiple hypothesis testing using Benjamini-Hochberg method. P-Value ≤ 0.05 are considered strongly enriched (highlighted in bold).

To assure that selection of one hydrazide protocol did not exclude information, the overlap of all protocols was examined and is shown in the Venn diagram of figure 20. The limited additional identifications confirmed that the hydrazide protocol with 1.5 M in the wash steps outcompeted the other versions of the hydrazide protocol.

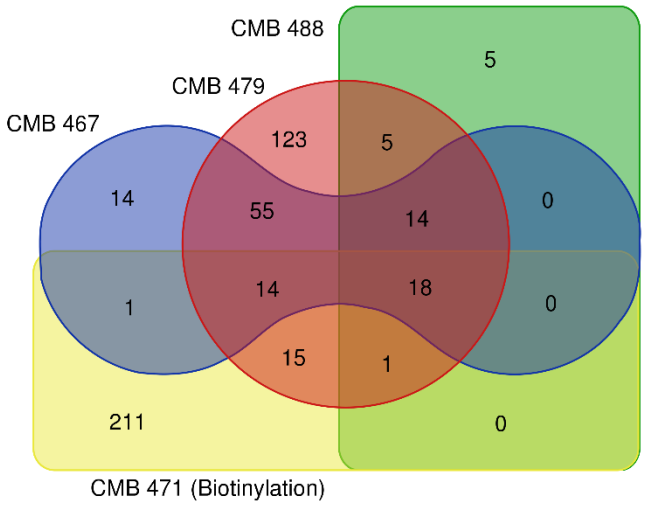

Figure 20: Venn diagram showing protein identifications across all experiments. Limited additional overlap between the other hydrazide protocols (CMB 467/488) and the biotinylation protocol (CMB 471) confirms the superiority of the hydrazide protocol used in CMB 479.

4. Discussion

The biotin approach (CMB 471) was able to identify 584 protein groups that were present in all three replicates of at least one group. Using the hydrazide approach this number of identifications was 2485 for the baseline protocol (CMB 467), 601 for the hydrazide protocol with 1.5 M NaCl in the wash steps (CMB 479) and 649 in the hydrazide protocol with 1.5 M NaCl in the lysis buffer. These are all similar results with the exception of the high amount of protein groups in the baseline hydrazide protocol (CMB 467). Functional enrichment analysis showed a significant enrichment against the *Homo sapiens* background for cell surface proteins in all four experiments: p-value of $2.2E^{-6}$, 3.1 FE (CMB 471), $1.9E^{-36}$, 13.0 FE (CMB 467), $8.9E^{-56}$, 10.4 FE (CMB 479) and $5.0E^{-9}$, 9.5 FE (CMB 488). Elevating the NaCl concentration in the wash steps to 1.5 M resulted in less protein identifications than in the baseline protocol, but with a higher amount of significant identifications. The higher salt concentration probably helped to remove the nonspecific interactors from the beads resulting in a lower background signal. This prompted the use of a high concentration of NaCl in the lysis buffer and although this resulted in similar numbers of identifications as in the hydrazide protocol with high concentrations of NaCl in the wash steps, there were a lot less significant enriched protein groups identified (43 significantly enriched protein groups compared with 245 significantly enriched protein groups). Secondly, there was a major difference between the significance of the enrichment of surface proteins: p-value of $8.9E^{-56}$ (CMB 479) compared to $5.0E^{-9}$ (CMB 488). This was also reflected in the absolute amount of proteins that were catalogued as surface proteins 77 (CMB 479) versus 13 (CMB 488). Moreover, 91.6% of the identified protein groups in CMB 479 were annotated as glycoproteins and there was a 10.4-fold enrichment of cell surface proteins against the *Homo sapiens* background. Based on these superior results the hydrazide protocol, with 1.5 M NaCl in the wash steps, (CMB 479) was selected as the hydrazide approach of choice to combine with the biotin approach.

The combination of the biotin (CMB 471) and the hydrazide approach (CMB 479) was evaluated by the screening for the presence of known surface markers for breast cancer. The specific selection of these tumour associated antigens was based on the studies of Ziegler *et al.* [25] and Criscitiello [73]. The presence of ERBB2, mucin-1 (MUC-1), carcinoembryonic antigen-related cell adhesion molecule 5 & 6 (CEACAM5, CEACAM6) and multiple tyrosine kinase receptors (RTKs) in the dataset, indicate that these enrichment approaches enable the detection of these markers.

When the hydrazide approach was compared with the biotin protocol, there was only a limited overlap between these two (Figure 18). This was expected, because both approaches target a distinctive subset of proteins. The hydrazide approach only targets glycosylated proteins while the biotinylation approach targets every protein that has an extracellular primary amine exposed to the surface. As a consequence, proteins that lack these sugar moieties could be more accessible for biotin labelling resulting in a higher presence in the subset of identifications using the biotin approach. Combining both approaches should result in a more complete characterisation of the surfaceome. However, adding the biotin approach to the results obtained in the superior hydrazide protocol (CMB 479) only resulted in four additional identifications of surface proteins (81 instead of 77). Furthermore, the biotin approach was unable to detect all eight of the biomarkers suggesting that the biotin protocol, as it has been setup, provides limited added value. A possible explanation for this finding, might be that the number of exposed lysines is low compared to the amount of glycans. However, as it was seen in the optimisation of the hydrazide approach that small alterations in the protocol result in major differences in outcome, it is likely that the biotin approach is subject to similar influences or equally prone to experimental variability. The relative low amount of enrichment for cell surface proteins 2.2 E^{-6} in the biotinylation approach could also be indicative of a high background and a high amount of nonspecific interactors. Which in turn can explain the limited overlap. Therefore, it would be prudent to subject the biotin approach to similar alterations to have experimental evidence on the robustness of the biotin protocol. Further optimisation of

the biotin protocol, like was done for the hydrazide approach, could also improve the enrichment efficiency.

It has to be noted that data analysis was performed using a false discovery rate (FDR) of 1% and with filtration of three valid values in at least one group. These are both very stringent parameters making it plausible that some correct protein identifications were missed. However, the objective was not a complete characterisation the surfaceome, but rather the development of an approach that could identify membrane proteins that are distinctive for a tumour type.

Hormann *et al.* [52] conducted a study which compared sulfo-NHS-SS-biotinylation with glycocapture using aminooxy-biotin. Using sulfo-NHS-SS-biotinylation resulted in identification of 1306 protein IDs of which 49% were annotated as plasma membrane proteins. In contrast, the aminooxy-biotin protocol was able to identify 468 proteins of which 74% had an annotation containing plasma membrane as a characteristic. Mind that the glycocapture using aminooxy-biotin is not completely the same as the hydrazide approach used in this dissertation. However, it is of interest because it also targets glycosylated proteins to enrich for membrane proteins. These findings are more or less comparable to the results of this paper as they confirm the global observation that the biotin approach has the highest number of identifications, but results in a lower enrichment of surface proteins because of a higher amount of unspecific interactors. The biotin approach was also used in a study of Hanke *et al*. [49] that reported on the use of sulfo-NHS labelling *in vivo*. Interestingly they noticed a major difference in efficiency between the *in vivo* results compared to the *in vitro* applications of the protocol. The *in vivo* perfusion experiments resulted in an identification between 1000-1200 proteins per sample with 40-45% of predicted membrane proteins, compared to ~600 identified proteins in the *in vitro* condition (30% suspected membrane proteins). This might suggest that biotinylation could be applicable for *in vivo* screenings in a clinical setting.

The objective of this Master dissertation was to optimise the enrichment of surface membrane proteins as this enables the characterisation of the surface of breast cancer cells. Despite high initial responses, almost all cancers develop resistance to targeted therapies due to the highly dynamic nature of cancer [18]. Tumour heterogeneity is the main reason for therapeutic failure and this is especially the case for targeted therapies as several studies demonstrate that higher tumour heterogeneity predispose patients to inferior outcome [21, 22]. This also means that the use of a single targeted therapy might entail risks in the case of very heterogenic tumours [18]. Furthermore, cancer originating from the same tissue can have a completely different molecular background, while cancers originating from different tissues can be very similar [74]. Effectiveness of targeted therapies depends on the presence or absence of distinct markers and this requires the extensive and serial characterization of tumour composition [75].

This research was only performed in one breast cancer cell line without thorough optimisation of the biotin approach. The next step should be the optimisation of the biotinylation protocol to determine if the parallel use of both the hydrazide and the biotin enrichment strategies is indeed opportune. After this was determined a more extensive experiment with multiple cell lines should be set-up. The use of both cancer cell and non-cancerous cell lines will allow to detect proteins that are only present on the cancer cells providing potential therapeutic targets.

The detection of several membrane associated tyrosine kinase receptors (ERBB2, ERBB3, PTK7, IGF1R, EPHA1, EPHA2 and EPHA7) is another interesting result. These receptors initiate signal transduction cascades which are critical in normal cell biology and can fuel several oncogenic processes by promoting proliferation and metastasis [76]. Ziegler *et al*. [25] reported that the hard to treat TNBC express a broad array of tyrosine kinases. It is promising that the enrichment approaches used are able to incorporate these tyrosine kinase receptors, as they may be one of the keys to treat TNBC.

Despite the predominant rhetoric in proteomic literature that membrane proteins are easily accessible one can wonder to what extent this is actually the case. Many cells are highly glycosylated *in vivo*, which might limit their accessibility for both labels that are used to detect them as well as therapies that target them. Furthermore, the accessibility of a cell in the middle of a tissue also possesses issues which makes the findings of the improved identification of *in vivo* biotinylation events by Hanke *et al*.[49] even more interesting.

Finally, the complexity that is currently seen at the surfaceome level is probably only the tip of the iceberg. Genetic variability that manifests itself at protein termini and results in different proteoforms is now overlooked with the peptide searches in MaxQuant. Moreover, the mind-blowing diversity of carbohydrates should also be taken into account and whereas current protocols only look at quantitative differences at a single time point is cancer a highly dynamic system that interacts with its surrounding. Analysis of this complex data will require extended databases, improved search tools that can automate the searches and a lot of brainpower.

5. Conclusion

Tumour heterogeneity and the highly dynamic nature of cancer are the main reasons for therapeutic failure of targeted therapies. Despite initial strong responses, almost all cancers develop resistance to therapy. The eradication of all tumour cells therefore requires the sequential or combinatorial use of treatments. The challenge of therapy resistance, posed by evolving tumours, is even more important in targeted therapies. This highlights the importance of this research because the effectiveness of targeted therapies is highly dependent on the presence or absence of distinctive molecular patterns. The clinical use of mass spectrometry could prove to be essential for the characterisation of tumours. Surface membrane proteins are of special interest because they are easily accessible and form the main interface along which cells interact with their environment. However, membrane proteins are underrepresented in shotgun proteomics data because of their high dynamic range and hydrophobicity. Therefore this dissertation looked at two possible approaches to enrich for the membrane proteins present on the surface of cancer cells. The hypothesis was that combination of these approaches could provide additional information on the surfaceome of breast cancer cells and of other cancer cells in general. The data suggests that both approaches indeed work complementary. However, the hydrazide approach outcompetes the biotinylation protocol (in its current form) in the enrichment of surface proteins (fold enrichment of 10.4 versus 3.1). Although the combination of both protocols results in higher absolute numbers of surface proteins identifications (81 instead of 77), was the biotin approach only able to identify four out of the eight selected biomarkers. Furthermore, none was uniquely identified using this approach, suggesting that the biotin protocol provides little added value. Although the biotin approach could marginally improve the complete characterisation of the surfaceome, this comes at the cost of higher variability and more false positives. Suggesting that the single use of the hydrazide protocol is more effective to screen for possible drug targets.

6. References

1. Torre LA, Siegel RL, Ward EM, Jemal A (2016) Global Cancer Incidence and Mortality Rates and Trends--An Update. *Cancer Epidemiol Biomarkers Prev* **25**: 16-27

2. Baccelli I, Trumpp A (2012) The evolving concept of cancer and metastasis stem cells. *J Cell Biol* **198**: 281-93

3. Badve S, Nakshatri H (2012) Breast-cancer stem cells-beyond semantics. *Lancet Oncol* **13**: e43-8

4. Scott AM, Allison JP, Wolchok JD (2012) Monoclonal antibodies in cancer therapy. *Cancer immunity* **12**: 14

5. Garcin G, Paul F, Staufenbiel M, Bordat Y, Van der Heyden J, Wilmes S, Cartron G, Apparailly F, De Koker S, Piehler J*, et al.* (2014) High efficiency cell-specific targeting of cytokine activity. *Nature communications* **5**: 3016

6. Ahmed AA, Abedalthagafi M (2016) Cancer diagnostics: The journey from histomorphology to molecular profiling. *Oncotarget* **7**: 58696-58708

7. Maryas J, Faktor J, Dvorakova M, Struharova I, Grell P, Bouchal P (2014) Proteomics in investigation of cancer metastasis: functional and clinical consequences and methodological challenges. *Proteomics* **14**: 426-40

8. Roesli C, Borgia B, Schliemann C, Gunthert M, Wunderli-Allenspach H, Giavazzi R, Neri D (2009) Comparative analysis of the membrane proteome of closely related metastatic and nonmetastatic tumor cells. *Cancer Res* **69**: 5406-14

9. Singh S, Kumar N, Dwiwedi P, Charan J, Kaur R, Sidhu P, Chugh VK (2017) Monoclonal Antibodies: A Review. *Current clinical pharmacology* **13**: 85-99

10. NCI (2019) Targeted therapies fact sheet: https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/targeted-therapies-fact-sheet#q8 Last updated April 29, 2019, Accessed on May 3, 2019. In NIH

11. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* **144**: 646-74

12. Geiger T, Cox J, Mann M (2010) Proteomic changes resulting from gene copy number variations in cancer cells. *PLoS genetics* **6**: e1001090

13. Alfaro JA, Ignatchenko A, Ignatchenko V, Sinha A, Boutros PC, Kislinger TJGM (2017) Detecting protein variants by mass spectrometry: a comprehensive study in cancer cell-lines. **9**: 62

14. Smith LM, Kelleher NL (2013) Proteoform: a single term describing protein complexity. *Nature methods* **10**: 186-7

15. Gawron D, Gevaert K, Van Damme P (2014) The proteome under translational control. *Proteomics* **14**: 2647-62

16. Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH (2008) The functional consequences of alternative promoter use in mammalian genomes. *Trends in genetics : TIG* **24**: 167-77

17. Sendoel A, Dunn JG, Rodriguez EH, Naik S, Gomez NC, Hurwitz B, Levorse J, Dill BD, Schramek D, Molina H*, et al.* (2017) Translation from unconventional 5' start sites drives tumour initiation. *Nature* **541**: 494-499

18. Dagogo-Jack I, Shaw AT (2018) Tumour heterogeneity and resistance to cancer therapies. *Nature reviews Clinical oncology* **15**: 81-94

19. Turner NC, Reis-Filho JS (2012) Genetic heterogeneity and cancer drug resistance. *Lancet Oncol* **13**: e178-85

20. Greaves M, Maley CC (2012) Clonal evolution in cancer. *Nature* **481**: 306-13

21. Harbst K, Lauss M, Cirenajwis H, Isaksson K, Rosengren F, Torngren T, Kvist A, Johansson MC, Vallon-Christersson J, Baldetorp B*, et al.* (2016) Multiregion Whole-Exome Sequencing Uncovers the Genetic Evolution and Mutational Heterogeneity of Early-Stage Metastatic Melanoma. *Cancer Res* **76**: 4765-74

22. Oxnard GR, Thress KS, Alden RS, Lawrance R, Paweletz CP, Cantarini M, Yang JC, Barrett JC, Janne PA (2016) Association Between Plasma Genotyping and Outcomes of Treatment With Osimertinib (AZD9291) in Advanced Non-Small-Cell Lung Cancer. *J Clin Oncol* **34**: 3375-82

23. Bozic I, Reiter JG, Allen B, Antal T, Chatterjee K, Shah P, Moon YS, Yaqubie A, Kelly N, Le DT*, et al.* (2013) Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife* **2**: e00747

24. Zhang J, Adrian FJ, Jahnke W, Cowan-Jacob SW, Li AG, Iacob RE, Sim T, Powers J, Dierks C, Sun F*, et al.* (2010) Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* **463**: 501-6

25. Ziegler YS, Moresco JJ, Tu PG, Yates JR, 3rd, Nardulli AM (2014) Plasma membrane proteomics of human breast cancer cell lines identifies potential targets for breast cancer diagnosis and treatment. *PloS one* **9**: e102341

26. Groenendijk FH, Bernards R (2014) Drug resistance to targeted therapies: deja vu all over again. *Mol Oncol* **8**: 1067-83

27. Redmond KL, Papafili A, Lawler M, Van Schaeybroeck S (2015) Overcoming Resistance to Targeted Therapies in Cancer. *Semin Oncol* **42**: 896-908

28. Cabezon T, Gromova I, Gromov P, Serizawa R, Timmermans Wielenga V, Kroman N, Celis JE, Moreira JM (2013) Proteomic profiling of triple-negative breast carcinomas in combination with a three-tier orthogonal technology approach identifies Mage-A4 as potential therapeutic target in estrogen receptor negative breast cancer. *Molecular & cellular proteomics : MCP* **12**: 381-94

29. Jorgensen JT (2015) Companion diagnostics: the key to personalized medicine. Foreword. *Expert Rev Mol Diagn* **15**: 153-6

30. Kuhlmann L, Cummins E, Samudio I, Kislinger T (2018) Cell-surface proteomics for the identification of novel therapeutic targets in cancer. *Expert Rev Proteomics* **15**: 259-275

31. Elschenbroich S, Kim Y, Medin JA, Kislinger T (2010) Isolation of cell surface proteins for mass spectrometry-based proteomics. *Expert Rev Proteomics* **7**: 141-54

32. Vuckovic D, Dagley LF, Purcell AW, Emili A (2013) Membrane proteomics by high performance liquid chromatography-tandem mass spectrometry: Analytical approaches and challenges. *Proteomics* **13**: 404-23

33. Tan S, Tan HT, Chung MC (2008) Membrane proteins and membrane proteomics. *Proteomics* **8**: 3924-32

34. Bausch-Fluck D, Hofmann A, Bock T, Frei AP, Cerciello F, Jacobs A, Moest H, Omasits U, Gundry RL, Yoon C*, et al.* (2015) A mass spectrometric-derived cell surface protein atlas. *PloS one* **10**: e0121314

35. Macher BA, Yen TY (2007) Proteins at membrane surfaces-a review of approaches. *Mol Biosyst* **3**: 705-13

36. Aebersold R, Mann M (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* **537**: 347-55

37. Altelaar AF, Munoz J, Heck AJ (2013) Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* **14**: 35-48

38. Horth P, Miller CA, Preckel T, Wenz C (2006) Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Molecular & cellular proteomics : MCP* **5**: 1968-74

39. Figeys D, Ducret A, Aebersold R (1997) Identification of proteins by capillary electrophoresis-tandem mass spectrometry. Evaluation of an on-line solid-phase extraction device. *J Chromatogr A* **763**: 295-306

40. Manadas B, Mendes VM, English J, Dunn MJ (2010) Peptide fractionation in proteomics approaches. *Expert Rev Proteomics* **7**: 655-63

41. Banerjee S, Mazumdar S (2012) Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem* **2012**: 282574

42. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, NY)* **246**: 64-71

43. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods* **4**: 709-12

44. Vaudel M, Sickmann A, Martens L (2010) Peptide and protein quantification: a map of the minefield. *Proteomics* **10**: 650-70

45. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**: 1017-31

46. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics : MCP* **13**: 2513-26

47. Zhu W, Smith JW, Huang CM (2010) Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010**: 840518

48. Elia G (2008) Biotinylation reagents for the study of cell surface proteins. *Proteomics* **8**: 4012-24

49. Hanke SA, Kerner A, Nadler WM, Trumpp A, Zhang Y, Rosli CP (2016) The pivotal role of reactivity in the design of novel biotinylation reagents for the chemical-proteomics-based identification of vascular accessible biomarkers. *J Proteomics* **141**: 57-66

50. Go YM, Chandler JD, Jones DP (2015) The cysteine proteome. *Free Radic Biol Med* **84**: 227-245

51. Noguera-Salva MA, Guardiola-Serrano F, Martin ML, Marcilla-Etxenike A, Bergo MO, Busquets X, Escriba PV (2017) Role of the C-terminal basic amino acids and the lipid anchor of the Ggamma2 protein in membrane interactions and cell localization. *Biochim Biophys Acta Biomembr* **1859**: 1536-1547

52. Hormann K, Stukalov A, Muller AC, Heinz LX, Superti-Furga G, Colinge J, Bennett KL (2016) A Surface Biotinylation Strategy for Reproducible Plasma Membrane Protein Purification and Tracking of Genetic and Drug-Induced Alterations. *Journal of proteome research* **15**: 647-58

53. Gahmberg CG, Tolvanen M (1996) Why mammalian cell surface proteins are glycoproteins. *Trends Biochem Sci* **21**: 308-11

54. Stanley P (2011) Golgi glycosylation. *Cold Spring Harb Perspect Biol* **3**: a005199

55. Moremen KW, Tiemeyer M, Nairn AV (2012) Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* **13**: 448-62

56. Nairn AV, York WS, Harris K, Hall EM, Pierce JM, Moremen KW (2008) Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* **283**: 17298-313

57. Wollscheid B, Bausch-Fluck D, Henderson C, O'Brien R, Bibel M, Schiess R, Aebersold R, Watts JD (2009) Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nature biotechnology* **27**: 378-86

58. Zeng Y, Ramya TN, Dirksen A, Dawson PE, Paulson JC (2009) High-efficiency labeling of sialylated glycoproteins on living cells. *Nature methods* **6**: 207-9

59. Bausch-Fluck D, Hofmann A, Wollscheid B (2012) Cell surface capturing technologies for the surfaceome discovery of hepatocytes. *Methods in molecular biology (Clifton, NJ)* **909**: 1-16

60. Hofmann A, Bausch-Fluck D, Wollscheid B (2013) CSC technology: selective labeling of glycoproteins by mild oxidation to phenotype cells. *Methods in molecular biology (Clifton, NJ)* **951**: 33-43

61. McDonald CA, Yang JY, Marathe V, Yen TY, Macher BA (2009) Combining results from lectin affinity chromatography and glycocapture approaches substantially improves the coverage of the glycoproteome. *Molecular & cellular proteomics : MCP* **8**: 287-301

62. Soule HD, Vazguez J, Long A, Albert S, Brennan M (1973) A human cell line from a pleural effusion derived from a breast carcinoma. *J Natl Cancer Inst* **51**: 1409-16

63. Scheurer SB, Roesli C, Neri D, Elia G (2005) A comparison of different biotinylation reagents, tryptic digestion procedures, and mass spectrometric techniques for 2-D peptide mapping of membrane proteins. *Proteomics* **5**: 3035-9

64. Karhemo PR, Ravela S, Laakso M, Ritamo I, Tatti O, Makinen S, Goodison S, Stenman UH, Holtta E, Hautaniemi S*, et al.* (2012) An optimized isolation of biotinylated cell surface proteins reveals novel players in cancer metastasis. *J Proteomics* **77**: 87-100

65. Zhang H, Li XJ, Martin DB, Aebersold R (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nature biotechnology* **21**: 660-6

66. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**: 976-89

67. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**: 731-40

68. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10**: 1794-805

69. Arnesen T, Van Damme P, Polevoda B, Helsens K, Evjenth R, Colaert N, Varhaug JE, Vandekerckhove J, Lillehaug JR, Sherman F*, et al.* (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 8157-62

70. Varland S, Osberg C, Arnesen T (2015) N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics* **15**: 2385-401

71. Tyanova S, Cox J (2018) Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Methods in molecular biology (Clifton, NJ)* **1711**: 133-148

72. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**: 44-57

73. Criscitiello C (2012) Tumor-associated antigens in breast cancer. *Breast Care (Basel)* **7**: 262-6

74. Lili LN, Matyunina LV, Walker LD, Daneker GW, McDonald JF (2014) Evidence for the importance of personalized molecular profiling in pancreatic cancer. *Pancreas* **43**: 198-211

75. Sadelain M, Riviere I, Riddell S (2017) Therapeutic T cell engineering. *Nature* **545**: 423-431

76. Meyer AS, Miller MA, Gertler FB, Lauffenburger DA (2013) The receptor AXL diversifies EGFR signaling and limits the response to EGFR-targeted inhibitors in triple-negative breast cancer cells. *Sci Signal* **6**: ra66

7. Supplementary information

7.1 Biotinylation protocol

Day 1: Biotinylation + affinity purification

Take biotin label from freezer (needs to be at RT for 30')[2].

Lysis buffer (= modified Ripa for the actual cell lysis) (pH 7.4):

Need 7 mL of lysis buffer (for 6 samples)

$$\frac{V_{Buffer} \times [\,]_{Buffer}}{[\,]_{Stock\ solution}} = \boldsymbol{X\ mL}$$

- SDS$_{[0,1\%]} \rightarrow \frac{7\ \text{mL} \times 0{,}1\%}{10\%} = 70\ \mu L$
- EDTA$_{[1\text{mM}]} \rightarrow \frac{7\ \text{mL} \times 1\ \text{mM}}{100\ mM} = 70\ \mu L$
- DOC$_{[0,5\%]} \rightarrow \frac{7\ \text{mL} \times 0{,}5\%}{10\%} = 350\ \mu L$
- Phosphate Buffer$_{[50\ \text{mM}]} \rightarrow \frac{7\ \text{mL} \times 50\ \text{mM}}{1\ \text{M}} = 350\ \mu L$
- NaCl$_{[150\ \text{mM}]} \rightarrow \frac{7\ \text{mL} \times 150\ \text{mM}}{3\ \text{M}} = 350\ \mu L$
- IGEPAL$_{[1\%]} \rightarrow \frac{7\ \text{mL} \times 1\%}{10\%} = 700\ \mu L$
- Iodoacetamide$_{[30\ \text{mM}]} \rightarrow \frac{7\ \text{mL} \times 0{,}03\text{M}}{1\text{M}} = 210\ \mu L$
- protease inhibitor$_{[50\text{X}]} \rightarrow \frac{7\ \text{mL} \times 1\text{X}}{50\ \text{X}} = 140\ \mu L$

= 2,240 mL => add 4,760 mL Milli-Q (= ultrapure H_2O) to reach 7 mL

Lysis buffer without IAA and PI (for the washing of the beads):

Need 5 mL (4x 1mL for washes + 1 spare)

- SDS$_{[0,1\%]} \rightarrow \frac{5\ \text{mL} \times 0{,}1\%}{10\%} = 50\ \mu L$
- EDTA$_{[1\text{mM}]} \rightarrow \frac{5\ \text{mL} \times 1\ \text{mM}}{100\ mM} = 50\ \mu L$
- DOC$_{[0,5\%]} \rightarrow \frac{5\ \text{mL} \times 0{,}5\%}{10\%} = 250\ \mu L$
- Phosphate Buffer$_{[50\ \text{mM}]} \rightarrow \frac{5\ \text{mL} \times 50\ \text{mM}}{1\ \text{M}} = 250\ \mu L$
- NaCl$_{[150\ \text{mM}]} \rightarrow \frac{5\ \text{mL} \times 150\ \text{mM}}{3000\ \text{mM}} = 250\ \mu L$
- IGEPAL$_{[1\%]} \rightarrow \frac{5\ \text{mL} \times 1\%}{10\%} = 500\ \mu L$

= 1,350 mL => add 3,650 mL Milli-Q (= ultrapure H_2O) to reach 5 mL

[2] Otherwise condensation would deteriorate the remaining biotin label for the following experiments.

Iodoacetamide (= IAA):

Need 10 mL DPBS/plate$_{(n)}$ containing IAA$_{[30mM]}$ for the incubation

$$\frac{10\ mL\ \times\ 0{,}03\text{M}}{1\text{M}} = 0{,}3\ mL \Rightarrow \frac{300\ \mu\text{L of IAA (1M)}}{plate}$$

6 plates => 1,8 mL of IAA (1M)

Need 7 mL of lysis buffer containing IAA$_{[30mM]}$

$$\frac{7\ mL\ \text{x}\ 0{,}03\text{M}}{1\text{M}} = 0{,}210\ mL$$

=> need 2,01 mL of IAA (1M) stock solution

MW$_{IAA}$ = 184,96 mg/mL

$$184{,}96\frac{mg}{mL} \times 2{,}01\ mL = 371{,}77\ mg \Rightarrow weigh\ at\ least\ 372\ mg$$

$$\frac{(X)\ mg}{184{,}96\ \frac{mg}{mL}} = (V)\ mL$$

Cover with aluminium foil because it is light sensitive.

Glycine:

Need 6 x 7 mL DPBS containing Glycine$_{[20\ mM]}$ to block the non-reacted biotin.

$$\frac{42\ mL\ \text{x}\ 20\ \text{mM}}{1000\ \text{mM}} = 0{,}840\ mL \Rightarrow dissolve\ 840\ \mu L\ of\ Glycine_{[1\ M]}\ in\ 41{,}160\ mL\ DPBS$$

Biotinylation of membrane proteins on adherent cells	**Negative control (C^-)**
Wash: (3x) with 7mL DPBS[3] on ice (4 °C)	
Prevent biotin reduction[4]: Incubate cells with 10 mL $IAA_{[30mM]}$ in DPBS for 5 min @ 4 °C covered by aluminium foil.	
Wash: (1x) with 7mL DPBS on ice (4 °C)	
Prepare biotin stock solution$_{[100mM]}$	
Incubate cells for 30 min with 7mL Sulfo-NHS[5]-SS-Biotin $_{[500\mu M]}$ @ 4° C => 35µL of Biotin$_{[0,1\ M]}$	Incubate with 7 mL PBS for 30 min @ 4 °C
Remove biotinylating reagent/PBS: Wash 2x with DPBS 2000 x g for 2min @ 4 °C Use suction pump to remove fluid layer	
Blocking of non-reacted biotin: Block non-reacted biotin with glycine$_{[20\ mM]}$ for 15 min @ 4 °C.	
=> 140µL of Glycine$_{[1M]}$ In 7mL DPBS	=> 140µL of Glycine$_{[1M]}$ In 7mL DPBS
Remove glycine: Wash 1x with DPBS on ice	
Cell lysis: Add 1mL lysis buffer scrape into 2mL Eppendorf and sonicate Place 30 min on rotor @ 4 °C. *Prepare Bradford solution in the meanwhile and wash the beads (1x LB)*	
Centrifugate to clear the lysate: (= 16000 x g) for 10 min @ 4 °C use cleared lysate (= INPUT). IN- (B | C^-)	
Bradford protein assay Calculate the volume we need from the INPUT samples (IN-) to add an equal amount to the beads. Leave a volume (90 µL) for WB and place these in the freezer or start the WB protocol.	
Affinity purification: Add equal amounts of protein from each sample to the Neutravidin beads (40 µL). AP- (B_x| C_x) Incubate 3h @ 4 °C on rotor.	

[3] Use DPBS with ions. The formulation without is more expensive and is used when a chelator is used later on to detach the cells. Cells require Mg^{2+} or Ca^{2+} for adherence.
[4] Incubate with iodoacetamide (IAA) to prevent the reduction of the disulphide bridge in the biotin label during cell lysis by carbamylation of free -SH in cysteines in the reducing intracellular environment.
[5] N-hydroxysuccinimide (NHS)-biotin reacts with primary amines in physiologic to slightly alkaline conditions (pH 7.2-9) by releasing NHS and couple the biotin with a stable amide bond.

- Prepare 25 mL of lysis buffer without Iodoacetamide and protease inhibitor, but with higher [NaCl] (for stringent washing the Neutravidin beads to remove interactors)
  - IGEPAL CA-630 (Nonidet P40)$_{[1\%]}$
  - DOC$_{[0,5\%]}$
  - SDS$_{[0,1\%]}$
  - EDTA$_{[1mM]}$
  - **NaCl$_{[300\ mM]}$**
  - Phosphate Buffer$_{[50\ mM]}$ (pH 7.4)

- Prepare 15 mL of TEAB$_{[100\ mM]}$ (= Tetraethylammonium bicarbonate)

- Prepare 1 mL of DTT$_{[15mM]}$ (= Dithiothreitol)

Removing unbound proteins: Centrifugate for 2 min with 2000 x g @ 4 °C Remove and freeze the supernatants AP-SN (B_n| C^-_n) This can be used to check if the AP captured every biotinylated protein (=measure for capturing efficiency)
Washing: 4x with lysis buffer containing NaCl$_{[300\ mM]}$ (3min on rotor @ 4 °C between washes) 2x TEAB$_{[100\ mM]}$ (pH = 8) (to go to a detergent free buffer) *Use each time 1 mL of the appropriate solution and centrifugate 2 min with 2000g @ 4 °C*
Eluting captured proteins: Elute (2x) with 50µL of DTT$_{[15\ mM]}$ dissolved in TEAB$_{[100mM]}$ (pH = 8,0)[6] @ 55° C for 15 min Pool both elutions in **Low protein binding Eppendorf's** (TRP- B_x| C_x) and add 300µL TEAB$_{[10mM]}$[7] Freeze the beads[8] (AP- B_x| C_x).
Trypsin overnight: Add 10 µL of Trypsin [0.2 µg/µL] (1:100 / w:w) Overnight in heat block @ 37 °C / 750 rpm

[6] Trypsin and endoproteinase Lys-C work better at slightly basic pH.
[7] To dilute the DTT 4x (higher concentrations interfere with the enzyme digest)
[8] These are still bound to the glycosylated peptide. This can be removed by PNGase F (= enzyme that removes all N-linked oligosaccharides from glycoproteins) and can be used for further analysis.

Biotin stock solution[100 mM]:

Need 3 x7 mL of PBS with Biotin[500µM]

$$\frac{21\ mL \text{ x } 0{,}5\ \text{mM}}{100\ \text{mM}} = 0{,}105\ mL$$

=> need 105 µL of Biotin[100mM] stock solution

$$MW_{Biotin} = \frac{607\ g}{mol} => 1M = \frac{607\ g}{L} = \frac{607\ mg}{mL} => 100\ mM = 60{,}7\frac{mg}{mL}$$

$$0{,}105\ mL \times \frac{60{,}7\ mg}{mL} = 6{,}37\ mg => weigh\ at\ least\ 6{,}4\ mg\ Biotin$$

Wash buffer (= Lysis buffer without IAA and PI and higher [NaCl]):

Need 25 mL to remove the interactors from the captured proteins

- SDS (0,1%) → $\frac{25\ \text{mL} \times 0{,}1\%}{10\%} = 250\ \mu L$
- EDTA (1mM) → $\frac{25\ \text{mL} \times 1\ \text{mM}}{100\ mM} = 250\ \mu L$
- DOC (0,5%) → $\frac{25\ \text{mL} \times 0{,}5\%}{10\%} = 1{,}250\ mL$
- Phosphate Buffer (50 mM) → $\frac{25\ \text{mL} \times 50\ \text{mM}}{1000\ \text{M}} = 1{,}250\ mL$
- NaCl (300 mM) → $\frac{25\ \text{mL} \times 300\ \text{mM}}{3000\ \text{M}} = 2{,}5\ mL$
- IGEPAL (1%) → $\frac{25\ \text{mL} \times 1\%}{10\%} = 2{,}500\ mL$

= 8 mL => add 17 mL Milli-Q (= ultrapure H_2O) to reach 25 mL

Tetraethylammonium bicarbonate (TEAB):

Need 15mL of TEAB[100 mM]:

Commercial stock solution of 1M => 1,5 mL of TEAB[1M] + 13,5 mL milli-Q.

Dithiothreitol (DTT):

Need 6x 100 µL of DTT[15 mM] dissolved in TEAB[100 mM] to elute the captured proteins from the beads.

$$\frac{600\ \mu L \times 15\ mM}{1000\ M} = 9\ \mu L => need\ at\ least\ 9\ \mu L\ DTT_{[1\ M]}\ stock\ solution$$

$$MW_{DTT} = \frac{154{,}25\ g}{mol} => 1M = \frac{154{,}25\ g}{L} = \frac{154{,}25\ mg}{mL}$$

$$0{,}090\ mL \times \frac{154{,}25\ mg}{mL} = 13{,}88\ mg => weigh\ at\ least\ 14\ mg\ DTT$$

$$\frac{(X)\ mg}{154{,}25\ \frac{mg}{mL}} = (V) => \boldsymbol{Dissolve\ (X)\ in\ (V)\ mL\ to\ have\ 1M\ stock\ solution}$$

FACULTY OF MEDICINE AND HEALTH SCIENCES

Day 2: MS-sample preparation

- Prepare 10 mL **pre-wash buffer**
  - 80% Acetonitrile (ACN) → 8 mL
  - 20% Milli-Q (H_2O) → 2 mL
  - 0,1% TFA → $TFA_{[20\%]}$ stock solution → $\frac{10\ mL \times 0{,}1\%}{20\ \%} = 50\ \mu L$

- Prepare 20 mL of **Solvent A** (= Wash buffer)
  - 100% Milli-Q (H_2O) → 20 mL
  - 0,1% TFA → $TFA_{[20\%]}$ stock solution → $\frac{20\ mL \times 0{,}1\%}{20\ \%} = 100\ \mu L$

- Prepare 1 mL of **Solvent B** (= Elution buffer)
  - 60% Acetonitrile (ACN) → 600 µL
  - 40% Milli-Q (H_2O) → 400 µL
  - 0,1% TFA → $TFA_{[20\%]}$ stock solution → $\frac{1\ mL \times 0{,}1\%}{20\ \%} = 5\ \mu L$

Inactivate trypsin:
Heat for 10 min in heat block @ 75 °C / 300 rpm
Use ice to bring the sample back to RT and use a quick spin to collect the condensation

Acidify the digested/peptide sample to 1% TFA:
=> Add 5,5 µL (= 1/20 of sample volume) of 20% TFA to each sample

Check the pH by spotting 0,5 µL of the peptide sample on a pH strip
(pH should be lower than pH = 3)[9]

Purify samples:
centrifugate acidified samples for 10 min with 1780 x g @ RT
(to precipitate contaminants in a pellet)
Only load the purified samples on C18 material

Use C18 Pipette-based solid phase extraction (SPE)
(Ziptip /Agilent Bond Elut OMIX)
to remove the buffer and salts from the peptide mixture:

Set pipette to volume smaller than the samples to avoid aspiration of air and only push till the first stop of the pipette (only push through for the second elution = the very last time !)

1. **Wet tip with pre-wash buffer (80% ACN / 20% Milli-Q / 0,1% TFA)**
   Pipette 5x 150 µL of pre-wash buffer and discard buffer between each time.
2. **Wash tips 5x with Solvent A (= wash buffer) (100% ACN / 0,1% TFA)**
   Pipette 5x 150 µL of wash buffer
   discard buffer between each time.
3. **Load acidified and cleared digest**
   Pipette 20-30 times up and down with the samples
4. **Wash tips 3x with Solvent A (= wash buffer) (100% ACN / 0,1% TFA)**
   Pipette 3x 150 µL of wash buffer and discard buffer between each time.
5. **Elute peptides with Solvent B (= elution buffer) (60% ACN / 40% Milli-Q / 0.1% TFA)**
   Sequential use 2x 75 µL of elution buffer in Low protein binding Eppendorf's
   Pipette 10x up and down in first Eppendorf and elute in the labelled MS vial repeat this for the second Eppendorf.

Transfer the samples to MS vials[10]:
Transfer the samples to labelled MS vials
and vacuum dry the purified peptides to dryness in a Speed Vac.

Re-suspend in loading buffer:
Re-suspend dried sample in 20µL loading buffer
(0.1% TFA, 2% ACN and 98% Milli-Q (v/v))
Store in -20 °C (MS ready freezer) till analysis.

[9] This is done to make sure that there is nothing left to precipitate at a low pH prior to loading it on C18 material. Otherwise this would clog the C18 material when you add 0.1% TFA. This already gives a pH = 2.
[10] Maximum volume of MS vial is 350 µL (use SpeedVac to concentrate the sample if volume is bigger)

7.2 Hydrazide protocol

Day 1: oxidation of membrane proteins

Lysis buffer:

Prepare 7 mL of lysis buffer (for 6 samples) $\frac{V_{Buffer} \times [\,]_{Buffer}}{[\,]_{Stock\ solution}} = \boldsymbol{X\ mL}$

- IGEPAL (1%) → $\frac{7\ \text{mL} \times 1\%}{10\%} = 0{,}7\ mL = 700 \mu L$
- SDS (0,1%) → $\frac{7\ \text{mL} \times 0{,}1\%}{10\%} = 0{,}07\ mL = \ 70\ \mu L$
- NaAcetate buffer 50 mM (pH = 7) → $\frac{7\ \text{mL} \times 50\ \text{mM}}{100\ mM} = 3{,}5\ mL$
- NaCl (150 mM) → $The\ NaAcetate\ buffere\ allready\ contains\ 150mM\ NaCl$
  $\rightarrow so\ just\ add\ for\ the\ remaining\ 3{,}5\ mL\ \rightarrow \frac{3{,}5\ \text{mL}\ \times\ 150\ \text{mM}}{3000\ \text{mM (stock solution)}} = 175\ \mu L$
- protease inhibitor (1X) → $\frac{7\ \text{mL} \times 1\text{X}}{50\ \text{X}} = 140\ \mu L$

Total of 4,735 mL -> add 2,265 mL Milli-Q (= ultrapure H_2O) to reach 7 mL.

Sodium periodate ($NaIO_4$):

Need 3 x 10 mL $NaIO_{4\,[2\ mM]}$

$$\frac{30\ mL \times 2\ mM}{1000\ mM} = 0{,}060\ mL\ \Rightarrow\ need\ at\ least\ 60\ \mu L\ NaIO_{4\,[1\ M]}\ stock\ solution$$

$$MW_{NaIO_4} = \frac{214\ g}{mol} \Rightarrow 1M = \frac{214\ g}{L} = \frac{214\ mg}{mL}$$

$$0{,}06\ mL\ \times \frac{214\ mg}{mL} = 12{,}84\ mg \Rightarrow weigh\ at\ least\ 13\ mg\ NaIO_4$$

$$\frac{(X) mg}{214\ \frac{mg}{mL}} = (V) \Rightarrow \boldsymbol{Dissolve\ (X)\ in\ (V)\ mL\ PBS\ (pH\ \approx 6{,}3)}$$

Cover with aluminium foil because it is light sensitive.

Stock solution of $NaCl_{[6\ M]}$[11]

$$MW_{NaCl} = \frac{58{,}4\ g}{mol} \Rightarrow 1M = \frac{58{,}4\ g}{L} = \frac{58{,}4\ mg}{mL} \Rightarrow 6M = \frac{350{,}4\ mg}{mL\ of\ stock\ solution}$$

After clearing the lysate we want to reach $NaCl_{[1{,}5\ M]}$.

$$(1\ mL\ of\ scraped\ cells + \boldsymbol{x}\ mL) \times 1{,}2M^{12} = \boldsymbol{x}\ mL\ \times 6M$$

$$\Leftrightarrow 1{,}2\ mL.M\ + 1{,}2\ \boldsymbol{x}\ mL.M\ = 6\boldsymbol{x}\ mL.M$$

$$\Leftrightarrow 1{,}2\ mL.M = 6\boldsymbol{x}\ mL.M - 1{,}2\ \boldsymbol{x}\ mL.M$$

$$\Leftrightarrow 1{,}2\ mL.M = 4{,}8\ \boldsymbol{x}\ mL.M$$

$$\Leftrightarrow \frac{1{,}2\ mL.M}{4{,}8\ mL.M} = \boldsymbol{x} = 0{,}250\ mL$$

[11] 6M is the maximal solubility of NaCl
[12] Already contains 300mM

Capturing glycosylated membrane proteins	**Negative control (C^-)**
Wash: (3x) with 7mL ice-cold DPBS (pH = 6-6,5)[13]	
Oxidation[14]: Incubate cells for 10 min with $NaIO_4$ [2 mM] in 10 mL DPBS (pH = 6-6,5) on ice in the dark	**Oxidation:** ice-cold DPBS (pH = 6-6,5)
Washing: Remove $NaIO_4$ reagent and wash cells (3x) with 7ml ice-cold DPBS (pH = 7).	
Cell lysis: Add 1mL of lysis buffer (pH = 7) and scrape into 2mL Eppendorf (easier for sonication) 30 min on rotor @ 4 °C.	
Start washing (6) x 100µL hydrazide beads slurry - Wash (3x) with 10 mL of H_2O (let the beads precipitate by gravity) - Re-suspend beads in Sodium Acetate buffer [50 mM] (Add 1,5 mL of Sodium Acetate buffer [100 mM] (pH = 5,5) + 1,5 mL H_2O)	
prepare 20 mL Bradford solution (1/5)	
Centrifugation: 16000 x g for 10 min @ 4 °C Put cleared lysate (INPUT) in new Eppendorf (IN – Ox_n | C^-_n.)	
Increase salt concentration to $NaCl_{[1,5\ M]}$[15]: Add 250 µL of $NaCl_{[6M]}$ to each sample	
Bradford protein assay: Measure protein concentration (don't forget to increase salt concentration in Bradford blanc)	
Acidify the samples: Should be around pH = 5,5 (need ~2,0 µL $HCl_{[3,7\%]}$)	
Divide beads: Mix washed beads well and divide in # samples (AP – Ox_n | C^-_n) (use 1,5 mL Eppendorf's because this is easier to separate the beads) Add an equal amount of proteins from the INPUT samples to the hydrazide beads.	
Capture of oxidized glycans[16]: Overnight affinity purification @ 4 °C on a rotor.	

[13] pH needs to be slightly acidic for higher efficiency of the oxidation reaction with $NaIO_4$
[14] Partial oxidation of vicinal diols of glycans to aldehydes
[15] Increasing the salt concentration from $NaCl_{[300\ mM]}$ to $NaCl_{[1500\ mM]}$ during coupling prevents nonspecific binding. (Also pay attention to leave the samples 5 min on the rotor during the washes.)
[16] Formation of a stable hydrazone linkage between the aldehydes and the hydrazide functional group of the agarose beads.

Day 2: elution of captured proteins from beads

- Prepare 10 mL of $TEAB_{[100\ mM]}$ (= Tetraethylammonium bicarbonate)
- Prepare 25 mL of **Urea buffer**
  - Urea [8 M]
  - TEAB [0,1 M]
  - 0,1% SDS

- Prepare 90µL stock solution **Dithiothreitol (DTT)$_{[1M]}$**
- Prepare fresh 180 µL **Iodoacetamide (IAA)$_{[1M]}$**

Removing unbound proteins: Centrifugate for 2 min with 2000 x g @ 4 °C Remove and freeze the supernatants AP-SN (Ox_n | C^-_n)
Denaturation of captured proteins: Re-suspend beads in 1 mL of Urea buffer (Urea [8 M] / TEAB [0.1 M] / 0.1%SDS)
Reduction of disulphide bridges: Rotate for 1h @ room temperature with DTT [15 mM] (or TCEP [10 mM])
Alkylation of free thiols: Rotate for 30 min @ room temperature with Iodoacetamide [30 mM][17] covered in aluminium foil.
Wash to remove interactors[18]: 3x with 1mL (Urea [8 M] / TEAB [0.1 M] / 0.1 % SDS) Centrifugate 2 min with 2000 x g @ RT (place 5' on rotor for each wash)
Wash to remove Urea[19]: 1 mL TEAB [100mM] Centrifugate 2 min with 2000 x g @ RT
Endoproteinase Lys-C[20]: Add 3 µL Endolysin C [1µg/µL] (1:100 / w:w)[21] in 200 µL TEAB [100mM] for 3 hours in heat block @ 37 °C / 750 rpm
Trypsin overnight: Add 10 µL of Trypsin [0.2 µg/µL] (1:100 / w:w) Overnight in heat block @ 37 °C / 750 rpm

[17] If you use TCEP you only need IAA [12 mM]
(= double the amount of DTT used because DTT contains a disulphide bridge)
[18] The denaturation of the captured proteins with Urea buffer and the reduction of the disulphide bridges followed by the alkylation of the free thiols aims to remove the interactors that are bound to the captured proteins, but are not bound covalently to **hydrazide** beads.
[19] Endoproteinase Lys-C and Trypsin only work below a concentration of Urea [4 M] and [2 M] respectively.
[20] Endoproteinase Lys-C only cleaves after lysine. It is used here to facilitates the work of Trypsin
[21] This means 1 µg of enzyme should be added for 100 µg of proteins

Tetraethylammonium bicarbonate (TEAB):

Need 10mL of TEAB$_{[100\ mM]}$:

Commercial stock solution of 1M => 1 mL of 1 M stock solution + 9 mL milli-Q.

Urea buffer:

Need 25 mL of Urea buffer (Urea $_{[8M]}$ / TEAB $_{[0,1M]}$ / SDS $_{[0,1\%]}$)

- $MW_{Urea} = \frac{60{,}06\ g}{mol} \Rightarrow 1M = \frac{60{,}06\ g}{L} = \frac{60{,}06\ mg}{mL}$

$$\Rightarrow 8M = \frac{480{,}48\ mg}{mL}$$

=> 25 mL contains 12,01 g Urea

$$\frac{\boldsymbol{V_{Buffer} \times [\,]_{Buffer}}}{\boldsymbol{[\,]_{Stock\ solution}}} = \boldsymbol{X\ mL}$$

- SDS $_{[0,1\%]} \rightarrow \frac{25\ \text{mL} \times 0{,}1\%}{10\%} = 250\ \mu L\ of\ SDS_{[10\%]}\ stock\ solution$
- TEAB$_{[100\ mM]} \rightarrow \frac{25\ \text{mL} \times 100\text{mM}}{1000\ \text{mM}} = 2{,}5\ mL\ of\ TEAB_{[1M]}\ commercial\ stock\ solution$

First dissolve in 10 mL of Milli-Q and then add till 25mL

Iodoacetamide:

Need 6 mL IAA$_{[30\ mM]}$

$$\frac{6\ mL \times 30\ mM}{1M} = 0{,}180\ mL \Rightarrow need\ at\ least\ 180\ \mu L\ IAA_{[1\ M]}\ stock\ solution$$

$$MW_{IAA} = \frac{184{,}96\ g}{mol} \Rightarrow 1M = \frac{184{,}96\ g}{L} = \frac{184{,}96\ mg}{mL}$$

$$0{,}180\ mL \times \frac{184{,}96\ mg}{mL} = 33{,}3\ mg \Rightarrow weigh\ at\ least\ 33{,}4\ mg\ IAA$$

$$\frac{(X) mg}{184{,}96\ \frac{mg}{mL}} = (V) \Rightarrow \boldsymbol{Dissolve\ (X)\ in\ (V)\ mL\ to\ have\ 1M\ stock\ solution}$$

Dithiothreitol (DTT):

Need enough stock solution of DTT$_{[1M]}$ to reach 15 mM in 6mL of re-suspended beads.

$$\frac{6\ mL \times 15\ mM}{1M} = 0{,}090\ mL \Rightarrow need\ at\ least\ 90\ \mu L\ DTT_{[1\ M]}\ stock\ solution$$

$$MW_{DTT} = \frac{154{,}25\ g}{mol} \Rightarrow 1M = \frac{154{,}25\ g}{L} = \frac{154{,}25\ mg}{mL}$$

$$0{,}090\ mL \times \frac{154{,}25\ mg}{mL} = 13{,}88\ mg \Rightarrow weigh\ at\ least\ 14\ mg\ DTT$$

$$\frac{(X)\ mg}{154{,}25\ \frac{mg}{mL}} = (V) \Rightarrow \boldsymbol{Dissolve\ (X)\ in\ (V)\ mL\ to\ have\ 1M\ stock\ solution}$$

Day 3: MS-sample preparation

- Prepare 10 mL **pre-wash buffer**
  - 80% Acetonitrile (ACN) → 8 mL
  - 20% Milli-Q (H_2O) → 2 mL
  - 0,1% TFA → $TFA_{[20\%]}$ stock solution → $\frac{10\ \text{mL} \times 0{,}1\%}{20\ \%} = 50\ \mu L$

- Prepare 20 mL of **Solvent A** (= Wash buffer)
  - 100% Milli-Q (H_2O) → 20 mL
  - 0,1% TFA → $TFA_{[20\%]}$ stock solution → $\frac{20\ \text{mL} \times 0{,}1\%}{20\ \%} = 100\ \mu L$

- Prepare 1 mL of **Solvent B** (= Elution buffer)
  - 60% Acetonitrile (ACN) → 600 µL
  - 40% Milli-Q (H_2O) → 400 µL
  - 0,1% TFA → $TFA_{[20\%]}$ stock solution → $\frac{1\ \text{mL} \times 0{,}1\%}{20\ \%} = 5\ \mu L$

Collect tryptic peptides:
Centrifugate for 2' with 2000 x g @ RT
Collect the tryptic peptides in **Low protein binding Eppendorf's** (TRP- Ox_n| C^-_n)
and freeze the beads[22] (AP- Ox_n| C^-_n).

Inactivate trypsin:
Heat for 10 min in heat block @ 75 °C / 300 rpm
Use ice to bring the sample back to RT and use a quick spin to collect the condensation

Acidify the digested/peptide sample to 1% TFA[23]:
(= Add 1/20 of sample volume of 20% TFA to each sample)
Check the pH by spotting 0,5 µL of the peptide sample on a pH strip

Purify samples:
centrifugate 15 min with 1780 x g @ RT to precipitate contaminants in a pellet
Only load the purified samples on C18 material

Use C18 Pipette-based solid phase extraction (SPE)
(Ziptip /Agilent Bond Elut OMIX)
to remove the buffer and salts from the peptide mixture:

Set pipette to volume smaller than the samples to avoid aspiration of air and only push till the first stop of the pipette (only push through for the second elution = the very last time!)

1. **Wet tip with pre-wash buffer (80% ACN / 20% Milli-Q / 0,1% TFA)**
   Pipette 5x 150 µL of pre-wash buffer and discard buffer between each time.
2. **Wash tips 5x with Solvent A (= wash buffer) (100% Milli-Q / 0,1% TFA)**
   Pipette 5x 150 µL of wash buffer and discard buffer between each time.
3. **Load acidified and cleared digest**
   Pipette 20-30 times up and down with the samples
4. **Wash tips 3x with Solvent A (= wash buffer) (100% Milli-Q / 0,1% TFA)**
   Pipette 3x 150 µL of wash buffer and discard buffer between each time.
5. **Elute peptides with Solvent B (= elution buffer)**
   (60% ACN / 40% Milli-Q / 0.1% TFA)
   Sequential use 2x 75 µL of elution buffer in Low protein binding Eppendorf's
   Pipette 10x up and down in first Eppendorf and elute in the labelled MS vial repeat this for the second Eppendorf.

Transfer the samples to MS vials[24]:
Transfer the samples to labelled MS vials
and vacuum dry the purified peptides to dryness in a SpeedVac.

Re-suspend in loading buffer:
Re-suspend dried sample in 20µL loading buffer
(0.1% TFA, 2% ACN and 98% Milli-Q (v/v))
Store in -20 °C (MS ready freezer) till analysis.

[22] These are still bound to the glycosylated peptide. This can be removed by PNGase F (= enzyme that removes all N-linked oligosaccharides from glycoproteins) and can be used for further analysis.
[23] pH should be lower than pH=3. This is done to make sure that there is nothing left to precipitate at a low pH prior to loading it on C18 material. Otherwise this would clog the C18 material when you add 0.1% TFA.
[24] Maximum volume of MS vial is 350 µL (use SpeedVac to concentrate the sample if volume is bigger)