Palacký University Olomouc

Bachelor‘s thesis

Olomouc 2016 **Tomáš Oliva**

Palacký University Olomouc

Faculty of Science

Department of cell biology and genetics

Method development and optimization for the discovery of serum based biomarkers in a xenograft mouse model

Bachelor‘s thesis

Tomáš Oliva

Study programme: Biology

Field of study: Molecular and cell biology

Form of study: Full time

Olomouc 2016 **Supervisor: M.Sc. Lakshman Varanasi, Ph.D.**

Declaration

I hereby declare that I have written this bachelor thesis on my own and that I have used only the sources listed in the references.

In Olomouc to date ……………………… ………………………………

Tomáš Oliva

Summary

Colorectal cancer (CRC) is the fourth most common cause of death worldwide. Each year more than 600,000 people die from this disease. Although CRC can be easily cured by surgical resection in early stages, most of the cases are discovered late, when the cancer starts to metastasize. Therefore, there is an urgent need to discover new and more sensitive methods of early diagnosis and prognosis.

The purpose of this study is to compare two protocols for the general improvement of the solid-phase N-glycopeptide extraction workflow in the lab. The current workflow is based on that by Tian *et al* (Protocol 1) and it is compared with one based on Ossola *et al* (Protocol 2). The newer workflow can also be adapted to a robotic platform and this is another reason why the comparison is of value. Serum from SCID mice was used as sample for the workflows and data acquired on an Orbitrap mass spectrometer. Raw data was analyzed using software developed in-house. From the samples measured for the comparison, Protocol 1 yields slightly more N-glycopeptides than Protocol 2. This suggests that the sonication of serum prior to the workflow, or the higher temperature of the reduction reaction, or both, contributes to the results. Additionally, for the same samples, HCD fragmentation also doesn’t appear to work as well as fragmentation by CID. These experiments will need to be repeated on additional biological replicates to verify if the aforesaid results.

Souhrn

Kolorektální karcinom (CRC) je čtvrtou nejčastější příčinou smrti na světě, přičemž ročně tomuto onemocnění podlehne více než 600,000 lidí. Ačkoliv může být kolorektální karcinom chirurgicky jednoduše odstraněn v počátečních fázích, většina případů CRC je diagnostikována později, když začíná rakovina vytvářet metastáze. Je tedy nutné vyvinout nové a citlivější metody včasné diagnózy.

Tato práce si klade za cíl srovnat dva protokoly extrakce na pevné fázi za účelem zlepšení procesu izolace N-glykopeptidů v naší laboratoři. Námi rutinně využívaný protokol odvozený od Tiana *et al* (Protokol 1) je srovnáván s druhým, odvozeným od Ossoly *et al* (Protokol 2). Protokol 2 je možné zautomatizovat, což je dalším důvodem, proč je toto srovnání opodstatněné. Sérum z SCID myší bylo použito jako vzorek pro oba protokoly. Identifikace N-glykopeptidů byla provedena pomocí Orbitrap hmotnostního spektrometru. Data byla analyzována softwarem vyvinutým v naší laboratoři. Z výsledků vyplynulo, že pomocí Protokolu 1 bylo izolováno větší množství N-glykopeptidů než s využitím Protokolu 2. Z toho lze usoudit, že sonifikace, či vyšší teplota při redukci disulfidických můstků, nebo obojí, má vliv na množství identifikovaných N-glykopeptidů ve vzorku. Dále bylo zjištěno, že fragmentace pomocí HCD není tak efektivní jako fragmentace pomocí CID. Pro ověření výše zmíněných výsledků je nutné oba experimenty opakovat na dalších biologických replikátech.

Acknowledgement

I would like to thank my supervisor Lakshman Varanasi, Ph.D., for his mentorship and for his patience during the work on my thesis. I am grateful to other members of the proteomics facility, namely Mgr. Martina Jakoubková, Ing. Miroslav Hruška and Mgr. Dušan Holub, for their help and support.

Finally I would also like to thank doc. MUDr. Marian Hajdúch, Ph. D., for the opportunity to do my bachelor thesis at the Institute of Molecular and Translational Medicine in Olomouc.

Table of contents

1 GOALS OF THE THESIS ... 7
2 INTRODUCTION ... 8
3 LITERATURE REVIEW ... 9
3.1 Post-translational modification ... 9
3.1.1 Glycosylation ... 9
3.2 Isolation of N-glycosylated proteins and peptides ... 11
3.2.1 Solid-phase extraction of glycopeptides ... 11
3.2.2 Hydrazide chemistry ... 12
3.2.3 PNGase F ... 13
3.3 Aberrant glycosylation ... 13
3.4 Protein biomarkers ... 14
3.4.1 Types of biomarkers ... 14
3.4.2 Biomarker development ... 15
4 MATERIAL AND METHODS ... 18
4.1 Biological material ... 18
4.2 Reagents ... 18
4.3 Equipment ... 18
4.4 Serum preparation ... 19
4.5 Protein concentration measurement ... 19
4.6 Solid-phase extraction protocol 1 ... 20
4.7 Solid-phase extraction protocol 2 ... 22
4.8 Mass spectrometric analysis ... 24
4.9 Data analysis ... 24
5 RESULTS ... 26
6 DISCUSSION ... 29
7 CONCLUSION ... 32
8 REFERENCES ... 33
9 ABREVIATIONS ... 38
10 APPENDIX ... 40

1 GOALS OF THE THESIS

The goal of the thesis was to compare two different solid-phase extraction protocols (Protocol 1 and Protocol 2) in order to discover whether Protocol 2 is suitable for high-throughput automation.

The theoretical part focuses on literature review of the topic of protein post-translational modification, mainly N-linked glycosylation and methods associated with isolation of N-linked glycopeptides. The literature review describes the biomarker development process.

The experimental part focuses on preparation of the sample for mass spectrometry using two different solid-phase extraction methods, to identify glycopeptides in those samples using mass spectrometry and to compare the results.

2 INTRODUCTION

Colorectal cancer (CRC) is one of the leading causes of dead worldwide. More than 1.2 million people are diagnosed with colorectal cancer each year and about half of them die from this disease (Jemal *et al*., 2011). The prognosis of patients with colorectal cancer mainly depends on the stage of CRC at the time of a diagnosis (Siegel *et al*., 2012). Current diagnostic methods include mainly various types of colonoscopy and fecal occult blood testing. These methods are commonly used but are invasive and may require several conditions to be performed effectively. In recent years, there is an endeavor to find methods of CRC diagnosis that are more effective and minimally invasive. Those methods mainly include the use of blood serum or plasma based biomarkers.

Currently, there are only two clinically used biomarkers in CRC diagnosis – carcinoembryonic antigen and cancer antigen 19-9 used separately or together. Nevertheless, their specificity to colorectal cancer is not optimal (Polat *et al*., 2014; Vukobrat-bijedic *et al*., 2013). More biomarkers are therefore needed for diagnosis, for prognosis and for determining the efficacy of a treatment regimen.

Liquid chromatography-mass spectrometry (LC-MS) is the method of choice for proteomic biomarker development studies. Sample preparation for analysis by LC-MS determines to a large extent the number of candidate protein biomarkers identified, and consequently, the entire biomarker development process. We conduct here a comparison of two protocols for sample preparation for mass spectrometry- one currently in use in our laboratory, and the other, a slightly modified one, reported more recently. The purpose of this general workflow improvement study is to determine if the newer protocol can identify a comparable, or greater, number of N-glycopeptides. The study is part of a larger project for identifying N-glycoprotein indicators of gastrointestinal cancer.

3 LITERATURE REVIEW

3.1 Post-translational modification

Proteins are biological macromolecules consisting of one or more chains of amino acids. The complement of all proteins in one cell or one organism is called proteome. The study of the proteome is called proteomics.

Post-translational modifications are covalent and generally enzymatic modifications of the precursor protein after biosynthesis. Several post-translational modifications have been reported in recent decades.

3.1.1 Glycosylation

Glycosylation is one of the most common and important post-translational modification of proteins. More than fifty percent of all proteins occurring in the human proteome are estimated to be glycosylated (Apweiler *et al*., 1999, Christiansen *et al*., 2014). Variations in oligosaccharide structure are associated with various physiological functions, for instance, cell differentiation and migration, cell signaling, immune response, cell adhesion and proliferation (Reis *et al*., 2010; Sethi *et* Fanayan, 2015).

Glycosylation is a process occurring in both the endoplasmic reticulum (ER) and Golgi apparatus (GA) where chains of monosaccharides are attached to proteins, lipids or other organic compounds. Maturation and differentiation of glycoproteins take place in the Golgi apparatus. The whole process is catalyzed by a wide range of glycosyltransferases, glycosidases and other specific glycosylation enzymes (Alberts *et al.*, 2007).

3.1.1.1 N-linked glycosylation

N-linked glycosylation was first observed in the early 1960s in hen egg ovalbumin (Johansen *et al*., 1961). It has been predicted that more than half of all eukaryotic proteins are glycosylated and that more than 90% of these are likely to be N-glycosylated (Apweiler *et al*., 1999, Christiansen *et al*., 2014). Later, it was postulated that glycans are linked to the asparagine residue in a conservative acceptor sequence called sequon. The sequon for N-linked glycosylation consist of Asn-Xaa-Ser/Thr, where Xaa stands for any amino acid except proline. Nevertheless, not every Asn-Xaa-Ser/Thr sequon in proteins is glycosylated. The processes behind this phenomenon are still unclear, although it has been reported that protein conformational changes can play a role (Apweiter *et al.*, 1999).

N-linked glycosylation is the most common protein modification of membrane and secreted proteins of eukaryotes (Mohorko *et al*., 2011). To re-iterate, glycosylation (both N-linked and O-linked) occurs in both the ER and Golgi apparatus. In the ER, a core oligosaccharide chain is attached to a protein. This chain consists of three molecules of glucose, nine molecules of mannose and two of N-Acetylglucosamine (GlcNAc) (Alberts *et al*., 2007, Murray *et al.*, 1998). In the Golgi apparatus, manossyl groups may be removed from the core chain to generate a trimannosyl core. Depending on the various saccharide groups that are further attached to this core, three types of N-glycans are produced – complex, high-mannose and hybrid (**Figure 1**). Terminal sequences can be modified by attachment of sialic acid or fucose by sialyltransferases and fucosyltransferases, respectively. N-glycans play role in protein folding, protein conformation and affect solubility. They are also important determinants of antigenicity (Stanley *et al*., 2009).

Figure 1: Types of N-glycans (Stanley *et al.*, 2009).

3.1.1.2 O-linked glycosylation

O-linked glycosylation is a process where glycosylation linkage is created between a side-chain hydroxyl group of an amino acid (Ser, Thr, Tyr, Hyp, Hyl) and a carbohydrate (Spiro, 2002). In contrast to N-glycosylation, no conservative sequon has been identified for O-glycosylation. However, many studies have reported that O-glycosylation usually occurs in proteins with clusters of Ser/Thr residues (Hagen *et al*., 2001; Waren *et al.*, 1993). Typical examples of such proteins are secretory or membrane mucins containing Ser/Thr rich sites (Hagen *et al*., 2001; Reis *et al*., 2010).

3.2 Isolation of N-glycosylated proteins and peptides

With the recognition of glycosylation as an important feature of cancer and cancer pathogenesis, it is necessary to introduce better approaches for glycoprotein/glycopeptide isolation. The complexity of the N-glycoproteome is what complicates its isolation. A variety of sample preparation methods have been developed for this purpose (Ongay *et al*., 2012). For instance, lectin-based capture techniques, chromatography-based measurements such as hydrophilic interaction liquid chromatography (HILIC) and methods based on covalent interactions. Generally the focus of these approaches is to isolate the glycoproteome and in the process decrease sample complexity and enrich low abundant glycoproteins. The approach based on covalent interactions will be further discussed bellow (Ongay *et al*., 2012).

3.2.1 Solid-phase extraction of glycopeptides

Solid-phase extraction was first applied to the isolation of N-glycopeptides in 2003 (Zhang *et al*., 2003). Briefly, carbohydrate cis-diol groups are oxidized to aldehydes using an oxidizing agent such as Sodium periodate. The aldehyde groups then react with hydrazide groups immobilized on solid support. The immobilized glycopeptides are then released by cleavage of the hydrazide bond with a glycosidase such as PNGase F. The original protocol was later modified (Ossola *et al*., 2011; Tian *et al.*, 2007)

Although it is possible to capture both N-linked and O-linked glycopeptides, these methods are limited for isolation of N-linked glycopeptides. To date, there are no available glycosidases allowing to cleave the linkage between O-linked glycopeptides and hydrazide support with enough specificity (Yang *et al*., 2013). Therefore groups of enzymes or chemical methods like β-elimination are used (Harvey, 2005).

3.2.2 Hydrazide chemistry

Hydrazides are organic compounds having a covalent bond between two nitrogen atoms. Hydrazides are routinely used in molecular biology and biochemistry for making a crosslink between the hydrazine on the solid-phase matrix and oxidized carbohydrate moieties of glycoproteins or polysaccharides (Tian *et al.*, 2007).

Aldehydes and ketones are organic compounds containing one or more carbonyl groups. The polarity of the carbonyl bond makes the carbon atom electrophilic and reactive to nucleophiles like primary amines. Although aldehydes do not naturally occur in proteins, they can be generated in reducing sugars such as those commonly attached to proteins during glycosylation (Murray *et al*., 1998)

The aldehyde group can be generated from carbohydrate groups by treatment with strong oxidizing agents like Sodium periodate. The result of carbohydrate oxidation is a change of the hydroxyl group into the final aldehyde group. The aldehyde group reacts with the amino groups of hydrazine and creates a covalent bond (Ongay *et al*., 2012). The principle is described in **Figure 2**.

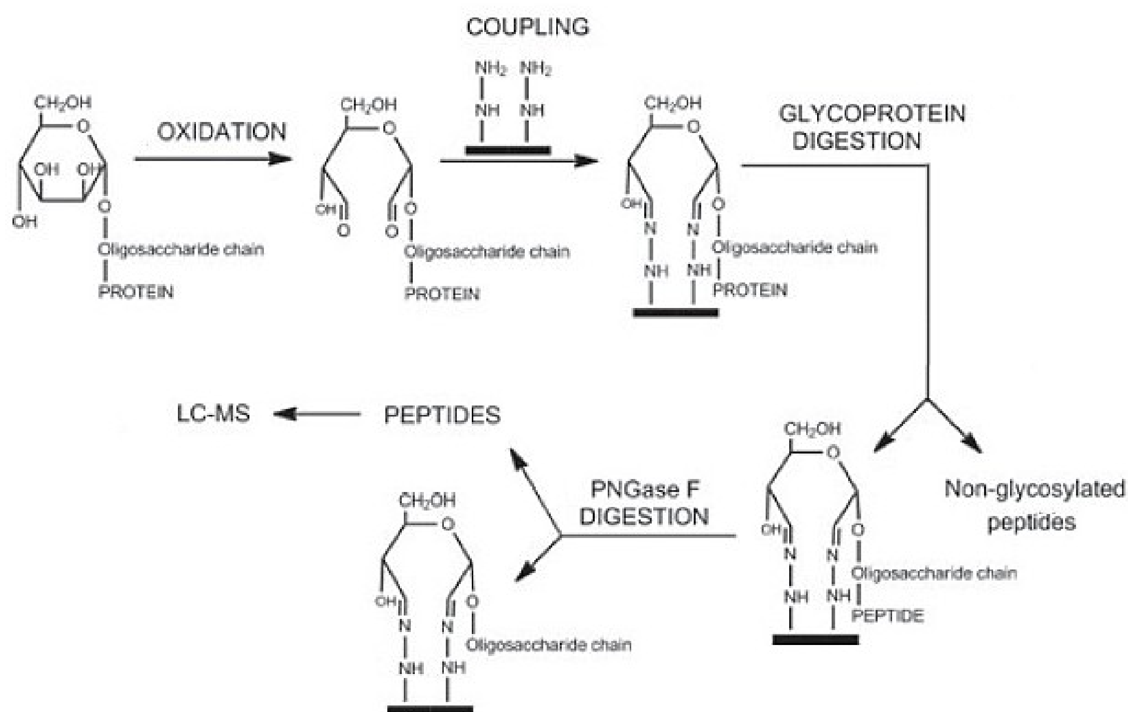

Figure 2: Principle of hydrazide chemistry in solid-phase extraction of glycopeptides (Ongay *et al*., 2012).

3.2.3 PNGase F

Peptide-N-Glycosidase F, shortly PNGase F, is one of the most commonly used enzymes for releasing N-linked carbohydrates from biological macromolecules. It was first discovered and isolated in 1984 from the Gram-negative soil bacteria *Flavobacterium meningosepticum* (Plummer *et al.*, 1984). PNGase F can specifically cleave the linkage between N-acetylglucosamine of high-mannose, hybrid and complex oligosaccharides and the asparagine residuum of N-linked glycoproteins. This cleavage results in deamidation of asparagine to asparate, and an accompanied mass change of 0.98 Da. This is characteristic of the cleavage of a glycosidic bond by PNGase F. This feature is made use of in the reliable identification of the N-glycopeptides (Yang *et al.*, 2010). PNGase F is most active in the pH range from 7.5 – 9.5, while the optimal pH is 8.5 (Tarentino *et al.*, 1990).

3.3 Aberrant glycosylation

It is well known, that due to the changes associated with protein functions, altered glycosylation leads to a number of diseases including cancer (Dube *et al.*, 2005). In terms of cancer, altered glycosylation is recognized as one of the hallmark events because a number of tumor-specific glycoproteins play a pivotal role in cancer growth, invasion and metastasis (Hoja-Lukowicz *et al.*, 2012; Kim *et al.*, 2012; Wei *et al.*, 2012). Aberrant glycosylation is usually a result of disruption of normal functions of enzymes in the glycosylation machinery of affected cells (Dube *et al.*, 2005).

Glycosylation is characterized and controlled by various factors (Holst *et al.*, 2015):

1. Differential expression of glycosyltransferases
2. Availability and localization of carbohydrate donors and transporters
3. Competition between glycosyltransferases
4. Aberrant expression of glycosidases
5. Transfer speed of proteins through ER-Golgi network

The differences in expression of glycosyltransferases are considered as their main cause of aberrant glycosylation in cancer (Meany *et al.*, 2011). Presently the best characterized glycosyltransferase is N-acetylglucosaminyltransferase V (GnT-V). Increased expression of GnT-V and its products is commonly observed in malignancies and it is associated with tumor growth and metastasis (Murata *et al.*, 2000; Pinho *et al.*, 2012; Taniguchi *et* Korekane, 2011).

The competition between two glycosylation enzymes with the same substrate can affect the structure of the final glycan and can also lead to altered glycosylation (Zhao *et al*., 2006).

Although altered glycosylation is considered as one of the hallmark of carcinogenesis, it is unclear whether those changes are a cause or result of cancer (Sethi *et* Fanayan, 2015). Nevertheless, as many of the clinically used biomarkers are glycoproteins, the significance of altered glycosylation has been recognized. The examples of those glycoproteins are Her2/Neu in breast cancer, CA-125 in ovarian cancer, prostate specific antigen (PSA) for prostate cancer or carcinoembryonic antigen (CEA) and cancer antigen 19-9 (CA 19-9) in colorectal cancer (Bottoni *et* Scatena, 2015; Kos *et* Dabbs, 2016; Peracaula *et al*., 2008).

3.4 Protein biomarkers

Biomarkers were at first defined as a cellular, biochemical or molecular changes which can be measured in biological media like tissues, cells or body fluids (Hulka, 1990). This definition was accepted for more than one decade. After that it was changed into its current formulation. Biomarker is a measurable indicator of normal or aberrant physiological processes (Atkinson *et al*., 2001).

Biomarkers can be identified in various types of biological samples such as blood plasma, serum, urine, saliva, cerebrospinal fluid, peritoneal lavage or cancer tissue (Oldenhuis *et al*., 2008).

3.4.1 Types of biomarkers

Biomarkers can be divided into three main types: prognostic, predictive and diagnostic (Mayeux, 2004). Currently, the aim of a biomarker development process is to identify and validate diagnostic, prognostic or predictive biomarkers. The type of the biomarker depends on the patient cohort.

Prognostic biomarkers are parameters useful in determining patient prognosis or disease outcome. The presence or absence of prognostic biomarkers gives us information which can be useful in the selection of patients for a specific treatment, but it does not predict the response to this treatment. Prognostic biomarkers can be divided into two groups: (1) Biomarkers which inform us about possible recurrence and (2) biomarkers that correlate with the duration of survival of metastatic patients (Oldenhuis *et al*., 2008).

Predictive biomarkers are parameters that predict the effect of a therapeutic intervention (Oldenhuis *et al*., 2008).

Diagnostic biomarkers are proteins that indicate the presence of a disease. Some rare genetic variations, posttranslational modifications or an increased or decreased level of specific protein may be a diagnostic biomarker of a disease. The issue with biomarker development is that current biomarkers lack specificity or sensitivity (Lech *et al*., 2016)

When a biomarker is specific enough, it can be used for overall screening. Successful screening tests are generally beneficial to clinical outcome. In some cases, namely colorectal cancer, early detection of a malignancy provides opportunities to remove an affected tissue before it metastasizes to other organs. In case of CRC there are two clinically used diagnostic biomarkers: CEA which can be used individually or in combination with CA 19-9 to provide more accurate information. However, use of those proteins as screening biomarkers is debatable due to their low specificity (Yamashita *et* Watanabe, 2009).

3.4.2 Biomarker development

To re-iterate, biomarker is a parameter which can be objectively measured and evaluated as an indicator of normal or aberrant physiological processes (Atkinson *et al*., 2001).

For many centuries now, signs of illness such as body temperature, tremors or rashes have been known to a medical science. However, better scientific approaches, methods and techniques have pushed biomarker discovery to the molecular level. One of the goals of clinical proteomics is to identify proteins and to develop assays that enable use of these proteins in the clinics (Surinova *et al*., 2010).

Biomarkers can be determined from large variety of biological samples such as blood plasma, serum, urine and cancer tissues or peritoneal lavage fluid (Oldenhuis *et al*., 2008). Some of these samples can be obtained less invasively than others and this is an importance in the biomarker development process. The discovery of new biomarkers is a complex process and it is necessary to divide the development process into separate phases. Those include identification of possible candidates, preclinical verification and validation of these candidates and finally clinical trials (Drabovich *et al*., 2014; Surinova *et al*., 2010)

3.4.2.1 Identification

Protein biomarker identification is one of the three crucial parts of biomarker development. The instrument of choice used in this phase is mass spectrometry (MS). Particularly in recent years, mass spectrometry based proteomics has become powerful

technology for discovery new biomarker candidates (Surinova *et al*., 2010). This advancement is partly caused by improvement of the mass spectrometric instrumentation itself, but also by the development of new associated techniques and methods. The technology has advanced far enough to be able to detect proteins in even picomolar concentrations (Domon *et* Aebersold, 2006).

The most commonly used technique for protein biomarker discovery is the “shot-gun” proteomic approach for complex protein analysis. “Shot-gun” proteomics uses high performance liquid chromatography (HPLC) in combination with mass spectrometry. Samples are processed using a bottom-up proteomic workflow (Meissner *et al*., 2014). The current workflows yield thousands of proteins from biological sample in a single run (Aye *et al*., 2010).

The reason for using blood as a sample for biomarkers is the fact that blood circulates through the body and its composition reflects the physiological state of the body (Surinova *et al*., 2010). However, what makes the biomarker identification process more difficult is the presence of a few abundant proteins such as serum albumin, which does not allow identification of peptides present in trace quantities. This is why methods for identification must be sensitive enough to identify protein in low concentrations and in presence of abundant proteins (Anderson, 2003; Surinova *et al*., 2010).

3.4.2.2 Verification and validation

In the past, the main verification and validation method was ELISA (Parker *et* Borchers, 2014). A disadvantage of this method is the need for one or two different antibodies, depending on a type of ELISA, against studied protein or peptide. These antibodies are not always available for all proteins in the proteome and expensive and time-consuming research is needed for their synthesis (Haab *et al*., 2006). Nevertheless, ELISA is still used when it is required to verify and validate amount of tens biomarkers. There is a need for new techniques that enable the simultaneous validation of hundreds or even thousands of candidate biomarkers. In other words, there is a need for a technique that can deliver higher throughput.

When ELISA is not suitable, an alternative targeted proteomic method called selected reaction monitoring (SRM) or sometimes multiple reaction monitoring (MRM) is used. SRM is performed on a triple quadrupole mass spectrometer. SRM is based on specific fundamental attributes of peptides. The mass to charge ratio (m/z) of a peptide, the energy applied to fragment and the m/z ratio of the product ion fragment are all constants

and characterize the peptides. These attributes are used by SRM to measure the abundance of a peptide in a given sample (Addona *et al*., 2009; Cohen-Freue *et* Borchers, 2012; Lange *et al*., 2008). Modern SRM approaches allow the measurement of more than 100 single peptides in a one hour run. In recent years, SRM has become the "gold standard" of biomarker validation (Stahl-Zeng *et al*., 2007).

After verification and validation biomarkers are advanced to clinical trials.

4 MATERIAL AND METHODS

4.1 Biological material

- SCID mice – purchased from Envigo (formerly Harlan Laboratories)

4.2 Reagents

- Bicinchoninic Acid (Sigma)
- Copper (II) Sulfate Pentahydrate 2% solution (Sigma)
- Hydrazine resin: Affi-Prep beads (Bio-Rad Laboratories)
- Sodium periodate (Sigma)
- Tris (2-carboxyethyl) phosphine (Sigma)
- Iodoacetamide (Sigma)
- Potassium phosphate (Sigma)
- Ammonium bicarbonate (Sigma)
- Urea (Sigma)
- Sodium dodecyl sulfate (Sigma)
- Hydrochloric acid: 17.3% (v/v) (Sigma)
- PNGase F (New England BIolabs)
- Trypsin (Promega)
- Acetonitrile (J.T. Baker)
- Trifluoroacetic acid (TFA): 99% (v/v) (Sigma)
- Dithiothreitol (Sigma)
- Tris-HCl (Sigma)
- Bovine serum albumin (Sigma)
- Formic acid: 98% (v/v) (Sigma)
- Deionized water

4.3 Equipment

- 96-well plate (Sigma)
- Incubator (Biotech MS Incubator)
- Plate reader (Perkin-Elmer EnSpire)
- Vacuum concentrator (Eppendorf concentrator 5301)
- Orbitrap mass spectrometer (Thermo)

- SepPak C-18 Column (Waters)
- Centrifuge (Eppendorf centrifuge 5810R)
- Vortex (Heidolph reax control)
- pH meter (Denver instrument)
- Microfuge (Eppendorf minispin)
- Probe sonicator (Branson digital sonifier)
- Rotator (BioSan Multi RS-60)

4.4 Serum preparation

Blood was processed into serum within 60 minutes after collection. It was centrifuged at 840 g for 10 minutes at 4 °C, aliquoted and stored at -80 °C.

4.5 Protein concentration measurement

Protein concentration was measured using BCA assay in 96-well plate. Firstly, according to **Table 1**, BCA working reagent containing Bicinchoninic acid (Reagent 1, 98% (v/v)) and Cu(II) sulfate pentahydrate solution (Reagent 2, 2% (v/v)) was prepared.

Seven different standard dilutions and a blank were prepared in duplicate in a 96-well plate. The wells contained quantities of BSA ranging from 100 µg to 1200 µg (**Table 2**). 200 µl of BCA reagent was added quickly to each well and the plate then incubated at 37 °C for 30 minutes. The absorbance of the samples was measured at 562 nm on a Perkin-Elmer EnSpire plate reader and plotted against protein quantity in the manner shown (**Figure 5** and **Table 3**). The data was fit to a linear regression (R^2 = 0.9987) to generate the equation,

$$y = 0.0518x + 0.1333 \qquad \text{Eqn 1}$$

(**Figure 5**), where y denotes absorbance and x protein quantity. The equation may be rearranged to solve for x, where:

$$x = (y - 0.1333)/0.0518 \qquad \text{Eqn 2}$$

Average absorbance values of the different samples can be substituted for y and total protein quantity calculated.

Table 1: Representative volume of BCA working reagents.

Number of wells	Amount of each reagent used		
	Reagent 1 [ml]	**Reagent 2 [ml]**	**Total volume [ml]**
10	2	0.04	2.04

Table 2: Quantity of BSA standards.

Serial No.	Protein standard [µg]	BCA working reagent [µl]
1	0	200
2	100	200
3	200	200
4	400	200
5	600	200
6	800	200
7	1000	200
8	1200	200

Serum samples from SCID mice are diluted ten-fold and 1 µl of diluted sample used for the assay in the manner described above. The protein quantity calculated from Eqn 2 is then divided by the volume of the sample used for the assay (1 µl) to obtain the protein concentration. The concentration values are then increased by a factor of 10 to account for the ten-fold dilution and to obtain the actual concentrations (**Table 4**).

4.6 Solid-phase extraction protocol 1

Protocol 1 is based on Tian *et al* (2007). The comparison of Protocol 1 and Protocol 2 with the original N-glycopeptides solid-phase extraction protocol is showed in **Figure 3** (Ossola *et al.*, 2011; Zhang *et al.*, 2003).

After a determination of protein concentration, samples were diluted 10-fold in freshly prepared denaturing buffer containing 8 M urea in ammonium bicarbonate with 0.1% (w/v) SDS. Total amount of 1.2 mg of protein was used for the protocol. Samples were first sonicated for 1 minute at room temperature using a duty cycle of 0.5 and amplitude of 35%. 100 mM stock solution of TRIS-Phosphine was then added to the sample to a final concentration of 10 mM. Samples were incubated at 60 °C for 60 minutes. After incubation, 200mM iodoacetamide stock solution was added to the sample to a final concentration of 12 mM and the incubated at room temperature in the dark for 30 minutes.

Samples were then diluted with potassium phosphate buffer (100 mM stock solution pH 8.0) to reduce urea concentration to 1.6 M. Finally trypsin was added in a ratio 1 µg of trypsin per 200 µg of protein and incubated at 37 °C overnight with agitation.

An appropriate number of C-18 SepPak columns was conditioned by washing twice with 1 ml of 50% Acetonitrile (ACN) in 0.1% Trifluoroacetic acid (TFA) and then twice with 1 ml of 0.1% TFA. Digested peptides were acidified with 17.3% HCl to final pH 3.0. Samples were loaded onto conditioned SepPak column and washed three times with 1 ml of 0.1% TFA. Peptides were eluted twice with 0.2 ml of 50% ACN in 0.1% TFA. Sodium periodate to a final concentration of 10 mM was added to each samples and incubated in the dark at 4 °C for 1 hour.

Samples were then diluted 10-fold in 0.1% TFA. New C-18 SepPak columns were conditioned by washing twice with 1 ml of 50% ACN in 0.1% TFA and then twice with 1 ml of 0.1% TFA. Samples were loaded onto conditioned SepPak column and then washed three times with 1 ml of 0.1% TFA. Elution of peptides was done twice using 0.2 ml of 80% ACN in 0.1% TFA.

Then 25 µl of hydrazide resin (50 µl of 50% slurry) per samples was prepared. The resin was briefly centrifugated (2500 g for 30 seconds) and the solution was removed from the resin. The resin was then washed three times by resuspending the resin in 250 µl of deionized water. Water was then removed after a brief centrifugation. The oxidized peptides were added to the hydrazide support, conjugation was left overnight at room temperature with mixing.

The resin was washed four times with 250 µl of freshly prepared ammonium bicarbonate buffer. After the wash, the resin was resuspended in 25 µl of ammonium bicarbonate buffer and then 3 µl of PNGase F (500U/µl) was added to a vial. Samples were incubated at 37 °C for 4 hours with agitation. After the incubation, the resin was briefly centrifuged (1500 g for 30 seconds) and supernatant was collected. The resin was then washed twice with 100 µl of ammonium bicarbonate buffer. The washes were pooled with the previously collected supernatant.

The new C-18 SepPak column was conditioned by washing it twice with 1 ml of 50% ACN in 0.1% TFA and then twice with 1 ml of 0.1% TFA. Glycopeptides were the acidified with 17.3% HCl to final pH 3.0 and then loaded onto the conditioned C-18 SepPak column. The column was washed three times with 1 ml of 0.1% TFA. Peptides were eluted twice with 0.2 ml of 80% ACN in deionized water. After elution the released glycopeptides

were dried in the microtube within a SpeedVac in room temperature until complete dryness and then stored at -20 °C.

4.7 Solid-phase extraction protocol 2

Protocol 2 is based on Ossola *et al* (2011). The comparison of used protocols with the original protocol is showed in **Figure 3**.

Samples were diluted 10-fold in freshly prepared denaturing buffer containing 8 M urea in ammonium bicarbonate with 0.1% (w/v) SDS. Total 1.2 mg of protein was used for protocol. For a reduction of disulfide bonds 1 M DTT solution was added to the sample to a final concentration of 5 mM. Samples were incubated at 37 °C for 60 minutes. After disulfide bonds reduction 0.5 M iodoacetamide was added to a sample to a final concentration of 25 mM. Incubation was left in dark at room temperature for 30 minutes. Proteins were then diluted with 0.2 M Tris-HCl buffer pH 8.5 to get a final urea concentration of 1.6 M. Trypsin was added to the samples in a ratio 1 µg of trypsin per 200 µg of protein and incubated at 37 °C overnight with agitation.

The C-18 SepPak column was conditioned by washing twice with 1 ml of 50% ACN in 0.1% TFA and then twice with 1 ml of 0.1% TFA. Digested peptides were acidified with 17.3% HCl to final pH 3.0. Samples were then loaded onto conditioned SepPak column and washed three times with 1 ml of 0.1% TFA. Elution was performed with 0.2 ml of 50% ACN in 0.1% TFA. 10 mM freshly prepared sodium periodate was added to each samples and incubated in the dark at 4 °C for 1 hour.

After incubation, samples were diluted 10-fold in 0.1% TFA. A new C-18 SepPak column was conditioned by washing twice with 1 ml of 50% ACN in 0.1% TFA and then twice with 1 ml of 0.1% TFA. Samples were loaded onto column and then washed three times with 1 ml of 0.1% TFA. Peptides were eluted twice using 0.2 ml of 80% ACN in 0.1% TFA.

25 µl of hydrazide resin (50 µl of 50% slurry) per samples was prepared. The resin was centrifuged briefly (2500 g for 30 seconds) and the solution was removed from the resin. The resin was then washed three times by resuspending in 250 µl of deionized water. After a brief centrifugation redundant water was removed. The oxidized peptides were added to the hydrazide support, conjugation was left overnight at room temperature with mixing.

The resin was washed four times with 250 µl of freshly prepared ammonium bicarbonate buffer and then resuspended again in the same buffer. Then 3 µl of PNGase F (500U/µl) was added to the resin. The resin was incubated at 37 °C for 4 hours with agitation.

After the incubation the resin was centrifuged (1500 g for 30 seconds) and a supernatant was collected. The hydrazide was washed twice with ammonium bicarbonate buffer and the washes were polled with previously acquired supernatant.

A new C-18 SepPak column was conditioned by washing twice with 1 ml of 50% ACN in 0.1% TFA and then twice with 1 ml of 0.1% TFA. Acquired glycopeptides were acidified with 17.3% HCl to a final pH 3.0. Samples were loaded onto the column followed with the wash of the conditioned column three times with 1 ml of 0.1% TFA. Peptides were eluted twice with 0.2 ml of 80% ACN in deionized water and then dried in the microtube within a SpeedVac in room temperature until complete dryness. Samples were stored at -20 °C.

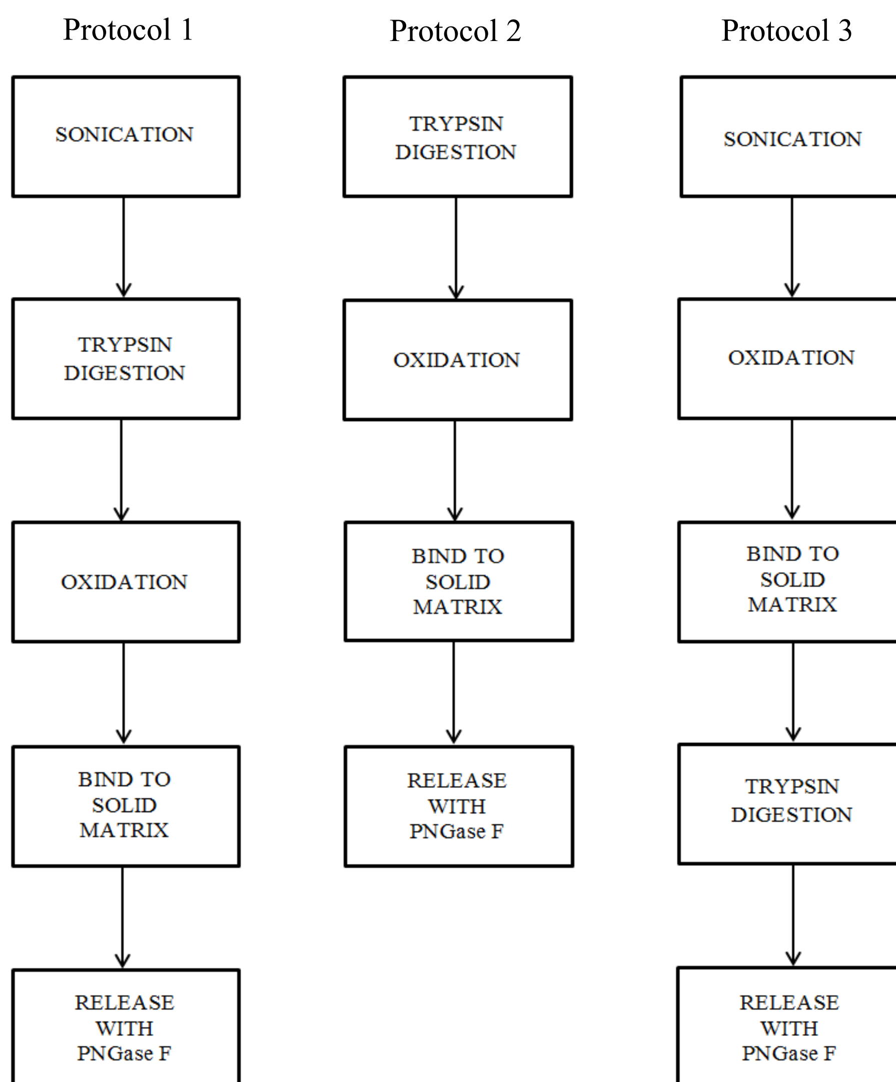

Figure 3: Comparison of the original SPE protocol (Protocol 3) by Zhang *et al* (2003) with Protocol 1 based on Tian *et al* (2007) and Protocol 2 based on Ossola *et al* (2011).

4.8 Mass spectrometric analysis

A total number of eight samples and one BSA control were processed. The samples were resuspended in 100 µl of 5% ACN in 0.1% Formic acid (FA). The control was comprised of 5 µl of BSA standard peptide and 195 µl of 5% ACN in 0.1% FA.

Mass spectrometric analysis was performed on an Orbitrap Elite instrument (Thermo Fisher Scientific) with a Proxeon Easy-Spray ionization source coupled with an Ultimate 3000 RSLCnano chromatograph. Ten microliter of the sample were loaded onto a PepMap 100 desalting column (Thermo Fisher Scientific) and on a PepMap RSLC analytical column (Thermo Fisher Scientific) and heated to 35 °C. Peptides were separated on the analytical column using the organic phase ranging from 5% to 35% with a flow rate of 300 nl per minute for a total of 150 minutes. The organic phase reagents are 0.1% Formic acid and 80% ACN in 0.1% Formic acid.

Data was acquired using a top-20 method. The MS data was collected in the ion-trap in the positive ion mode with a source voltage of 2000 V. Precursor ions with charges 2-6 were selected and scanned with an m/z ranging from 300 to 1500 at a resolution of 120,000 full width at half maximum nominal resolution settings. The cycle was 3 seconds and data was collected in profile mode.

For the first discovery run the twenty most intense ions were selected for collision-induced dissociation (CID) type of fragmentation in the Orbitrap. Normalized collision energy was 30 eV. For a precursor mass window of 10 ppm a dynamic exclusion was applied for 40 seconds. An automatic gain control target of 2.0e5 ions and an accumulation time of 50 ms were applied. MS spectra were collected in the Orbitrap in the normal scan mode using a precursor m/z selection window of 1.6 Th.

A second discovery run was performed using High-energy collision dissociation (HCD). An m/z range from 400 to 1600 was scanned for precursor ions. MS2 was performed in the orbitrap with HCD collision energy of 28% at a resolution of 15,000 full width at half maximum with an automatic gain of control target count of 50,000. MS2 data was collected in centroid mode.

4.9 Data analysis

Collected data was further analyzed using in-house developed software. Mass spectrometric RAW data files were converted to the universal HUPO-recommended mzML format with the MSConvert utility (v 3.0.9098). Identification of peptides was done using software *Dymka* developed in-house in Python. The actual identification

was performed on X!Tandem search engine (Cyclone, 2013.02.01.1). Results were then processed using a number of in-house developed scripts. Visualization was done using Matplotlib (v. 1.3.1).

The algorithm for data analysis is as follows. Identified peptides are scanned for the N-glycosylation consensus sequence/motif (NXS/T). If the peptides exhibit the motif and also a mass change of 0.98 Da, they are considered N-glycopeptides. Murine and human N-glycopeptides are then separated on the basis of their origin (mouse or xenograft). Because the experiment described above uses SCID mice lacking a human xenograft, the discussion is restricted to murine N-glycopeptides (**Figure 4**).

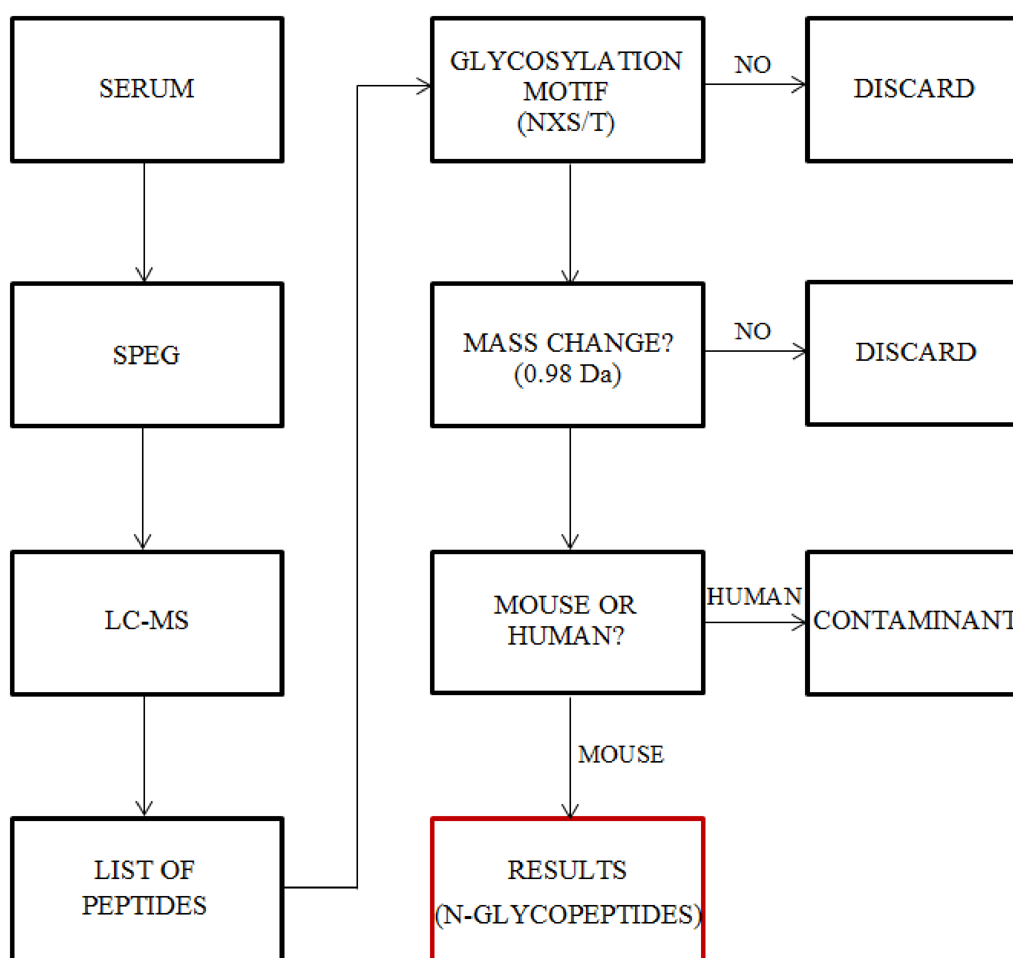

Figure 4: Complete workflow from mice serum to mass spectrometric results.

5 RESULTS

The focus of the thesis was to perform two different solid-phase extraction methods, Tian based protocol (Protocol 1) and Ossola based protocol (Protocol 2) on a series of SCID serum samples to isolate glycoproteome and then to perform mass spectrometric analysis to identify glycopeptides.

Before the both solid-phase extraction protocols were performed, it was necessary to determine protein concentration in all samples so as to use the same quantity of protein as the starting material. The protein concentration was measured using standard BCA assay, as described in the methods section (subsection 4.5). Two replicates of each sample were used and the absorbance measured at 562 nm. Protein concentration was determined from the equation of linear regression of BSA standard curve (**Table 2, 3** and **Figure 5**). Concentrations of proteins in all samples are listed in **Table 4**.

Table 3: Concentration and absorbance of BSA standards.

Protein concentration [µg/µl]	Average absorbance
0	0.1215
1	0.182
2	0.2395
4	0.3505
6	0.453
8	0.5555
10	0.6455
12	0.7475

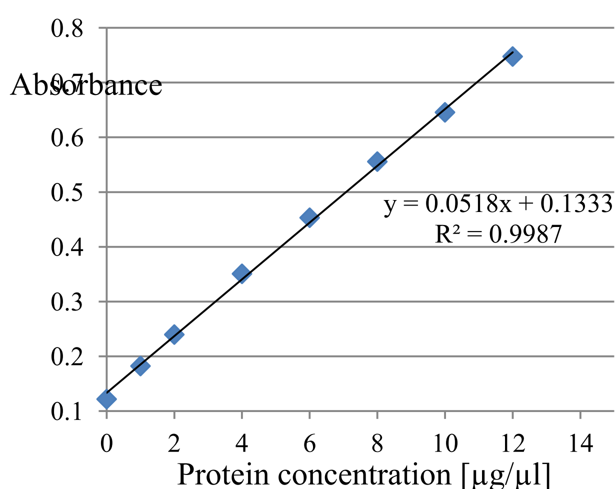

Figure 5: BSA standard curve.

Table 4: Protein concentrations determined from average absorbance of each sample. Abs = absorbance.

Sample	Abs 1	Abs 2	Average abs	Protein quantity [µg]	Protein concentration [µg/µl]	Protein concentration [µg/µl] (10x)
1	0.522	0.483	0.5025	7.13	7.13	71.27
2	0.582	0.521	0.5515	8.07	8.07	80.73
3	0.581	0.604	0.5925	8.68	8.68	86.76
4	0.535	0.574	0.5545	8.12	8.12	81.21

A total quantity of 1.2 mg of serum protein was used as a starting material for both methods. Both protocols were performed together in order to maintain the same work conditions.

The Orbitrap mass spectrometer (Thermo) was used for acquiring data. All samples were measured using two different types of fragmentation techniques, CID and HCD. Mass spectrometric data acquisition is described in more detail in the Methods section.

The numbers of glycosylated peptides identified by each protocol are listed in **Table 5** and **6**. The total number of glycopeptides identified in the CID mode by Protocol 1 was higher by a factor of 1.26. In case of HCD, the Protocol 1 exceeds Protocol 2 by a factor of 1.16. Protocol 1 also yields 1.22 times more glycopeptides than Protocol 2 does.

Although both protocols are used for isolation of glycopeptides, a number of non-glycosylated peptides were identified as well (**Table 7**). This phenomenon will be further discussed in the next section. The relevance of this data for the experiment is being investigated.

Table 5: Total number of glycopeptides identified.

Sample	Protocol 1 CID	Protocol 2 CID	Protocol 1 HCD	Protocol 2 HCD
1	647	133	415	117
2	530	116	268	63
3	483	485	231	208
4	298	566	93	352
Average	490	325	252	185
Total	**738**	**586**	**404**	**349**

Table 6: Unique glycopeptides identified by CID and HCD.

Protocol	CID unique	Both	HCD unique
Protocol 1	378	360	44
Protocol 2	285	301	48

Table 7: Total number of non-glycosylated peptides identified.

Sample	Protocol 1 CID	Protocol 2 CID	Protocol 1 HCD	Protocol 2 HCD
1	2051	681	1045	309
2	1649	697	611	185
3	1541	1613	558	494
4	1172	2170	270	1084
Average	**1603**	**1290**	**621**	**518**

The degree of contamination of the samples by *human* glycopeptides and non-glycosylated peptides was also assessed (**Table 8** and **9**). It is likely that there is greater contamination in Protocol 1 samples because of greater exposure during the initial sonication step. A total of 530 non-glycosylated human peptides and 26 human glycopeptides were identified. The full list of identified contaminants is provided in the appendix (**Table 10** and **11**). According to the results, Protocol 2 shows significantly lower numbers of glycosylated and non-glycosylated peptide contaminants in almost all samples.

Table 8: Total number of human glycopeptides identified.

Sample	Protocol 1 CID	Protocol 2 CID	Protocol 1 HCD	Protocol 2 HCD
1	14	8	6	2
2	7	0	2	1
3	10	5	1	0
4	8	8	0	0
Average	**10**	**5**	**2**	**1**

Table 9: Total number of human non-glycosylated peptides identified.

Sample	Protocol 1 CID	Protocol 2 CID	Protocol 1 HCD	Protocol 2 HCD
1	218	60	21	9
2	224	56	17	15
3	228	155	5	16
4	142	141	8	28
Average	**203**	**103**	**13**	**17**

6 DISCUSSION

Our larger project at the proteomics facility at the Institute of Molecular and Translational Medicine is to discover and validate human protein biomarkers in the serum from SCID mice xenografted with human tumors and to eventually advance those biomarkers into clinics. The rationale for a xenograft mouse model is that any human protein identified in murine serum must have necessarily originated in the implanted tumor.

Automation is essential for processing tens or hundreds of samples daily. The manual processing of samples is not feasible in the clinical setting because of the volume of clinical samples. Another reason is that automation precludes human error during the process and improves the reproducibility of the method.

The focus of this work is a comparison of two different solid-phase extraction techniques of glycopeptides in order to improve the discovery workflow. In the process we try to determine if an automation compatible protocol is better than the protocol we currently use.

A total of eight samples were tested, four biological replicates for each method. The Orbitrap mass spectrometer was used for the analysis. Two types of fragmentation – collision-induced dissociation (CID) and higher-energy C-trap dissociation (HCD) were applied to all samples in order to determine if one type of fragmentation is better for identification of murine glycopeptides than the second. CID and HCD are two of the ways by which the mass spectrometer fragments a peptide that is present in the sample in order to obtain sequence information. The two methods differ in the energy applied to the fragmentation and can yield complementary and confirmatory peptide sequence information.

The glycopeptides and non-glycosylated peptides identified from the two fragmentation modes of the two protocols were compared.

Overall, the number of glycopeptides from Protocol 1 is greater than those from Protocol 2 for both fragmentation techniques. The total number of identified glycopeptides for Protocol 1 using CID was 738, while the average number of identified glycopeptides for Protocol 2 was 586 (**Table 5**). The total number of glycopeptides discovered by Protocol 1 was 404, as compared to 349 by Protocol 2 in HCD mode (**Table 5**). There appears to be significant variation in the results for each sample for a given protocol and fragmentation mode (for instance, **Column 2** of **Table 5**). However, the reason for the variation is unclear. The greater number of glycopeptides identified with Protocol 1

is probably because of the sonication step. Many proteins in serum are likely to aggregate and prevent complete tryptic digestion. These aggregates need to be separated by sonication before proceeding with solid-phase extraction (Ray *et al*., 2011).

These results also show that in this instance HCD is not as effective as CID, since the numbers of identified peptides are lower. Several studies have compared these two types of fragmentation. It appears that the experiment dictates the use of one method over another (Nagaraj *et al*., 2010; Olsen *et al*., 2009; Jedrychowski *et al*., 2011).

Although both of these protocols are used for isolation of glycopeptides, a large number of non-glycosylated peptides were identified as well (**Table 7**). This may be because of natural binding of the two types of peptides to each other. However, more studies are needed for understanding this phenomenon.

Another different factor considered during software data analysis was possible human contamination. The main difference between those two protocols is the absence of sample sonication in case of Protocol 2. Sonication is the step where the samples are exposed to laboratory environment for a significant amount of time. Thus, the samples may have been contaminated during sonication. The numbers of identified contaminating human glycopeptides and peptides are listed in **Table 8** and **9.** The average number of discovered contaminants in case of Protocol 1, where sonication is an essential step, is 203 peptides and 10 glycopeptides for CID. The average number of contaminants for Protocol 2 is two-fold lesser, 103 peptides and 5 glycopeptides (**Table 8** and **9**). A greater number of contaminants were identified by Protocol 1, than by Protocol 2. This is in keeping with the general trend observed in the actual results. The same is true of the two types of fragmentation modes - CID yields more than HCD. The typical contaminants such as the most common types of keratin, products of digestion of used enzymes, such as trypsin or PNGase F, were not identified because they were automatically excluded during computational analysis (Keller *et al*., 2008). It has been suggested that most of the contaminants do not originate in the environment where the glycopeptide isolation was performed, but in the environment where the model organism was grown (Hodge *et al*., 2013). In our case, we cannot tell if the contaminations come from the mice preparation stage or from the sample preparation. The full list of the contaminants is attached in the appendix (**Table 10** and **11**).

Another different aspect of the comparison is the overall time and labor needed for performing each protocol. Both these methods take approximately three days to perform. However, the initial sonication step at the beginning of the Protocol 1 adds one or two hours

to the total time, but more importantly, it renders automation difficult, especially for a large number of samples. Comparison of both protocols is showed in **Figure 3**.

Although the efficiency of Protocol 2 compared to Protocol 1 is lower, Protocol 2 appears to be suitable for automation for the reasons mentioned above.

7 CONCLUSION

The results from the two protocols are comparable with Protocol 1 yielding a slightly larger number of candidate markers in both the CID and HCD modes of peptide fragmentation. CID also appears to result in a larger number of identifications than HCD. There are fewer human contaminants from Protocol 2, possibly because it lacks the initial sonication step. For the same reason, *i.e.* for the sonication step, Protocol 1 requires approximately 1-2 hours more, and this varies with the number of samples.

Because automation of the extraction procedure is desirable, and because Protocol 2 does not require the sonication step, it was of interest to determine if Protocol 2 yielded more candidate markers. It does not. However, the difference in the number of candidate markers from the two protocols is not considerable. More samples and more optimization may lessen the difference. In either case, Protocol 2 will be adapted to the robotic platform for preliminary tests.

8 REFERENCES

Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., Spiegelman, C. H., Zimmerman, L. J., Ham, A. J. L, Keshishian, H., Hall, S. C., Allen, S., Blackman, R. K., Borchers, C. H., Buck, C., Cardasis, H. L., Cusack, M. P., Dodder, N. G., Gibson, B. W., Held, J. M., Hiltke, T., Jackson, A.,. Johansen, E. B, Kinsinger, C. R., Li, J., Mesri, M., Neubert, T. A., Niles, R. K., Pulsipher, T. C., Ransohoff, D., Rodriguez, H., Rudnick, P. A., Smith, D., Tabb, D. L., Tegeler, T. J., Variyath, A. M., Vega-Montoto, L. J., Wahlander, A., Waldemarson, S., Wang, M., Whiteaker, J. R., Zhao, L., Anderson, N. L., Fisher, S. J., Liebler, D. C., Paulovich, A. G., Regnier, F. E., Tempst, P., Carr, S. A. (2009): Multi-site Assessment of the Precision and Reproducibility of Multiple Reaction Monitoring–based Measurements of Proteins in Plasma. *Nature Biotechnology*, 27: 633-641.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2007): Molecular Biology of the Cell. 5th Edition (Anderson, M., Granum, S., eds). Garland Science, Taylor & Francis Group, New York.

Anderson, N. L. (2003) The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular & Cellular Proteomics,* 2: 50.

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

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., Woodcock, J., Zeger, S. L. (2001): Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology*, 69: 89-95.

Aye, T. T., Scholten, A., Taouatas, N., Varro, A., van Veen, T. A. B., Vos, M. A., Heck, A. J. R. (2010): Proteome-wide protein concentrations in the human heart. *Molecular BioSystems,* 6: 1917-1927.

Bottoni, P., Scatena, R. (2015): The Role of CA 125 as Tumor Marker: Biochemical and Clinical Aspects. *Advances in Experimental Medicine and Biology*, 867: 229-244.

Cohen-Freue, G. V., Borchers, C. H.. (2012): Multiple Reaction Monitoring (MRM): Principles and Application to Coronary Artery Disease. *Circulation: Cardiovascular Genetics,* 5: 378-378.

Domon, B., Aebersold, R. (2006): Mass spectrometry and protein analysis. *Science* 312: 212-217.

Drabovich, A. P., Martínez-Morillo, E., Diamandis, E. P. (2014): Toward an integrated pipeline for protein biomarker development. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1854: 677-86.

Dube, D. H., Bertozzi, C. R. (2005): Glycans in Cancer and Inflammation — Potential for Therapeutics and Diagnostics. *Nature Reviews Drug Discovery*, 4: 477-88.

Haab, B. B., Paulovich, A. G., Anderson, N. L., Clark, A. M., Downing, G. J., Hermjakob, H., Labaer, J., M. Uhlen. (2006): A Reagent Resource to Identify Proteins and Peptides of Interest for the Cancer Community: A workshop report. *Molecular & Cellular Proteomics*, 5: 1996-2007.

Hagen, K. G. T., Bedi, G. S., Tetaert, D. (2001): Cloning and Characterization of a Ninth Member of the UDP-GalNAc: Polypeptide N-Acetylgalactosaminyltransferase Family, ppGaNTase-T9. *Journal of Biological Chemistry*, 276: 17395-17404.

Harvey D.J. (2005): Proteomic analysis of glycosylation: structural determination of N- and O-linked glycans by mass spectrometry. *Expert Review of Proteomic*, 2: 87–101.

Hodge, K., Have, S. T., Hutton L., Lamond, A. I. (2013): Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *Journal of Proteomics*, 88: 92-103.

Hoja-Lukowicz, D., Link-Lenczowski, P., Carpentieri, A., Amoresano, A., Pochec, E., Artemenko, K. A., Bergquist, J., Litynska, A. (2012): L1CAM from Human Melanoma Carries a Novel Type of N-glycan with Galβ1-4Galβ1- Motif. Involvement of N-linked Glycans in Migratory and Invasive Behaviour of Melanoma Cells. *Glycoconjugate Journal*, 30: 205-25.

Holst, S., Wuhrer, M., Rombouts, Y. (2015): Glycosylation Characteristics of Colorectal Cancer. *Advances in Cancer Research*, 126: 203 – 256.

Hulka, B. S. (1990): Overview of biological markers. In: Biological markers in epidemiology (Hulka, B. S., Griffith, J. D., Wilcosky, T. C., eds.) Oxford University Press, New York.

Christiansen, M. N., Chik, J., Lee, L., Anugraham, M., Abrahams, J. L., Packer, N. H. (2014): Cell surface protein glycosylation in cancer. *Proteomics,* 14: 525-546.

Jedrychowski, M. P., Huttlin, E. L., Haas, W., Sowa, M. E., Rad, R., Gygi, S. P. (2011): Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics. *Molecular & Cellular Proteomics* 10(12): M111.009910

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., Forman, D. (2011): Global cancer statistics. *CA: A Cancer Journal for Clinicians*, 61: 69-90.

Johansen, P. G., Marshall, R. D., and Neuberger, A. (1961): Carbohydrates in protein. The preparation and some of the properties of a glycopeptide from hen egg ovalbumin. *Biochemical Journal,* 78: 518–527.

Keller, B. O., Sui, J., Young, A. B., Whittal, R. M. (2008): Interferences and Contaminants Encountered in Modern Mass Spectrometry. *Analytica Chimica Acta*, 627: 71-81.

Kim, Y.-S., Ahn, Y. H., Song, K. J.,. Kang, J. G, Lee, J. H., Jeon, S. K., Kim, H. C., Yoo, J. S., Ko, J. H. (2012): Overexpression and -1,6-N-Acetylglucosaminylation-initiated Aberrant Glycosylation of TIMP-1: A "double whammy" strategy in colon cancer progression. *Journal of Biological Chemistry*, 287: 32467-32478.

Kos, Z., Dabbs, D. J. (2016): Biomarker assessment and molecular testing for prognostication in breast cancer. *Histopathology*, 68: 70-85.

Lange, V., Picotti, P., Domon, B., Aebersold, R. (2008): Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology*, *4*: 222.

Lech, G., Slotwinski, R., Slodkowski, M., Krasnodebski, I. W. (2016): Colorectal Cancer Tumour Markers and Biomarkers: Recent Therapeutic Advances. *World Journal of Gastroenterology*, 22: 1745 – 1755.

Mayeux, R. (2004): Biomarkers: Potential uses and limitations. *NeuroRX*, 1: 182-188.

Meany, D. L., Chan, D. W. (2011): Aberrant glycosylation associated with enzymes as cancer biomarkers. *Clinical Proteomics*, 8: 1-14.

Meissner, F., Mann, M. (2014): Quantitative Shotgun Proteomics: Considerations for a High-quality Workflow in Immunology. *Nature Immunology,* 15: 112-17.

Mohorko, E., Glockshuber, R., Aebi, M. (2011): Oligosaccharyltransferase: The central enzyme of N-linked protein glycosylation. *Journal of Inherited Metabolic Disease,* 34: 869 –878.

Murata, K., Miyoshi, E., Kameyama, M., Ishikawa, O., Kabuto, T., Sasaki, Y., Hiratsuka, M., Ohigashi, H., Ishiguro, S., Ito, S. (2000): Expression of N-acetylglucosaminyltransferase V in colorectal cancer correlates with metastasis and poor prognosis. *Clinical Cancer Ressearch*, 6: 1772–1777.

Murray, R. K., Granner, D. K., Mayes, P. A., Rodwell, V. W. (1998) Harper's Biochemistry. 23rd Edition. Appleton & Lange, East Norwalk.

Nagaraj, N., D'Souza, R. C. J., Cox, J., Olsen, J. V., Mann, M. (2010): Feasibility of Large-Scale Phosphoproteomics with Higher Energy Collisional Dissociation Fragmentation. *Journal of Proteome Research*, 9: 6786-6794.

Oldenhuis, C. N. A. M., Oosting, S. F., Gietema, J. A., de Vries, E. G. E. (2008): Prognostic versus predictive value of biomarkers in oncology. *European Journal of Cancer*, 44: 946-953.

Olsen, J. V., Schwartz, J. C., Griep-Raming, J. (2009): A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Molecular & Cellular Proteomics*, 8: 2759-2769.

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

Ossola, R., Schiess, R., Picotti, P., Rinner, O., Reiter, L., Aebersold, R. (2011): Biomarker Validation in Blood Specimens by Selected Reaction Monitoring Mass Spectrometry of N-Glycosites. *Methods in molecular biology*, 728: 179 - 194.

Parker, C. E., Borchers, Ch. H. (2014): Mass spectrometry based biomarker discovery, verification, and validation – Quality assurance and control of protein biomarker assays. *Molecular Oncology*, 8: 840-858.

Peracaula, R., Barrabés, S., Sarrats, A., Rudd, P.M., de Llorens, R. (2008): Altered glycosylation in tumours focusedto cancer diagnosis. *Disease Markers,* 25: 207–218.

Pinho, S. S., Oliveira, P., Cabral, J., Carvalho, S., Huntsman, D., Gartner, F., Seruca, R., Reis, C. A., Oliveira, C. (2012): Loss and recovery of Mgat3 and GnT-III Mediated E-cadherin

N-glycosylation is a mechanism involved inepithelial-mesenchymal-epithelial transitions. *PLoS ONE*, 7: e33191.

Plummer, T. H., Elder, J. H., Alexander, S., Phelan, A. W., Tarentino, A. (1984): Demonstration of peptide: N-glycosidase F activity in endo-beta-N-acetylglucosaminidase F preparations. *The Journal of Biological Chemistry*, 259: 10700-10704.

Polat, E., Duman, U., Duman, M., Atici, A. E., Reyhan, E., Dalgic T., Bostanci, E. B., Yol, S. (2014): Diagnostic value of preoperative serum carcinoembryonic antigen and carbohydrate antigen 19-9 in colorectal cancer. *Current Oncology*, 21: e1-7.

Ray, S., Reddy, P. J., Jain, R., Gollapalli, K., Moiyadi, A., Srivastava S. (2011): Proteomic technologies for the identification of disease biomarkers in serum: Advances and challenges ahead. *Proteomics*, 11: 2139-2161.

Reis, C. A, Osorio, H., Silva, L., Gomes C., David, L. (2010): Alterations in glycosylation as biomarkers for cancer detection. *Journal of Clinical Pathology,* 63: 322-329.

Sethi, M., Fanayan, S. (2015): Mass Spectrometry-Based N-Glycomics of Colorectal Cancer. *International Journal of Molecular Sciences*, 16: 29278-29304.

Siegel, R., Desantis, C., Virgo, K., Stein, K., Mariotto, A., Smith, T., Cooper, D., Gansler, T., Lerro, C., Fedewa, S., Lin, C., Leach, C., Cannady, R. S., Cho, H., Scoppa, S., Hachey, M., Kirch, R., Jemal, A., Ward E. (2012): Cancer treatment and survivorship statistics, 2012. *CA: A Cancer Journal for Clinicians*, 62: 220-241.

Spiro, R. G. (2002): Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, 12: 43-56.

Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R., Domon, B. (2007): High Sensitivity Detection of Plasma Proteins by Multiple Reaction Monitoring of N-Glycosites. *Molecular & Cellular Proteomics*, 6: 1809-1817.

Stanley, P., Schachter, H., Taniguchi, N. (2009): N-Glycans. In: Essentials of Glycobiology, 2nd Edition (Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Har, G. W., Etzler, M. E., eds.) Cold Spring Harbor Laboratory Press, New York.

Surinova , R., Cerciello, F., Wollscheid, B., Aebersold, R. (2010): On the Development of Plasma Protein Biomarkers. *Journal of Proteome Research,* 10: 5-16.

Taniguchi, N., Korekane, H. (2011): Branched N-glycans and their implications for cell adhesion, signaling and clinical applications for cancer biomarkers and in therapeutics. *BMB Reports*, 44: 772–781.

Tarentino, A., Quinones, G., Trumble, A., Changchien, L., Duceman,B., Maley, F., Plummer, T. (1990): Molecular cloning and amino acids equence of peptide-N4-(N-acetyl-beta-D-glucosaminyl) asparagine amidase from *Flavobacterium meningosepticum*. *The Journal of Biological Chemistry*, 265: 6961-6966.

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

Vukobrat-Bijedic, Z., Husicselimovic, A., Sofic, A., Bijedic, N., Bjelogrlic, I., Gogov, B., Mehmedovic, A. (2013): Cancer Antigens (CEA and CA 19-9) as markers of advanced stage of colorectal carcinoma. *Medical Archives*, 67: 397-401.

Warren E.C. (1993): Glycosylation. *Current Opinion in Biotechnology*, 4: 596–602.

Wei, T., Liu, Q., He, F., Zhu, W., Hu, L., Guo, L., Zhang, J. (2012): The role of N-acetylglucosaminyltransferases V in the malignancy of human hepatocellular carcinoma. *Experimental and Molecular Pathology,* 93: 8-17.

Yamashita, K., Watanabe, M. (2009): Clinical significance of tumor markers and an emerging perspective on colorectal cancer. *Cancer Science,* 100: 195-199.

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

Yang, S., Li, Y., Shah, P., Zhang, H. (2013): Glycomic Analysis Using Glycoprotein Immobilization for Glycan Extraction. *Analytical Chemistry*, 85: 5555-5561.

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

Zhao, Y., Nakagawa, T., Itoh, S. (2006): N-Acetylglucosaminyltransferase III Antagonizes the Effect of N-Acetylglucosaminyltransferase V on 3beta1 Integrin-mediated Cell Migration. *Journal of Biological Chemistry*, 281: 32122-32130.

9 ABREVIATIONS

ACN	Acetonitrile
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
CA 19-9	Cancer antigen 19-9
CA-125	Cancer antigen 125
CEA	Carcinoembryonic antigen
CID	Collision-induced dissociation
CRC	Colorectal cancer
DTT	Dithiothreitol
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
FA	Formic acid
Fuc	Fucose
GA	Golgi apparatus
Gal	Galactose
GlcNAc	N-Acetylglucosamine
GnT-V	N-acetylglucosaminyltransferase V
HCD	High-energy collision dissociation
Her2/Neu	Erb-b2 receptor tyrosine kinase 2
HILIC	Hydrophilic interaction liquid chromatography
HPLC	High performance liquid chromatography
LC-MS	Liquid chromatography–mass spectrometry
Man	Mannose
MRM	Multiple reaction monitoring
MS	Mass spectrometry
PSA	Prostate-specific antigen
RSLC	Rapid separation liquid chromatography
SCID	Severe combined immunodeficiency
SDS	Sodium dodecyl sulfate
Sia	Sialic acid
SRM	Selected reaction monitoring
TFA	Trifluoroacetic acid

10 APPENDIX

Table 10: Full list of identified human glycopeptide contaminants; 1,2,3,4 – sample, P1 – protocol 1, P2 – protocol 2, C – CID, H – HCD.

Number of samples with contaminant	Names of samples with contaminant	Glycopeptide
6	1P1C, 2P1C, 3P2C, 4P1C, 1P2H, 3P1C	Phosphatidylinositol-glycan-specific phospholipase D
4	1P1C, 2P1C, 1P1H, 3P1C	Maltase-glucoamylase, intestinal
3	4P1C, 4P2C, 3P1C	Ig gamma-4 chain C region
3	1P2C, 2P2H, 1P2H	Leucine-rich alpha-2-glycoprotein
2	1P1H, 4P2C	Isoform 2 of Probable G-protein coupled receptor 116
1	4P1C	Zinc finger protein 532
1	1P1C	Isoform 3 of Roundabout homolog 2
1	1P2C	Isoform 4 of Double-stranded RNA-specific editase 1
1	3P2C	Isoform 2 of Unconventional myosin-IXa
1	4P1C	Ig gamma-1 chain C region
1	3P1C	Cathepsin D
1	4P1C	Insulin-like growth factor 1 receptor
1	2P1C	Isoform 11 of Myocardial zonula adherens protein
1	2P1C	Isoform Beta-4B of Integrin beta-4
1	2P1C	Tripeptidyl-peptidase 2 O
1	3P1H	Serine/arginine repetitive matrix protein 2
1	4P1C	F-box/LRR-repeat protein 3
1	3P1C	Proteolipid protein 2
1	1P2C	Ankyrin-3
1	1P2C	Stromal interaction molecule 1
1	4P2C	Glycine receptor subunit alpha-4
1	1P1C	Isoform 5 of Isthmin-2
1	1P1C	Isoform 3 of Transmembrane and ubiquitin-like domain-containing protein 2
1	3P2C	Histone-lysine N-methyltransferase SUV420H2
1	4P2C	Isoform 4 of Nesprin-1
1	2P1H	DNA excision repair protein ERCC-6-like 2

Table 11: Full list of identified human peptide contaminants; 1,2,3,4 – sample, P1 – protocol 1, P2 – protocol 2, C – CID, H – HCD.

Number of samples with	Names of samples with	Peptide

contaminant	contaminant	
6	3P2C, 1P1C, 2P1C, 4P1C, 1P2H, 3P1C	Phosphatidylinositol-glycan-specific phospholipase D
5	3P1C, 2P1C, 1P1C, 4P1C, 2P1H	Isoform 3 of Conserved oligomeric Golgi complex subunit 5
5	3P1C, 2P1C, 1P1C, 4P1C, 3P2H	Isoform 2 of Serum albumin
5	1P1C, 2P1C, 4P2H, 1P1H, 3P1C	Maltase-glucoamylase, intestinal
4	2P1C, 1P2C, 2P2C, 4P2C	Zinc finger protein 442
4	4P1C, 4P2H, 1P2H, 3P1C	Uncharacterized protein C1orf94
4	4P2C, 3P2C, 2P1C, 1P1C	Cyclic AMP-dependent transcription factor ATF-6 alpha
4	2P1C, 1P2C, 2P2C, 3P1C	Isoform 2 of Kunitz-type protease inhibitor 1
4	2P1C, 3P2C, 1P1C, 4P2C	Isoform 2 of DNA endonuclease RBBP8
3	1P2C, 2P2H, 1P2H	Leucine-rich alpha-2-glycoprotein
3	4P1C, 4P2C, 3P1C	Ig gamma-4 chain C region
3	3P2C, 4P2C, 4P2H	Hemopexin
3	2P1C, 1P1C, 1P1H	Fibrinogen alpha chain
2	1P1C, 1P2H	Transcription factor AP-2-delta
2	3P2C, 4P2C	Isoform 2 of Guanine nucleotide-binding protein G(olf) subunit alpha
2	3P2C, 4P2C	Isoform 3 of Periostin
2	2P1C, 3P1C	Isoform 4 of Aldehyde dehydrogenase family 8 member A1
2	4P1C, 3P1C	Isoform 2 of Putative uncharacterized protein C1orf145
2	1P1C, 4P2C	Galectin-4
2	1P1C, 3P1C	Dehydrogenase/reductase SDR family member 7C
2	3P2C, 4P2C	Small integral membrane protein 17
Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
2	2P2C, 3P2C	G-protein coupled receptor-associated sorting protein 1
2	4P2H, 1P1H	Isoform 2 of Rho guanine nucleotide exchange factor 5
2	4P1C, 3P1C	Isoform 2 of Echinoderm microtubule-associated protein-like 3
2	3P1H, 1P1H	Proenkephalin-A

2	3P2C, 4P2C	Uncharacterized protein FLJ14100
2	2P1C, 3P1C	Isoform 2 of Receptor-type tyrosine-protein phosphatase H
2	1P1C, 2P1C	CAD protein
2	3P2C, 3P1C	Isoform 4 of Probable E3 ubiquitin-protein ligase HERC4
2	3P2C, 2P2C	FYVE, RhoGEF and PH domain-containing protein 2
2	1P2C, 4P2C	Inter-alpha-trypsin inhibitor heavy chain H2
2	2P1C, 4P2C	Thrombospondin type-1 domain-containing protein 7B
2	2P1C, 4P2C	Zinc finger C2HC domain-containing protein 1B
2	2P2C, 4P2C	Isoform 3 of Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
2	1P2C, 3P2C	RNA-binding motif protein, X-linked-like-3
2	1P2C, 3P1C	Isoform 3 of NAD-dependent protein deacetylase sirtuin-2
2	4P1C, 1P1H	Peroxisomal 2,4-dienoyl-CoA reductase
2	4P1C, 3P1C	Mucin-16
2	2P1C, 2P2C	POTE ankyrin domain family member D
2	2P1C, 2P2H	Zinc finger protein 836
1	2P1H	Isoform 2 of Cadherin EGF LAG seven-pass G-type receptor 3
1	1P1C	Isoform 2 of Dual specificity protein phosphatase 4
1	3P1C	Zinc finger protein 816
1	1P2H	Kelch-like protein 30
1	3P1C	Arylsulfatase D
1	1P1C	Isoform 2 of Zinc finger protein 75D
1	3P1C	Isoform 2 of Heterogeneous nuclear ribonucleoprotein M
1	2P1C	Isoform 5 of Calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1A
1	1P2C	Laminin subunit beta-2
1	2P1C	Hepatocyte nuclear factor 3-alpha
1	4P1C	ATP synthase subunit epsilon, mitochondrial
1	3P1C	RWD domain-containing protein 2B

Number of samples with contaminant	Names of samples with contaminant	Peptide
1	3P2C	Ubiquitin-associated and SH3 domain-containing protein A
1	2P2C	Sestrin-2
1	1P1H	Ribonuclease P protein subunit p38
1	4P2C	Polyhomeotic-like protein 1
1	3P1C	Isoform 2 of DNA-dependent protein kinase catalytic subunit
1	3P1C	Isoform B of Disintegrin and metalloproteinase domain-containing protein 17
1	1P2C	Isoform 4 of Double-stranded RNA-specific editase 1
1	3P2C	Interferon-related developmental regulator 2
1	2P1C	Isoform 3 of Chromodomain-helicase-DNA-binding protein 3
1	3P1C	Cystine/glutamate transporter
1	2P1C	C-Jun-amino-terminal kinase-interacting protein 3

1	4P2C	Phosphatidylinositol phosphatase PTPRQ
1	3P2C	Synergin gamma
1	3P1C	Calpain-11
1	2P1C	Isoform 4 of Unconventional myosin-VI
1	3P2H	E3 ubiquitin-protein ligase HECTD1
1	2P1C	Uncharacterized protein KIAA1257
1	3P1C	MORC family CW-type zinc finger protein 2
1	3P1C	Secernin-1
1	3P2C	Isoform 2 of tRNA (guanine(26)-N(2))-dimethyltransferase
1	4P2C	Isoform 3 of Protein flightless-1 homolog
1	3P2C	Protein-glutamine gamma-glutamyltransferase K
1	2P1C	Proteasome subunit alpha type-1
1	3P1C	Epidermal growth factor receptor substrate 15
1	4P1C	DNA repair protein complementing XP-G cells
1	2P1C	Tripeptidyl-peptidase 2
1	1P2C	SHC-transforming protein 1
1	2P1C	Sodium- and chloride-dependent GABA transporter 1
1	3P1C	Replication factor C subunit 4
1	2P1C	Histamine H1 receptor
1	4P1C	Chloride channel protein 1
1	3P2C	Keratin, type I cytoskeletal 9
1	2P1C	Transcription factor SOX-11
1	2P1C	Isoform 4 of Breast cancer type 1 susceptibility protein

Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
1	2P1C	Isoform ID-B of DNA-binding protein inhibitor ID-1
1	3P1C	Nebulin
1	3P1C	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform
1	4P2C	Isoform 3 of Centrosomal protein kizuna
1	1P1C	Surfeit locus protein 2
1	4P1C	DNA-binding protein RFX7
1	3P1C	Uncharacterized protein KIAA0753
1	1P1C	Tumor suppressor p53-binding protein 1
1	4P2C	Chromatin assembly factor 1 subunit B
1	3P1C	Tastin
1	4P1C	Forkhead box protein D4
1	1P2C	Isoform 2 of General transcription factor 3C polypeptide 1
1	1P2C	Ankyrin-3
1	3P2C	Prostaglandin D2 receptor
1	4P1C	F-box/LRR-repeat protein 3
1	3P1C	Sperm protein associated with the nucleus on the X chromosome B/F
1	3P2C	Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1
1	4P1C	Isoform 2 of Cyclic nucleotide-gated cation channel beta-3
1	4P1C	Putative heat shock protein HSP 90-beta 4

1	4P1C	Spermatogenesis-associated protein 31A4
1	4P2C	Sodium/hydrogen exchanger 10
1	3P2C	Otogelin-like protein
1	1P2C	Laminin subunit gamma-2
1	4P1C	Spermatogenesis-associated protein 31A6
1	2P1C	Isoform 23 of Voltage-dependent L-type calcium channel subunit alpha-1C
1	1P1C	Isoform 2 of Cyclin-dependent kinase 13
1	2P1C	Kelch-like ECH-associated protein 1
1	2P1H	LIM and SH3 domain protein 1
1	2P1C	Isoform 5 of DAZ-associated protein 2
1	3P2C	Histone-lysine N-methyltransferase SETDB1
1	3P1C	Isoform 2 of Homeobox protein MOX-1
1	2P1C	Isoform 3 of Xaa-Pro dipeptidase
1	1P1C	CD226 antigen
1	3P2C	Striated muscle preferentially expressed protein kinase
1	3P2C	Putative pro-MCH-like protein 1
1	4P2C	Isoform 2 of Guanylate kinase
1	4P1C	UPF0532 protein C7orf60
Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
1	1P2C	Isoform 4 of Interleukin-23 receptor
1	2P1C	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 7
1	3P1C	Collagen alpha-1(VII) chain
1	1P1C	Isoform 2 of Fibrous sheath-interacting protein 2
1	4P2C	Isoform 4 of Transcription factor Sp3
1	3P2C	Angiopoietin-1 receptor
1	4P1C	Ras association domain-containing protein 7
1	3P2C	Cyclic AMP-responsive element-binding protein 5
1	1P1C	Fc receptor-like protein 6
1	3P2C	Fc receptor-like B
1	1P1C	Cell division cycle-associated protein 2
1	1P1C	Signal-induced proliferation-associated 1-like protein 2
1	1P1C	Intraflagellar transport protein 46 homolog
1	4P2C	Putative nascent polypeptide-associated complex subunit alpha-like protein
1	4P1C	L-lactate dehydrogenase A-like 6B
1	2P1C	TBC1 domain family member 2A
1	2P2C	Gasdermin-C
1	1P2C	Isoform 3 of Polycomb group RING finger protein 6
1	4P2C	Jerky protein homolog-like
1	4P2C	Isoform 3 of Centrosomal protein of 170 kDa protein B
1	3P2C	Isoform 2 of Protocadherin gamma-C4
1	1P1C	Protocadherin alpha-8
1	2P1C	Isoform 2 of GMP reductase 2
1	3P1C	Leucine--tRNA ligase, cytoplasmic
1	3P2C	Disco-interacting protein 2 homolog B
1	1P1C	Isoform 3 of KN motif and ankyrin repeat domain-containing protein 2

1	3P2C	Denticleless protein homolog
1	3P2C	Protocadherin Fat 2
1	2P2C	Dynein intermediate chain 1, axonemal
1	4P1C	Isoform 2 of Zinc finger protein 200
1	2P1C	Voltage-dependent N-type calcium channel subunit alpha-1B
1	2P1C	Isoform 2 of OTU domain-containing protein 4
1	1P2C	43 kDa receptor-associated protein of the synapse
1	4P2C	PRAME family member 13
1	4P1C	Spermatogenesis-associated protein 31A3
1	1P2C	Isoform 4 of Terminal uridylyltransferase 7

Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
1	3P2C	N-alpha-acetyltransferase 35, NatC auxiliary subunit
1	1P1C	Isoform 4 of Armadillo repeat-containing protein 3
1	4P1C	Collagen alpha-1(XV) chain
1	1P1C	Selenoprotein P
1	4P1C	Protein kinase C gamma type
1	3P1C	Mediator of RNA polymerase II transcription subunit 14
1	3P2C	Isoform 2 of Unconventional myosin-IXa
1	3P1C	Collagen alpha-5(VI) chain
1	3P1C	Zinc finger protein 891
1	1P1C	Isoform 2 of Zinc finger protein 785
1	1P2C	RIMS-binding protein 3B
1	3P1C	RanBP2-like and GRIP domain-containing protein 3
1	1P2C	RIMS-binding protein 3C
1	2P1C	Isoform 3 of Protein FAM149A
1	3P1C	Ankyrin repeat domain-containing protein 34B
1	2P1H	DNA excision repair protein ERCC-6-like 2
1	3P1C	Reticulon-3
1	2P1C	Zinc finger and BTB domain-containing protein 11
1	3P1C	UHRF1-binding protein 1-like
1	3P2C	Homeobox protein aristaless-like 3
1	2P2C	Isoform 4 of Protein-methionine sulfoxide oxidase MICAL2
1	3P2C	Isoform 2 of A-kinase anchor protein SPHKAP
1	4P2C	Translin-associated factor X-interacting protein 1
1	1P1C	Cytoplasmic tRNA 2-thiolation protein 2
1	1P2C	SKI family transcriptional corepressor 2
1	3P2C	Isoform 4 of Leucine-rich repeat flightless-interacting protein 1
1	3P1C	UDP-N-acetylhexosamine pyrophosphorylase-like protein 1
1	3P1C	Isoform 2 of Mitochondrial 10-formyltetrahydrofolate dehydrogenase
1	4P1C	Putative POM121-like protein 1
1	1P1C	Fragile X mental retardation syndrome-related protein 1
1	3P2C	Afamin
1	3P2C	DNA ligase 4
1	4P2C	Neuralized-like protein 1A
1	3P2C	Transmembrane protein 114

Number of samples with contaminant	Names of samples with contaminant	Peptide
1	2P1C	Isoform 5 of Serine/threonine-protein kinase Sgk1
1	3P2C	Krueppel-like factor 4
1	4P1C	Plasminogen
1	1P2C	Isoform 4 of Titin
1	4P1C	Isoform 4 of Deoxynucleoside triphosphate triphosphohydrolase SAMHD1
1	4P2H	Zinc finger and BTB domain-containing protein 32
1	1P1C	Mitogen-activated protein kinase kinase kinase 2
1	1P1C	Isoform 2 of 5-azacytidine-induced protein 2
1	4P2C	Isoform C of Potassium channel subfamily K member 7
1	3P1C	Isoform 2 of Hepatocyte nuclear factor 3-beta
1	3P2C	TRAF3-interacting JNK-activating modulator
1	1P1C	Isoform Short of Protein jagged-2
1	4P1C	Tenascin-N
1	3P1H	Serine/arginine repetitive matrix protein 2
1	3P1C	Dynamin-3
1	4P1C	Ig gamma-1 chain C region
1	2P1C	Sodium/nucleoside cotransporter 1
1	4P1C	Ig alpha-1 chain C region
1	2P2C	Integrin beta-2
1	2P1C	Collagen alpha-1(IV) chain
1	1P1C	Spectrin alpha chain, erythrocytic 1
1	3P2C	Complement factor D
1	4P1C	Isoform 2 of Protein-glutamine gamma-glutamyltransferase 6
1	1P1C	U3 small nucleolar RNA-interacting protein 2
1	1P1C	Isoform 5 of Potassium voltage-gated channel subfamily KQT member 2
1	4P1C	Ig alpha-2 chain C region
1	4P2C	Metallo-beta-lactamase domain-containing protein 1
1	4P1C	Spermatogenesis-associated protein 31C2
1	2P2C	Dual specificity protein phosphatase CDC14C
1	2P2C	Isoform 2 of Protein SSX5 OS=Homo sapiens GN=SSX5
1	1P1C	Nucleolar pre-ribosomal-associated protein 1
1	2P1C	T-cell receptor beta chain V region CTL-L17
1	4P1C	Isoform 2 of Zinc finger E-box-binding homeobox 2
1	3P1C	Cathepsin D
1	4P1C	Putative uncharacterized protein LOC100131404
1	2P1C	Laminin subunit beta-1
Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
1	4P1C	Insulin-like growth factor 1 receptor
1	3P2C	Isoform 2 of 72 kDa type IV collagenase
1	3P1C	Isoform 2 of Complement factor H
1	3P2C	Complement C1s subcomponent
1	4P2C	Coiled-coil domain-containing protein 175

Number of samples with contaminant	Names of samples with contaminant	Peptide
1	1P1C	Zinc finger protein 806
1	2P1C	Isoform 11 of Myocardial zonula adherens protein
1	2P2C	Putative upstream-binding factor 1-like protein 6
1	4P1C	Isoform 2 of Spermatogenesis-associated protein 31C1
1	2P2C	Laminin subunit gamma-1
1	3P1C	Putative alpha-1-antitrypsin-related protein
1	3P2C	Cation-independent mannose-6-phosphate receptor
1	1P2C	Isoform 9 of Troponin T, cardiac muscle
1	2P1C	Melanoma-associated antigen 3
1	3P1C	Isoform 2 of Integrin beta-1-binding protein 1
1	3P1C	Isoform 2 of Colorectal mutant cancer protein
1	4P1C	Zinc finger protein 213
1	1P1C	Cyclin-G-associated kinase
1	3P2C	Isoform 2 of Rho guanine nucleotide exchange factor 10
1	1P1C	Histone-lysine N-methyltransferase SETD1A
1	4P1C	DNA-directed RNA polymerases I and III subunit RPAC1
1	1P1C	Isoform 2 of Kinesin-like protein KIF26B
1	1P1C	Eukaryotic translation initiation factor 3 subunit D
1	4P2C	Suppressor of cytokine signaling 1
1	3P2H	Cholinesterase
1	1P1C	Lysosomal alpha-glucosidase
1	2P1H	Isoform MLC3 of Myosin light chain 1/3, skeletal muscle isoform
1	4P1C	ATP-dependent DNA helicase Q5
1	3P2C	Isoform 2 of Peroxisome biogenesis factor 10
1	4P2C	Isoform 2 of Protein diaphanous homolog 2
1	4P2C	Isoform 3 of Beta-1,4-galactosyltransferase 2
1	4P1C	Isoform 2 of 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase eta-2
1	1P1C	SLIT-ROBO Rho GTPase-activating protein 2
1	3P1C	Obscurin-like protein 1
1	1P1C	Vacuolar protein sorting-associated protein 4B
1	1P1C	Serine/threonine-protein kinase/endoribonuclease IRE1
Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
1	2P1C	WD repeat and HMG-box DNA-binding protein 1
1	2P1C	Stanniocalcin-2
1	2P1C	Neurocan core protein
1	1P1C	Isoform 2 of Striatin
1	4P2C	Acetyl-CoA carboxylase 2
1	1P1C	Cytochrome P450 2A7
1	1P1C	Ubiquitin D
1	2P2C	Segment polarity protein dishevelled homolog DVL-2
1	3P1C	Cyclin-dependent kinase 1
1	3P2C	Isoform 3 of Yorkie homolog
1	1P1C	Ras GTPase-activating-like protein IQGAP1
1	4P2C	3-hydroxyanthranilate 3,4-dioxygenase
1	4P1H	Nestin

1	2P2C	Mannan-binding lectin serine protease 1
1	3P2C	Centromere protein F
1	3P1C	Isoform 2 of Cysteine--tRNA ligase, cytoplasmic
1	3P2C	Homeobox even-skipped homolog protein 1
1	2P2C	Filaggrin
1	2P1C	Isoform 8 of cAMP-responsive element modulator
1	3P1C	Proteolipid protein 2
1	1P1C	Isoform 2 of Coiled-coil domain-containing protein 67
1	1P2C	Isoform 2 of Elongation of very long chain fatty acids protein 1
1	1P1C	Zinc finger protein 304
1	4P1C	Isoform 1 of SH2 domain-containing protein 2A
1	2P2C	Matrix metalloproteinase-25
1	1P1C	Proline-rich nuclear receptor coactivator 2
1	3P1C	Transmembrane protein 9B
1	3P1C	Isoform 2 of Ubiquitin carboxyl-terminal hydrolase CYLD
1	4P2C	Isoform 5 of Uncharacterized protein KIAA1683
1	4P2C	Isoform 3 of Paternally-expressed gene 3 protein
1	4P2C	Isoform 5 of Pyrin and HIN domain-containing protein 1
1	1P1C	Isoform 3 of SET domain-containing protein 5
1	3P1C	Zinc finger protein 578
1	4P1C	Isoform 2 of Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 3
1	4P1C	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 1

Number of samples with contaminant	Names of samples with contaminant	Peptide
1	2P2H	Isoform 4 of Vacuolar protein sorting-associated protein 13A
1	3P1C	Pleckstrin homology domain-containing family F member 1
1	4P1C	Zinc finger protein 532
1	4P2C	Glutathione peroxidase 7
1	3P2C	Isoform 2 of CDK5 regulatory subunit-associated protein 2
1	4P1C	Zinc finger and SCAN domain-containing protein 10
1	4P1C	Isoform 2 of Selenocysteine insertion sequence-binding protein 2
1	3P1C	Isoform 17 of Mucin-4
1	3P1C	Protein kinase C-binding protein NELL2
1	2P1C	Isoform 4 of Collagen alpha-1(XII) chain
1	3P2C	BRCA1-associated RING domain protein 1
1	4P1C	Selenoprotein S
1	4P1C	Putative uncharacterized protein LOC642776
1	3P1C	Ribonuclease P protein subunit p25
1	4P2C	Isoform 5 of Protein LAP2
1	2P1C	Mucolipin-3
1	4P1C	E3 ubiquitin-protein ligase RNF14
1	3P2C	Isoform 2 of E3 ubiquitin-protein ligase rififylin

1	1P1C	Izumo sperm-egg fusion protein 1
1	1P1C	Calcineurin subunit B type 2
1	2P2C	Proline-rich protein 22
1	1P2C	Zinc finger protein 431
1	2P2C	Zinc finger protein 473
1	3P2C	General transcription factor 3C polypeptide 2
1	1P1C	Butyrophilin subfamily 2 member A2
1	1P1C	Secretogranin-3
1	1P1C	Voltage-dependent calcium channel gamma-8 subunit
1	2P1C	Contactin-associated protein-like 5
1	3P2C	Ras and Rab interactor 2
1	1P1C	Isoform 2 of Uncharacterized protein C22orf15
1	4P1C	Translational activator GCN1
1	3P1C	Uncharacterized protein DKFZp434B061
1	3P2C	Minor histocompatibility protein HA-1
1	1P1C	Isoform 2 of NEDD4-binding protein 2-like 2
1	2P1C	TRAF family member-associated NF-kappa-B activator
1	3P2C	Rho guanine nucleotide exchange factor 2
1	1P1C	Isoform 2 of Adrenocortical dysplasia protein homolog
Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
1	2P1H	Isoform 2 of RUN domain-containing protein 1
1	1P1C	Cysteine protease ATG4C
1	2P1C	Isoform 2 of Leucine-rich repeat-containing protein 51
1	2P1C	Golgi apparatus membrane protein TVP23 homolog C
1	1P1C	Diamine acetyltransferase 2
1	1P2C	Target of EGR1 protein 1
1	3P1C	RNA-binding protein 41
1	3P2C	Isoform 3 of 3-hydroxybutyrate dehydrogenase type 2
1	1P1C	Isoform 3 of Protein LSM14 homolog B
1	3P1C	Engulfment and cell motility protein 2
1	1P1C	Isoform 2 of Roundabout homolog 2
1	2P1C	Collectrin
1	1P2C	Isoform 3 of Otoferlin
1	4P1C	Oxysterol-binding protein-related protein 1
1	1P2C	Protein BEX2
1	3P1C	RING finger protein 26
1	1P1C	Isoform 2 of Serine palmitoyltransferase 3
1	2P1C	Isoform 3 of Ubiquitin carboxyl-terminal hydrolase 16
1	2P1C	Peptidyl-prolyl cis-trans isomerase FKBP7
1	2P1C	Transforming acidic coiled-coil-containing protein 3
1	3P1C	Brefeldin A-inhibited guanine nucleotide-exchange protein 1
1	4P2C	Isoform 3 of Protein MRVI1
1	1P1C	Cohesin subunit SA-3
1	1P1C	Isoform 2 of N-acetyl-D-glucosamine kinase

1	1P1C	ATPase inhibitor, mitochondrial
1	3P1C	Isoform 2 of Anaphase-promoting complex subunit 11
1	3P2C	Potassium voltage-gated channel subfamily D member 1
1	4P2H	Phenylalanine--tRNA ligase beta subunit
1	2P1C	Isoleucine--tRNA ligase, mitochondrial
1	1P1C	UDP-glucuronic acid/UDP-N-acetylgalactosamine transporter
1	3P2C	E3 ubiquitin-protein ligase RNF146
1	2P2C	Probable 8-oxo-dGTP diphosphatase NUDT15
1	1P1C	39S ribosomal protein L9, mitochondrial
1	3P1C	F-box only protein 28
1	2P1C	ATP-dependent RNA helicase DDX18

Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
1	2P2C	Protein kintoun
1	2P1C	Isoform 6 of Sphingomyelin phosphodiesterase 4
1	1P1C	Isoform 2 of Testis-expressed sequence 10 protein
1	1P2C	Protein BEX1
1	2P2C	Potassium channel subfamily K member 12
1	2P1C	Isoform 4 of Pleckstrin homology domain-containing family A member 5
1	4P1C	Kv channel-interacting protein 4
1	1P2C	Protocadherin-9
1	2P1C	Isoform 3 of Methyl-CpG-binding domain protein 2
1	2P1C	Zinc finger protein 106
1	3P2C	CD59 glycoprotein
1	2P2H	Isoform L-VEGF189 of Vascular endothelial growth factor A
1	3P1C	Isoform Alpha-1 of N-chimaerin
1	2P1C	Integrin beta-4
1	2P1C	Zinc finger protein 823
1	1P1C	Isoform 4 of Cyclic AMP-dependent transcription factor ATF-7
1	4P1C	5,6-dihydroxyindole-2-carboxylic acid oxidase
1	1P1C	Isoform E of Plasma membrane calcium-transporting ATPase 1
1	1P1C	Isoform 5 of Isthmin-2
1	1P1C	Isoform 4 of Serine/threonine-protein phosphatase 4 regulatory subunit 3A
1	1P1H	DNA-directed RNA polymerase I subunit RPA2
1	1P1C	WW domain-containing transcription regulator protein 1
1	1P1C	Spectrin beta chain, non-erythrocytic 4
1	4P2C	Transcription elongation factor A protein-like 2
1	4P1C	Echinoderm microtubule-associated protein-like 4
1	4P2C	Mucin-13
1	4P1C	Isoform 2 of Band 4.1-like protein 1
1	2P1C	Zinc fingers and homeoboxes protein 3
1	4P2C	Occludin/ELL domain-containing protein 1
1	1P1C	Synapse differentiation-inducing gene protein 1

1	2P2C	Isoform 2 of Methyltransferase-like protein 8
1	1P1C	Zinc finger protein 395
1	3P1C	Isoform 4 of Queuine tRNA-ribosyltransferase subunit QTRTD1
1	3P1C	SET and MYND domain-containing protein 4
1	4P1C	MAP/microtubule affinity-regulating kinase 4

Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
1	1P1C	Isoform P of Kinesin light chain 1
1	1P1C	5'-AMP-activated protein kinase subunit gamma-2
1	3P1C	Chondroadherin-like protein
1	1P2C	Mis18-binding protein 1
1	4P2C	Tetratricopeptide repeat protein GNN
1	2P1C	Mitochondrial fission regulator 2
1	2P1C	Zinc finger protein 530
1	3P1C	Isoform 5 of Protein piccolo
1	4P1C	Zinc finger CCCH domain-containing protein 14
1	3P2C	Sushi domain-containing protein 1
1	3P1C	Hydroxyacid-oxoacid transhydrogenase, mitochondrial
1	2P1C	VEGF co-regulated chemokine 1
1	2P2C	Isoform 2 of Hydroxysteroid dehydrogenase-like protein 2
1	3P2C	Leucine-, glutamate- and lysine-rich protein 1
1	1P1C	Isoform 3 of E3 ubiquitin-protein ligase Arkadia
1	2P2C	Isoform 3 of Probable helicase senataxin
1	2P1C	Isoform 2 of Beclin 1-associated autophagy-related key regulator
1	1P1C	Putative uncharacterized protein FLJ45999
1	4P1C	Rho guanine nucleotide exchange factor 18
1	1P1C	Uncharacterized protein KIAA0408
1	4P2C	Maestro heat-like repeat family member 5
1	3P1C	Isoform 6 of Serine/threonine-protein kinase Nek10
1	3P2C	Ubiquitin carboxyl-terminal hydrolase 34
1	2P2C	Isoform 2 of 5-hydroxytryptamine receptor 3D
1	1P1C	Centromere protein U
1	3P1C	Isoform 4 of Ankyrin repeat and KH domain-containing protein 1
1	2P2C	Serine/threonine-protein kinase LMTK2
1	1P2C	Peflin
1	3P1C	Isoform 2 of Protein TILB homolog
1	2P2H	Isoform 3 of Uncharacterized protein CXorf57
1	4P2H	Zinc finger protein 782
1	4P2C	Isoform 2 of Ubiquitin-protein ligase E3B
1	3P1C	Isoform 2 of Coiled-coil domain-containing protein 18
1	4P2C	Zinc finger protein 415
1	4P2H	Probable methyltransferase TARBP1
1	2P1C	Zinc finger protein 211
1	4P1C	Spermatogenesis-associated protein 31A1
1	3P1C	Fidgetin

Number of samples with	**Names of samples with**	**Peptide**

contaminant	contaminant	
1	3P1C	Uncharacterized protein C1orf189
1	4P2C	A-kinase anchor protein 4
1	4P2C	Glycine receptor subunit alpha-4
1	4P1C	Spermatogenesis-associated protein 31A2
1	4P1C	Isoform 2 of Ventral anterior homeobox 1
1	4P2C	PRAME family member 14
1	1P1C	RNA-binding protein 20
1	3P2C	Isoform 2 of Centrosomal protein of 162 kDa
1	1P1C	Isoform 3 of Transient receptor potential cation channel subfamily M member 1
1	1P1C	Protein SOGA3
1	2P2C	Isoform 6 of Meiosis inhibitor protein 1
1	2P1C	Isoform 3 of Coiled-coil domain-containing protein 181
1	4P1C	Spermatogenesis-associated protein 31A5
1	2P1C	Plakophilin-3
1	1P1C	Isoform C of Collagen alpha-1(XI) chain
1	1P1C	Delta-aminolevulinic acid dehydratase
1	4P2C	Isoform 2 of Sodium/potassium-transporting ATPase subunit alpha-3
1	2P1C	Methionine aminopeptidase 1D, mitochondrial
1	2P1C	Cadherin-like protein 26
1	3P1C	Isoform 2 of Nitric oxide-inducible gene protein
1	4P2C	NANOG neighbor homeobox
1	1P2C	RIMS-binding protein 3A
1	1P2C	Isoform 3 of Sodium channel protein type 11 subunit alpha
1	2P1C	Coiled-coil domain-containing protein 17
1	2P2C	WD repeat-containing protein 52
1	2P1H	Isoform 3 of Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 3
1	4P1C	Isoform 3 of Sulfatase-modifying factor 1
1	1P1H	Isoform 2 of Probable G-protein coupled receptor 116
1	4P2C	Lysozyme g-like protein 1
1	2P1C	2-oxoglutarate and iron-dependent oxygenase domain-containing protein 1
1	2P1C	Potassium channel regulatory protein
1	4P2C	Isoform 2 of RING finger protein 10
1	2P1C	Isoform 2 of Ephexin-1
1	2P1C	Isoform 5 of Methyltransferase-like protein 13
1	3P1C	Zinc finger protein 611
1	4P2C	Zinc finger protein 713
1	3P1C	Zinc finger protein 565
1	1P2C	Interferon-induced very large GTPase 1

Number of samples with contaminant	Names of samples with contaminant	Peptide
1	3P2C	RBBP8 N-terminal-like protein
1	2P1C	Isoform 3 of Transmembrane channel-like protein 2
1	4P2C	Isoform 2 of Protein FAM98A
1	3P1C	Cytoplasmic dynein 2 heavy chain 1
1	2P1C	Isoform 2 of Coiled-coil domain-containing protein

		150
1	4P2C	GTPase IMAP family member 8
1	4P2C	Isoform 2 of B-cell scaffold protein with ankyrin repeats
1	4P2C	Isoform 3 of Nesprin-1
1	4P2H	Biorientation of chromosomes in cell division protein 1-like 1
1	3P2C	AT-rich interactive domain-containing protein 1B
1	3P1C	Isoform 2 of Alstrom syndrome protein 1
1	2P2C	Up-regulator of cell proliferation
1	3P2C	Isoform 2 of Chromodomain-helicase-DNA-binding protein 6
1	4P2C	Isoform 3 of Ubiquitin carboxyl-terminal hydrolase 33
1	3P2C	Centrosomal protein of 192 kDa
1	1P1C	Isoform 2 of 39S ribosomal protein L43, mitochondrial
1	2P1C	Putative HIN1-like protein
1	2P1C	Isoform 2 of Diacylglycerol kinase eta
1	3P2C	Uncharacterized protein C9orf85
1	4P2C	E3 ubiquitin-protein ligase HACE1
1	3P1C	tRNA-specific adenosine deaminase 2
1	2P1C	ATP-binding cassette sub-family A member 13
1	3P2C	Zinc finger protein 429
1	3P1C	Isoform 5 of Immunoglobulin-like and fibronectin type III domain-containing protein 1
1	2P1C	Isoform 2 of Nebulin-related-anchoring protein
1	4P2C	Glucocorticoid-induced transcript 1 protein
1	1P1C	Transmembrane protein 200A
1	3P1C	Uncharacterized protein CXorf67
1	3P1C	Iron-sulfur cluster co-chaperone protein HscB, mitochondrial
1	1P2C	Isoform 3 of Protein FAM131B
1	3P2C	Histone-lysine N-methyltransferase SUV420H2
1	1P1C	Butyrophilin subfamily 2 member A1
1	1P1C	Copine-8
1	4P2C	Isoform 2 of Transmembrane protease serine 6
Number of samples with contaminant	**Names of samples with contaminant**	**Peptide**
1	3P1C	Cartilage intermediate layer protein 2
1	2P2H	LysM and putative peptidoglycan-binding domain-containing protein 2
1	3P2C	Isoform 3 of Protein AHNAK2
1	2P2C	Isoform 8 of Neuron navigator 2
1	3P2H	Cyclin-dependent kinase-like 3
1	1P2C	Choline transporter-like protein 2
1	1P1C	Inositol 1,4,5-trisphosphate receptor-interacting protein
1	4P1C	Spermatogenesis-associated protein 31A7
1	4P2C	MAX gene-associated protein
1	1P1C	Isoform 3 of Transmembrane and ubiquitin-like domain-containing protein 2

1	1P2C	Stromal interaction molecule 1

Buffers preparation:

0,1% (w/v) SDS solution

SDS	0.25 g
Add deionized water to a final volume of:	250 ml

Denaturing buffer 1 (8 M Urea, 0.4 M ammonium bicarbonate, 0.1% (w/v) SDS)

Urea	4.8 g
Ammonium bicarbonate	0.316 g
0.1% (w/v) SDS	100 μl
Add deionized water to a final volume of:	10 ml

Denaturing buffer 2 (8 M Urea, 0.2 M Tris-HCl; 0.1% (w/v) SDS)

Urea	4.8 g
Tris-HCl	0.315 g
0.1% (w/v) SDS	100 μl
Add deionized water to a final volume of:	10 ml

Sodium periodate (100 mM)

Sodium periodate	0.021 g
Add deionized water to a final volume of:	1 ml

Tris (2-carboxyethyl) phosphine (100 mM)

Tris (2-carboxyethyl) phosphine	0.029 g
Add deionized water to a final volume of:	1 ml

Iodoacetamide (200 mM)

Iodoacetamide	0.037 g
Add deionized water to a final volume of:	1 ml

Iodoacetamide (0.5 M)

Iodoacetamide	0.093 g
Add deionized water to a final volume of:	1 ml

Potassium phosphate buffer (100 mM, pH 8.0)

Potassium phosphate monobasic	0.136 g
Add deionized water to a volume of:	8 ml
Adjust pH to 8.0	
Add deionized water to a final volume of:	10 ml

Ammonium bicarbonate buffer (100 mM, pH 8.3)

Ammonium bicarbonate	0.320 g
Add deionized water to a volume of:	35 ml
Adjust pH to 8.3	
Add deionized water to a final volume of:	40 ml

Dithiothreitol (1 M)

DTT	0.154 g
Add deionized water to a final volume of:	1 ml

Tris-HCl (0.2 M)

Tris-HCl	0.315 g
Add deionized water to a final volume of:	10 ml

0.1% (v/v) Trifluoroacetic acid

99% TFA	0.2 ml
Add deionized water to a final volume of:	200 ml

50% ACN in 0.1% TFA

ACN	20 ml
99% (v/v) TFA	40 µl
Add deionized water to a final volume of:	40 ml

80% ACN in 0.1% TFA

ACN	32 ml
99% (v/v) TFA	40 µl
Add deionized water to a final volume of:	40 ml

80% ACN in H_2O

ACN	32 ml
Add deionized water to a final volume of:	40 ml

5% ACN in 0.1% Formic acid

ACN	50 µl
98% (v/*v)* FA	1 µl
Add deionized water to a final volume of:	1000 µl